

UNIWERSYTET ŚLĄSKI  
WYDZIAŁ MATEMATYKI, FIZYKI, CHEMII  
Ośrodek dydaktyczny w Rybniku.

KIERUNEK: FIZYKA  
SPECJALNOŚĆ: EKONOFIZYKA

„Modelowanie szeregów czasowych za pomocą procesów ARMA i ARIMA”  
(aplikacje w systemie SAS)

Sebastian Zając

Praca licencjacka wykonana  
pod kierunkiem  
dr Jacka Syski

## Spis treści

Wstęp.....	4
Przedstawienie problemu identyfikacji .....	5
Rozdział 1. Modele ARMA i ARIMA .....	9
1.1. Stacjonarne szeregi ARMA.....	10
1.2. Przyczynowe i odwracalne procesy ARMA.....	10
• Definicja 1.2.2 Procesu ARMA.....	11
• Przykład 1.2.1 Proces średniej ruchomej MA(q).....	12
• Przykład 1.2.3 Proces autoregresji AR(1).....	14
• Definicja 1.2.3 Przyczynowego procesu ARMA.....	22
• Definicja 1.2.4 Procesu odwracalnego ARMA.....	22
1.3. Stacjonarny proces AR(p).....	23
1.4. Pierwiastki wielomianów $\phi(z)$ i $\theta(z)$ dla procesów ARMA(p, q) a przyczynowość, odwracalność i stacjonarność procesów ARMA.....	29
1.5. Procesy średniej ruchomej nieskończonego rzędu MA( $\infty$ ).....	31
• Definicja 1.5 Procesu MA( $\infty$ ).....	31
1.6. Wyznaczanie funkcji autokowariancji dla procesu ARMA(p, q).....	36
• Metoda pierwsza.....	36
• Metoda druga.....	38
• Przykład 1.6.1 (rozwiązanie metodą 1 i 2).....	40
1.7. Funkcja autokorelacji cząstkowej (PACF).....	43
• Definicja 1.7.1 Funkcji autokorelacji cząstkowej PACF.....	44
• Definicja 1.7.2 Funkcji autokorelacji cząstkowej z próby.....	44
• Przykład 1.7.1 Proces AR(1).....	45
• Przykład 1.7.2 Proces AR(p).....	47
• Przykład 1.7.3 Proces MA(1).....	47
1.8. Procesy niestacjonarne ARIMA.....	49
• Definicja 1.8.1 Procesu zintegrowanego.....	50
• Definicja 1.8.2 Procesu ARIMA .....	50
Rozdział 2. Estymacja parametrów $\mu$ oraz $\rho(\cdot)$ .....	52
2.1. Estymacja średniej $\mu$ .....	52
2.2. Estymacja funkcji autokowariancji $\gamma(\cdot)$ i autokorelacji $\rho(\cdot)$ .....	54

Rozdział 3. Predykcja modeli ARMA i ARIMA.....	60
3.1. Przewidywanie dla procesów stacjonarnych.....	60
3.2. Równanie predykcji.....	60
3.3. Równanie dla predykcji z jednym krokiem.....	61
3.4. Równania dla przewidywania z krokiem $h$ , ( $h \geq 1$ ).....	65
3.5. Metody rekurencyjne dla wyznaczania najlepszych przewidywań.....	65
• Przewidywanie rekurencyjne z wykorzystaniem algorytmu Durbin'a – Levinson'a.....	66
• Przewidywanie rekurencyjne z wykorzystaniem algorytmu innowacyjnego.....	67
3.6. Obliczenia rekurencyjne dla przewidywania $h$ – tego kroku, $h \geq 1$ .....	71
3.7. Przewidywanie rekurencyjne dla procesów ARMA z jednym krokiem.....	73
3.8. Przewidywanie rekurencyjne dla procesów ARMA $h$ – tego kroku, $h \geq 1$ .....	76
3.9. Predykcja niestacjonarnych procesów ARIMA.....	78
Rozdział 4. Estymacja parametrów $\phi$ , $\theta$ modeli ARMA.....	81
4.1. Estymacja procesów autoregresyjnych AR(p).....	82
• Równanie Yule'a - Walker'a.....	82
• Wstępna estymacja procesu autoregresji z wykorzystaniem algorytmu Durbin'a – Levinson'a.....	86
• Przykład 4.1 Proces AR (1).....	89
• Przykład 4.2 Proces AR (2) z wykorzystaniem algorytmu Durbin'a – Levinson'a.....	90
4.2. Estymacja procesów średniej ruchomej MA(q).....	92
• Przykład 4.3 Proces MA(2) z wykorzystaniem algorytmu innowacyjnego.....	96
4.3. Wstępna estymacja parametrów procesów ARMA.....	98
4.4. Estymatory metody największej wiarygodności i najmniejszych kwadratów dla procesów ARMA.....	99
4.5. Asymptotyczne własności estymatorów MNW i MNK.....	103
• Przykład 4.4 Proces AR(p).....	104
• Przykład 4.5 Proces MA(q).....	105
• Przykład 4.6 Proces ARMA(1,1).....	106
4.6. Przedziały ufności dla parametrów przyczynowych i odwracalnych procesów ARMA.....	106
4.7. Asymptotyczna normalność estymatorów parametrów.....	107
Rozdział 5. Kryteria wyboru rzędu parametrów $p$ i $q$ .....	110
5.1. Kryterium FPE.....	110
5.2. Kryterium AICC.....	111
Zakończenie.....	115
Uzupełnienie 1.....	116
Uzupełnienie 2.....	127
Uzupełnienie 3.....	130
Bibliografia .....	135

## Wstęp

Niniejsza praca zawiera wprowadzenie do modelowania szeregów czasowych za pomocą liniowych modeli ARIMA. Modelowanie dokonywane będzie z wykorzystaniem systemu statystycznego SAS (Statistical Analysis System). Program ten jest bogatym narzędziem analizy nie tylko szeregów czasowych, ale również bardzo różnych dziedzin analizy statystycznej.

W pierwszym rozdziale omówiono podstawowe definicje i własności modeli stacjonarnych szeregów ARMA, oraz niestacjonarnych szeregów ARIMA. Jest to grupa liniowych modeli służących do opisu rzeczywistych procesów stochastycznych. Okazuje się bowiem, że szeroka klasa stacjonarnych procesów może być utworzona z wykorzystaniem białego szumu jako wymuszającej siły pojawiającej się w (zbiorze) liniowych równań różnicowych, które są związane z pojęciem procesów ARMA (procesy z autoregresją (AR) i ruchomą średnią (MA)). Chociaż definicję tę podam jeszcze raz, to już tutaj sprecyzujemy zagadnienie, którego dotyczyć będzie praca, wprowadzając definicję (1.8) procesu ARMA ( $p, q$ ):

*Proces stochastyczny  $\{X_t, t = 0, \pm 1, \pm 2, \dots\}$  jest procesem ARMA( $p, q$ ), jeśli  $\{X_t\}$  jest stacjonarny, oraz jeśli dla każdego  $t$ , proces  $\{X_t\}$  spełnia równanie różnicowe:*

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}$$

*gdzie  $\{Z_t\} \sim \text{WN}(0, \sigma^2)$  jest białym szumem, a  $\phi_{(\cdot)}$  i  $\theta_{(\cdot)}$  są parametrami modelu.*

W pracy przedstawiono również sposoby wyznaczania funkcji autokowariancji, autokorelacji oraz autokorelacji częściowej wraz z przykładami, gdyż są one istotne w identyfikacji danego procesu.

W rozdziałach 2 i 4 podano metody estymacji parametrów  $\mu$ ,  $\rho(\cdot)$  (Rozdział 2), oraz  $\phi_{(\cdot)}$  i  $\theta_{(\cdot)}$  (Rozdział 4) wraz z przedziałami ufności. W rozdziale 3 przedstawiono główny cel analizy, czyli predykcję szeregów czasowych. Natomiast rozdział 5 przedstawia kryteria wyboru parametrów  $p$  i  $q$ .

W uzupełnieniu 2 podano metodę wyznaczania estymatorów metodą największej wiarygodności dla procesu Gaussowskiego, oraz wskazano przykładowe testy losowości reszt. (Test Durбина-Watsona wraz z przykładem opisany jest dokładnie w pracy [2]).

Niniejsza praca opiera się zasadniczo na obszernym opracowaniu teorii szeregów czasowych Brockwell'a i Davis'a [1], jednak liczne przeliczenia dowodów i przykładów zostały wykonane samodzielnie.

*Podstawowym celem pracy jest przedstawienie współczesnego podejścia do problemu identyfikacji właściwego modelu ARMA dla zaobserwowanego szeregu czasowego, estymacji parametrów tego modelu, oraz predykcji.*

## **Przedstawienie problemu identyfikacji.**

Niech  $\{X_t\}$  oznacza szereg skorygowany o średnią. Problemem jest znalezienie najbardziej satysfakcjonującego modelu ARMA( $p, q$ ) reprezentującego  $\{X_t\}$ . Jeśli  $p$  i  $q$  są znane z góry to należy skorzystać z techniki estymacji parametrów opisanych w Rozdziale 4. Jednakże zazwyczaj nie mamy do czynienia z taką sytuacją, więc podstawową sprawą jest identyfikacja odpowiednich wartości  $p$  i  $q$ . Na pierwszy rzut oka mogłoby się wydawać, że im większe wartości  $p$  i  $q$ , tym lepiej będzie model dopasowany. Na przykład, jeśli dopasowaliśmy całą sekwencję procesów AR( $p$ )  $p = 1, 2, \dots$ , wtedy wartość estymatora  $\hat{\sigma}^2$  metody największej wiarygodności (MNW) parametru  $\sigma^2$ , będącego wariancją białego szumu w modelu ARMA, spada zasadniczo w sposób monotoniczny wraz ze wzrostem  $p$ . Jednakże musimy uważać na niebezpieczeństwo przefitowania, czyli na zbyt dokładne dopasowanie modelu do poszczególnych obserwacji. Skrajny przypadek przefitowania zdarzyłby się, jeśli wielomian stopnia  $n$ -tego dopasujemy do  $n+1$  obserwacji wygenerowanych z modelu  $Y_t = a + bt + Z_t$ , gdzie  $\{Z_t\}$  jest niezależnym ciągiem standaryzowanych normalnych zmiennych losowych. *W takiej sytuacji dopasowanie będzie idealne dla obserwacji otrzymanych w próbie, ale obok skomplikowanej postaci modelu, użycie go do predykcji przyszłych wartości może być obarczone dużymi błędami.*

Rozwinięto szereg kryteriów, w szczególności kryterium AIC Akaike'a i kryterium FPE, które próbują zapobiegać przefitowaniu przez skuteczne wyznaczenie nakładu związanego z wprowadzeniem każdego dodatkowego parametru. Sprawa ta jest analogiczna do zastąpienia w zwykłych modelach regresji współczynnika determinacji  $R^2$  współczynnikiem  $R^2_{adj}$ , który w przeciwieństwie do  $R^2$  nie zawsze wzrasta przy dodaniu nowej zmiennej do modelu. Fakt ten wynika z równoczesnego spadku liczby stopni swobody dla reszt modelu, co wiąże się z ewentualnym spadkiem średniej wariancji wewnątrzgrupowej, oznaczanej zazwyczaj jako MSE.

W Rozdziale 5 omówimy poprawioną już, na brak systematycznego błędu, formułę kryterium AICC, zdefiniowaną następująco dla modelu ARMA( $p, q$ ) z wektorowymi współczynnikami parametrów  $\vec{\phi} = (\phi_1, \phi_2, \dots, \phi_p)$  i  $\vec{\theta} = (\theta_1, \theta_2, \dots, \theta_q)$  :

$$AICC(\vec{\phi}, \vec{\theta}) = -2 \ln L(\vec{\phi}, \vec{\theta}, S(\vec{\phi}, \vec{\theta})) / n + 2(p + q + 1)n / (n - p - q - 2) \quad (W.1)$$

gdzie  $L(\vec{\phi}, \vec{\theta}, \sigma^2)$  jest wiarygodnością danych (próbki) dla danych opisywanych Gaussowskim modelem ARMA z parametrami  $(\vec{\phi}, \vec{\theta}, \sigma^2)$ , a  $S(\vec{\phi}, \vec{\theta}, \sigma^2)$  jest sumą kwadratów reszt zdefiniowaną w Rozdziale 4.

Na podstawie analizy opisaney w Rozdziale 5, przedstawimy wybór modelu dla którego wartość AICC jest minimalna.

Intuicyjnie możemy myśleć, że wielkość  $2(p + q + 1)n / (n - p - q - 2)$  w powyższym równaniu (W.1) podaje graniczną wartość, której przekroczenie świadczy o przefitowaniu. Jeśli już znajdziemy model, który minimalizuje wartość AICC, to następnie musimy sprawdzić dobroć dopasowania (zasadniczo oznacza to sprawdzenie czy reszty to biały szum gaussowski). Problem ten nie będzie omawiamy w pracy (w Uzupełnieniu 2 podano podstawowe testy losowości reszt).

Wprowadzenie statystyki AICC (lub analogicznej) redukuje identyfikację modelu do dobrze zdefiniowanego zadania. Jednakże szukanie modelu, który minimalizuje AICC może być w przypadku braku jakiegoś pomysłu na klasę badanych modeli, bardzo czasochłonne. Dlatego należy użyć rozmaitych technik, które pomagają przyspieszyć proces wyboru modelu.

Najważniejszymi narzędziami używanymi do wskazywania wartości  $p$  i  $q$  są funkcja autokorelacji (ACF) i częściowa funkcja autokorelacji (PACF) w próbie, (Uzupełnienia oraz Rozdział 1), a w przypadku parametrów modelu, wstępne estymatory  $\hat{\phi}_m$  i  $\hat{\theta}_m$   $m = 1, 2, \dots$  parametrów  $\vec{\phi}$  i  $\vec{\theta}$ , które omówiono w Rozdziale 4. Metody estymacji opisane w tym rozdziale umożliwią nam znalezienie, dla zadanych wartości  $p$  i  $q$ , właściwego modelu ARMA( $p, q$ ), który dobrze dopasuje się do otrzymanej serii danych. Aby procedura ta miała zastosowanie musi istnieć co najmniej uzasadnione przypuszczenie, że dane są faktycznie realizacją procesu ARMA i szczególnie, że są one realizacją procesu stacjonarnego. Jeśli dane wykazują własności sugerujące niestacjonarność (np. trend i sezonowość), wtedy może być konieczne zrobienie transformacji, tak aby otrzymać po ich zastosowaniu nowy szereg, który jest bardziej kompatybilny z założeniami stacjonarności.

Odchylenie od stacjonarności może być sugerowane przez wykres szeregu czasowego lub funkcję autokorelacji.

Budowa wykresu szeregu ujawnia czasami silną zależność zmienności od poziomu szeregu, a w takim razie dane powinny zostać w pierw przekształcone, tak by zredukować bądź całkiem wyeliminować tę zależność. Gdy np. widać na wykresie liniowy wzrost amplitudy zmienności procesu  $\{U_t\}$  wraz ze wzrostem  $t$ , wtedy można dokonać transformacji  $V_t = \ln U_t$ , dzięki czemu nowy otrzymany wykres może nie ukazywać żadnego wzrostu zmienności  $V_t$  [3]. Transformacja logarymiczna jest faktycznie odpowiednia zawsze gdy  $\{U_t\}$  jest szeregiem którego odchylenie standardowe wzrasta liniowo ze średnią.

Wykaz ogólnych klas transformacji można znaleźć w różnych pozycjach literaturowych. Jedną z ogólnych klas transformacji tworzą transformacje Box – Cox’a. Równanie opisujące transformację  $f_\lambda$  Box–Cox’a ma postać [1]:

$$f_\lambda(U_t) = \begin{cases} \lambda^{-1}(U_t^\lambda - 1), & U_t \geq 0, \quad \lambda > 0 \\ \ln U_t, & U_t > 0, \quad \lambda = 0 \end{cases} \quad (\text{W.2})$$

Jednak pomimo dokonania eliminacji trendu i sezonowości wciąż możliwe jest, że funkcja autokorelacji w próbie może odpowiadać procesowi niestacjonarnemu lub prawie niestacjonarnemu, co jest sygnałem, że być może należałoby zastosować procedurę różnicowania opisaną w pierwszym rozdziale pracy (Rozdział 1, (1.142-146)).

Identyfikacja dla „czystego” procesu AR(p) i MA(q) jest rozsądna z bezpośrednim użyciem ACF i PACF i wstępnych estymatorów  $\hat{\phi}_m$  i  $\hat{\theta}_m$  oraz wskaźnika AICC.

Jednakże dla procesów ARMA(p, q) z obu wartościami  $p$  i  $q$  różnymi od zera, ACF i PACF jest dużo trudniejsza do interpretacji. Dlatego szukamy bezpośrednio wartości  $p$  i  $q$  takich, że AICC jest bliskie minimalnemu. Poszukiwanie może zostać przeprowadzone na wiele sposobów np. przez próbowanie wszystkich wartości  $p$  i  $q$  takich że  $p + q = 1$ , potem  $p + q = 2$  itd., lub alternatywnie postępować wg następujących kroków:

1. Użyj MNW dla dopasowania procesu ARMA klasy (1,1), (2,2) .... do danych, następnie wybierz model z najmniejszą wartością AICC
2. Zaczynij z modelu ARMA z minimalnym AICC eliminując jeden lub więcej współczynników, kierując się błędem standardowym estymowanych współczynników, maksymalizując prawdopodobieństwo dla każdego zredukowanego modelu i oszacuj wartość AICC
3. wybierz model z najmniejszym AICC

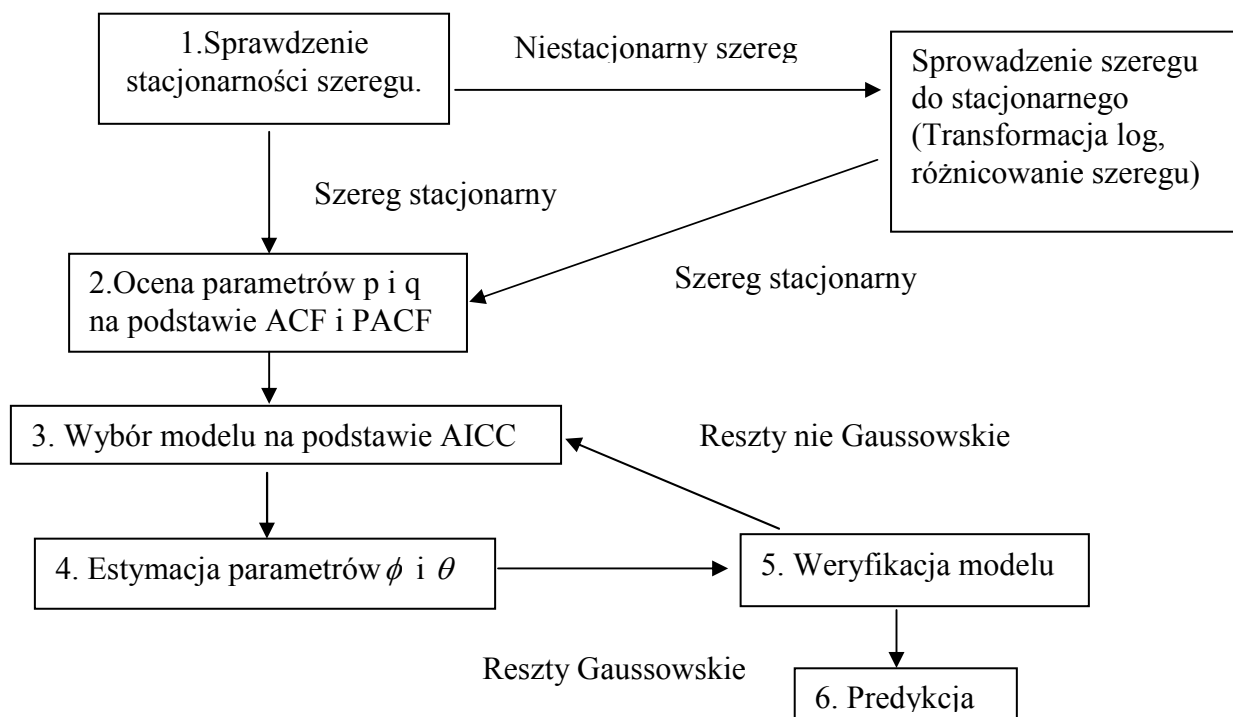
Jeżeli  $q - p < 0$  to funkcja autokorelacji składa się z kombinacji znikających funkcji wykładniczych i/lub sinusoid tłumionych, natomiast jeśli  $q - p \geq 0$  to wystąpi  $q - p + 1$  początkowych wartości współczynnika autokorelacji  $\rho_0, \rho_1, \dots, \rho_{q-p}$  nie podlegających temu ogólnemu schematowi. Funkcja autokorelacji cząstkowej procesu ARMA(p, q) rozciąga się w nieskończoność, przy czym dla dużych odstępów zachowuje się jak funkcja autokorelacji cząstkowej procesu średniej ruchomej, czyli – zależnie od rzędu średniej ruchomej i wartości parametrów – przeważają w niej człony typu znikających funkcji wykładniczych i/lub sinusoid tłumionych.

Box i Jenkins [4] proponowali dopasowanie procesów o coraz wyższych wielkościach parametrów  $p$  i  $q$  aż do momentu, kiedy reszty z tak dopasowanego procesu nie wykazują autokorelacji. Sugeruje się, że wystarczy do tego celu dopasować model, dla którego  $p + q < 3$ .

### Poniżej przedstawiono proces postępowania

1. Sprowadzenie procesu do szeregu stacjonarnego (transformacje log, różnicowanie)
2. Identyfikacja parametrów  $p$  i  $q$  (poprzez funkcję ACF i PACF)
3. Wybór modelu (wg kryterium AICC – model z wartością minimalną)
4. Estymacja parametrów  $\phi, \theta$ .
5. Weryfikacja modelu - sprawdzenie czy reszty są rozkładem Gaussowskim.
6. Predykcja szeregu

### Diagram postępowania





## Rozdział 1. Modele ARMA i ARIMA

W rozdziale tym przedstawiono ogólne definicje procesów ARMA i ARIMA dla opisu obserwacji  $\{X_t, t = 1, 2, \dots, n\}$ , bardzo ważnej klasy liniowych modeli do opisu procesów stochastycznych.

Jeśli dane

- (1) nie wykazują żadnych oczywistych odchyżeń od stacjonarności
- (2) mają szybko malejącą funkcję autokorelacji,

będziemy szukali odpowiednich procesów ARMA dla przedstawienia danych skorygowanych o średnią.

O ile warunki (1) i (2) nie są spełnione, wtedy szukamy transformacji danych, która wygeneruje nowy szereg, który będzie spełniał powyższe warunki (często przekształcenie wyjściowych danych sprowadza się do różnicowania, co sugeruje nam użycie ogólniejszej niż ARMA, klasy procesów ARIMA (auto regressive-integrated moving average)).

Kiedy dane zostały przekształcone i otrzymaliśmy szereg stacjonarny jedynym problemem staje się znalezienie satysfakcjonującego modelu ARMA( $p, q$ ) i wyznaczenie (zidentyfikowanie)  $p$  i  $q$ . Przydatnych wskazówek w tym wyborze mogą dostarczyć: funkcja autokorelacji i częściowej autokorelacji z próby (Rozdział 1) oraz wstępne estymatory  $\hat{\phi}_m$  i  $\hat{\theta}_m$  (Rozdział 4). Jednakże naszym najważniejszym kryterium w wyborze modelu będzie AICC, zmodyfikowana wersja kryterium Akaike'a AIC (Rozdział 5). Według tego kryterium obliczymy estymatory największej wiarygodności parametrów  $\phi$  i  $\theta$  oraz  $\sigma^2$  dla odpowiednie dobranego  $p$  i  $q$  (wybieramy dopasowany model z najmniejszym AICC).

Jeśli dopasowany model jest satysfakcjonujący reszty powinny przypominać biały szum Gaussowski (test sprawdzający tę własność opisany jest w [2], w uzupełnieniach wymieniono kilka testów) i powinny być przeprowadzone dla minimalnego modelu AICC (tzn. takiego, dla którego kryterium AICC ma najmniejszą wartość) tak żeby mieć pewność, że reszty są zgodne z ich oczekiwanym zachowaniem w tym modelu.

Jeśli reszty w „minimalnym” modelu AICC nie są zgodne z białym szumem Gaussowskim wtedy powinien zostać sprawdzony inny model (w którym kryterium AICC jest bliskie minimum), aż do momentu znalezienia takiego modelu, który przechodzi testy dobroci dopasowania.

W niektórych przypadkach mała różnica w wartości AICC (umownie mniej niż 2) między dwoma zadowalającymi modelami może zostać zignorowana dla prostoty modelu. W późniejszej części pracy rozważymy predykcję procesów ARIMA (Rozdział 3).

Wprowadzimy teraz pojęcie klasy modeli ARMA dla reprezentowania szeregów czasowych. Uogólnieniem tej klasy, która włącza szeroki zakres niestacjonarnych szeregów jest klasa procesów ARIMA tzn. takich procesów, które po skończeniu wielokrotnym różnicowaniu redukują się do procesów ARMA.

## 1.1 Stacjonarne procesy ARMA

Niech  $\{X_t\}$ ,  $(t=0, \pm 1, \pm 2 \dots)$  (1.1)

jest procesem spełniającym liniowe równanie różnicowe ze stałymi współczynnikami.

Narzucenie warunków stałych współczynników prowadzi do zdefiniowania parametrycznej rodziny stacjonarnych procesów, zwanych procesami ARMA.

### Twierdzenie 1.1

Dla dowolnej funkcji autokowariancji  $\gamma(\cdot)$  dla której  $\lim_{h \rightarrow \infty} \gamma(h) = 0$  (1.2)

i dla dowolnego  $k > 0$  można znaleźć proces ARMA z funkcją autokowariancji  $\gamma_x(\cdot)$ ,

która spełnia warunek  $\gamma_x(h) = \gamma(h)$ ,  $h = 0, 1, \dots, k$ . (1.3)

### Zastosowanie:

1. z powodu powyższego Twierdzenia 1.1 procesy ARMA odgrywają kluczową rolę w modelowaniu danych pochodzących z szeregów czasowych
2. liniowa struktura procesów ARMA prowadzi do prostej teorii liniowej predykcji.

## 1.2 Przyczynowe i odwracalne procesy ARMA

### Założenie:

Niech  $\{X_t\}$   $t = 0, \pm 1, \pm 2 \dots$  oraz niech zmienne losowe  $X_t$  są niezależne i mają takie same rozkłady ze średnią zero i wariancją  $\sigma^2$  tzn.  $\{X_t\} \sim iid(0, \sigma^2)$ . (1.4)

Z punktu widzenia własności drugiego rzędu, tzn. ignorując wszystkie własności rozkładów łącznych procesu  $\{X_t\}$ , za wyjątkiem własności drugiego rzędu tzn. tych, które możemy wyprowadzić z momentów  $E(X_t)$  oraz  $E(X_s X_t)$ , możemy procesy te identyfikować z klasą wszystkich stacjonarnych procesów mających średnią zero i funkcją autokowariancji :

$$\gamma(h) = \begin{cases} \sigma^2 & \text{dla } h=0 \\ 0 & \text{dla } h \neq 0 \end{cases} \quad (1.5)$$

### **Definicja 1.2.1 białego szumu (white noise)**

Proces  $\{Z_t\}$  jest nazywany białym szumem ze średnią 0 i wariancją  $\sigma^2$ , co zapisujemy:

$$\{Z_t\} \sim \text{WN}(0, \sigma^2), \quad (1.6)$$

wtedy i tylko wtedy, gdy  $\{Z_t\}$  ma średnią zero i kowariancję zadaną wzorem (1.5).

$$\text{Często zakładamy } \{Z_t\} \sim \text{iid}(0, \sigma^2), \quad (1.7)$$

co oznacza, że zmienne mają ten sam rozkład i są niezależne.

### **Zastosowanie:**

Szeroka klasa stacjonarnych procesów może być utworzona z wykorzystaniem białego szumu jako wymuszającej siły pojawiającej się w (zbiorze) liniowych równań różnicowych, które są związane z pojęciem procesów ARMA (procesy z autoregresją (AR) i ruchomą średnią (MA)).

### **Definicja 1.2.2 procesu ARMA(p, q)**

$\{X_t, t = 0, \pm 1, \pm 2 \dots\}$  jest procesem ARMA(p, q), jeśli  $\{X_t\}$  jest stacjonarny, oraz jeśli dla każdego t, proces  $\{X_t\}$  spełnia równanie różnicowe:

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q} \quad (1.8)$$

gdzie  $\{Z_t\} \sim \text{WN}(0, \sigma^2)$ .

Mówimy że proces  $\{X_t\}$  jest procesem ARMA(p, q) ze średnią  $\mu$ , jeśli

$$\{X_t - \mu\} \quad (1.9)$$

jest procesem ARMA(p, q).

Równanie (1.8) zapisujemy :

$$\phi(B)X_t = \theta(B)Z_t, \quad t = 0, \pm 1, \pm 2 \quad (1.10)$$

gdzie:

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p \quad (1.11)$$

jest wielomianem stopnia  $p$

$$\theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \dots + \theta_q z^q \quad (1.12)$$

jest wielomianem stopnia  $q$ , natomiast  $B$  jest operatorem cofnięcia:

$$B^j X_t = X_{t-j}, \quad \text{gdzie } B^j \equiv B B^{j-1}, \quad j = 0, \pm 1, \pm 2, \dots \quad (1.13)$$

Wielomian  $\phi$  nazywamy wielomianem autoregresji, a wielomian  $\theta$  nazywamy wielomianem średniej ruchomej, równania różnicowego (1.10).

W przypadku, jeśli w modelu nie występują elementy związane ze średnią ruchomą  $\theta_1 = \dots = \theta_q = 0$  ( $q = 0$ ), model sprowadza się do modelu autoregresyjnego AR( $p$ ).

Procesy dla których  $\phi_1 = \dots = \phi_p = 0$  ( $p = 0$ ) są tzw. procesami z ruchomą średnią rzędu  $q$  (MA( $q$ )).

Procesy ARMA służą jako klasa procesów porównawczych dla testowania stacjonarnych szeregów czasowych. Z ich pomocą można skutecznie estymować dużą klasę stacjonarnych procesów, a więc osiągnąć również ostateczny cel analizy, jakim jest predykcja nowych obserwacji.

### **Przykład 1.2.1 Proces średniej ruchomej MA( $q$ )**

Założenie:  $\phi(z) \equiv 1$ , wtedy równanie (1.10) daje:

$$X_t = \theta(B)Z_t \quad (1.14)$$

i proces nazywamy procesem z ruchomą średnią rzędu  $q$  (MA( $q$ )), czyli taki w którym tylko początkowe  $q$  wag jest różne od zera.

Równanie (1.14) można rozpisać następująco:

$$X_t = \theta_0 Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q} \quad \text{gdzie } \{Z_t\} \sim \text{WN}(0, \sigma^2) \quad (1.14')$$

natomiast  $\theta_0, \theta_1, \dots, \theta_q$  są parametrami modelu, a  $q$  - wielkością opóźnienia.

Równanie różnicowe (1.10) sprowadza się do (1.14) więc (1.14) jest wtedy jedynym rozwiązaniem równania (1.10).

Należy zwrócić uwagę, że nazwa „średnia ruchoma” jest nieco myląca, gdyż wagi  $1, \theta_0, \theta_1, \dots, \theta_q$  nie muszą się sumować do jedności ani nie muszą być dodatnie.

Proces  $MA(q)$  jest zawsze stacjonarny, niezależnie od wartości parametrów.

Istotnie:

$$E(X_t) = E(\theta(B)Z_t) = E\left(\sum_{j=0}^q \theta_j Z_{t-j}\right) = \sum_{j=0}^q \theta_j E(Z_{t-j}) = \sum_{j=0}^q \theta_j \cdot 0 = 0 \quad (1.15)$$

gdzie  $\theta_0 = 1$

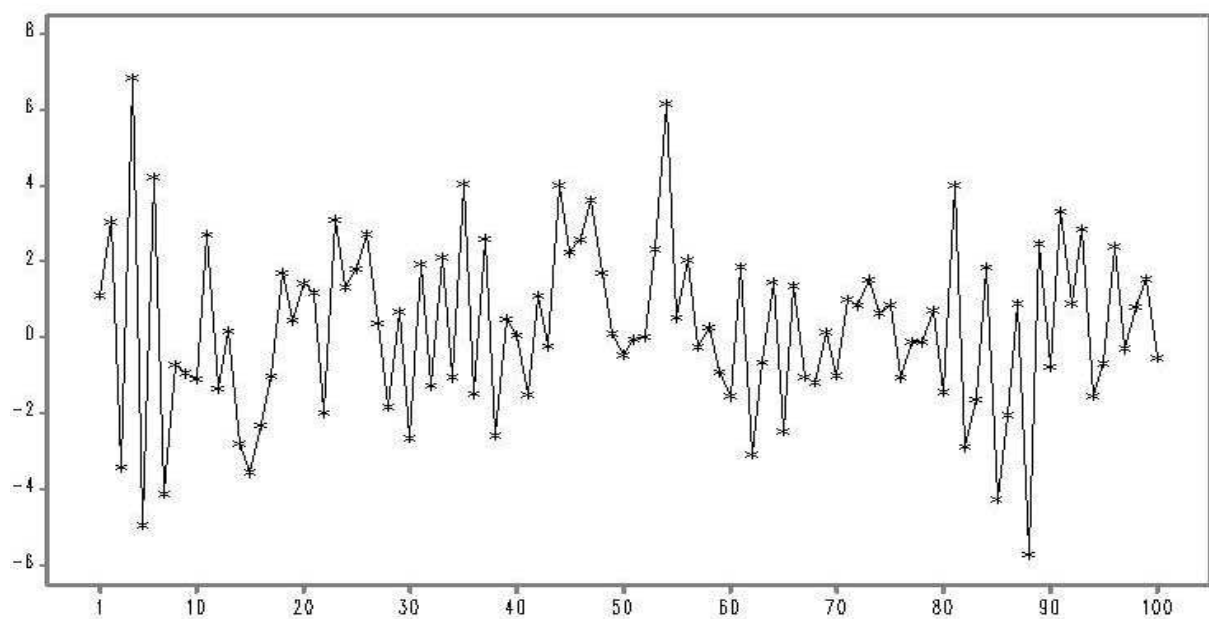
$$\text{cov}(X_{t+h}, X_t) = E(X_{t+h} \cdot X_t) - EX_{t+h} \cdot EX_t = \begin{cases} \sigma^2 \sum_{j=0}^{q-|h|} \theta_{j+|h|} \theta_j & |h| \leq q \\ 0 & |h| > q \end{cases} \quad (1.16)$$

gdzie skorzystano z faktu, że  $\theta_j = 0$  dla  $j > q$ .

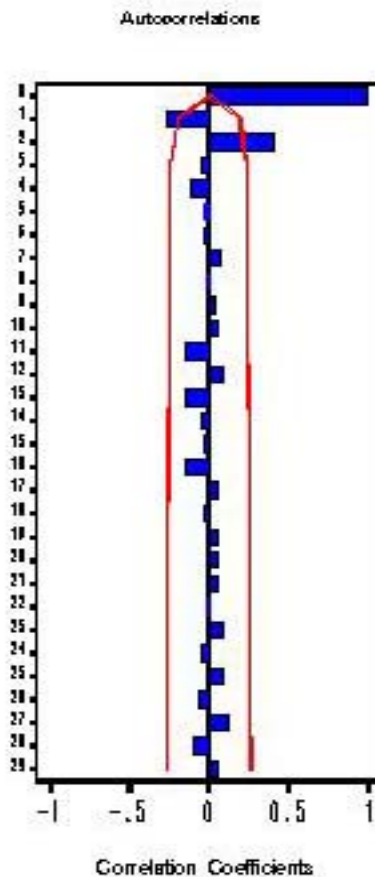
Dowód wymaga jedynie skorzystania z (1.14') i (1.5).

**Przykład 1.2.2.** Rozważmy proces  $MA(2)$   $X_t = Z_t + 0,3Z_{t-1} - 0,7Z_{t-2}$ , gdzie  $Z_t \sim WN(0,2)$ .

Wartości białego szumu wygenerowano dla 100 chwil z rozkładu normalnego  $N(0, 2)$  wykorzystując generator liczb losowych znajdujący się w programie SAS w module `Analyst->Data->Random Variates`. Poniżej zamieszczono wykres otrzymanej realizacji szeregu czasowego oraz odpowiadający mu wykres tzw. autokorelogramu.



Wykres 1. 100 obserwacji procesu  $X_t = Z_t + 0,3Z_{t-1} - 0,7Z_{t-2}$ , gdzie  $Z_t \sim WN(0,2)$



Wykres 2. Funkcja Autokorelacji (ACF) procesu  $X_t = Z_t + 0,3Z_{t-1} - 0,7Z_{t-2}$ , gdzie  $Z_t \sim WN(0,2)$ .

Otrzymany kształt autokorelogramu ma cechy, które wskazują na jego pochodzenie z procesu MA(2), tzn. nieistotność wartości funkcji ACF dla  $q > 2$ . Problem ten zostanie omówiony w później przy omawianiu Funkcji ACF dla procesów MA(q).

### **Przykład 1.2.3 Proces AR(p), $p = 1$ .**

Założenie:  $\theta(z) \equiv 1$ . Wtedy z (1.10) otrzymujemy:

$$\phi(B)X_t = Z_t \quad (1.17)$$

i proces nazywamy procesem autoregresji rzędu  $p$  (AR(p)).

Gdy  $p = 1$  wtedy:  $\phi(z) = 1 - \phi_1 z \quad (1.18)$

tzn.  $\phi(B) = 1 - \phi_1 B \quad (1.19)$

zatem:  $(1 - \phi_1 B)X_t = Z_t \quad (1.20)$

$$X_t - \phi_1 X_{t-1} = Z_t \quad (1.20')$$

$$X_t = Z_t + \phi_1 X_{t-1} \quad (1.21)$$

Zastosujemy procedurę iteracyjną do (1.21) z wykorzystaniem:

$$X_{t-1} = Z_{t-1} + \phi_1 X_{t-2}, \quad X_{t-2} = Z_{t-2} + \phi_1 X_{t-3}, \quad \dots, \quad X_{t-k} = Z_{t-k} + \phi_1 X_{t-k-1} \quad (1.22)$$

$$X_t = Z_t + \phi_1 Z_{t-1} + \phi_1^2 Z_{t-2} + \phi_1^3 Z_{t-3} + \dots + \phi_1^k Z_{t-k} + \phi_1^{k+1} X_{t-k-1} = \sum_{j=0}^k \phi_1^j Z_{t-j} + \phi_1^{k+1} X_{t-k-1} \quad (1.23)$$

$$\text{Jeśli proces } \{X_t\} \text{ jest stacjonarny (więc: } EX_t = 0\text{), wtedy } \|X_t\|^2 = E(X_t^2) \quad (1.24)$$

$$\text{jest stałe. Jeśli } |\phi_1| < 1, \text{ wtedy } \lim_{k \rightarrow \infty} \phi_1^k = 0 \quad (1.25)$$

i z (1.23) otrzymujemy:

$$\left\| X_t - \sum_{j=0}^k \phi_1^j Z_{t-j} \right\|^2 = \left\| \phi_1^{k+1} X_{t-k-1} \right\|^2 = \phi_1^{2k+2} \|X_{t-k-1}\|^2 \xrightarrow{k \rightarrow \infty} 0. \quad (1.26)$$

Sprawdźmy czy sumy częściowe szeregu

$$\sum_{j=0}^{\infty} \phi_1^j Z_{t-j} \quad (1.27)$$

tworzą ciąg Cauchy'ego.

$$\text{Suma częściowa dla tego szeregu ma postać } S_n = \sum_{j=0}^n \phi_1^j Z_{t-j} \quad (1.28)$$

$$\begin{aligned} \text{Niech } n \geq m: \quad & \|S_n - S_m\|^2 = \left\| \sum_{j=0}^n \phi_1^j Z_{t-j} - \sum_{j'=0}^m \phi_1^{j'} Z_{t-j'} \right\|^2 = \\ & = E \left\{ \left( \sum_{j=1}^n \phi_1^{j1} Z_{t-j1} - \sum_{j2=0}^m \phi_1^{j2} Z_{t-j2} \right) \left( \sum_{j3=0}^n \phi_1^{j3} Z_{t-j3} - \sum_{j4=0}^m \phi_1^{j4} Z_{t-j4} \right) \right\} = \\ & = \sum_{j1=0}^n \sum_{j3=0}^n \phi_1^{j1} \phi_1^{j3} \delta_{j1,j3} + \sum_{j2=0}^m \sum_{j4=0}^m \phi_1^{j2} \phi_1^{j4} \delta_{j2,j4} - \sum_{j1=0}^n \sum_{j4=0}^m \phi_1^{j1} \phi_1^{j4} \delta_{j1,j4} - \sum_{j2=0}^m \sum_{j3=0}^n \phi_1^{j2} \phi_1^{j3} \delta_{j2,j3} = \\ & = \sum_{j1=0}^n (\phi_1^{j1})^2 + \sum_{j4=0}^m (\phi_1^{j4})^2 - \sum_{j4=0}^m (\phi_1^{j4})^2 - \sum_{j2=0}^m (\phi_1^{j2})^2 = \sum_{j1=0}^n (\phi_1^{j1})^2 - \sum_{j2=0}^m (\phi_1^{j2})^2 \end{aligned} \quad (1.29)$$

Gdy  $|\phi_1| < 1$ , wtedy szereg 
$$\sum_{j=0}^n (\phi_1^j)^2 \quad (1.30)$$

jest szeregiem geometrycznym (z pierwszym wyrazem  $(\phi_1^0)^2 = 1$ ), zbieżnym

$$\sum_{j=0}^n (\phi_1^j)^2 \xrightarrow{n \rightarrow \infty} \frac{1}{1 - \phi_1^2}, \quad (1.31)$$

zatem dla  $n, m \rightarrow \infty$  z (1.29) otrzymujemy: 
$$\|S_n - S_m\|_{n, m \rightarrow \infty} \rightarrow 0, \quad (1.32)$$

co oznacza, że zgodnie z kryterium Cauchy'ego, szereg (1.27) tworzy ciąg Cauchy'ego.

Natomiast fakt, że ciąg  $\{S_n\}$  jest ciągiem Cauchy'ego, oznacza, że ze względu na zupełność przestrzeni Hilberta  $L^2$  (Uzupełnienie 1), której elementami są  $S_n$ , istnieje zmienna losowa  $S \in L^2$ , taka że 
$$S_m \xrightarrow[m.s.]{} S, \quad (1.33)$$

gdzie m.s. oznacza zbieżność średnio – kwadratową, tzn.

$$\|S_m - S\|^2 = E|S_m - S|^2 \rightarrow 0, \text{ gdy } m \rightarrow \infty. \quad (1.34)$$

Zgodnie z Uzupełnieniem 1 stwierdzamy, że ponieważ szereg  $S_n = \sum_{j=0}^n \phi_1^j Z_{t-j}$  tworzy ciąg Cauchy'ego, zatem z (1.26) i (1.34) wiemy, że istnieje zmienna losowa

$$X_t = S, \text{ a dokładnie } P(X_t = S) = 1, \quad (1.35)$$

a mianowicie:

$$X_t = \sum_{j=0}^{\infty} \phi_1^j Z_{t-j} \quad (1.36)$$

do której jest zbieżny szereg  $S_n$  w znaczeniu zbieżności średnio – kwadratowej i zbieżności bezwzględnej z prawdopodobieństwem równym 1.

Równość dwóch zmiennych  $S$  i  $X_t$  z prawdopodobieństwem równym jeden, tzn.

$$P(X_t = S) = 1, \text{ oznacza, że } X_t(\omega) = \sum_{j=0}^{\infty} \phi_1^j Z_{t-j}(\omega) \text{ dla wszystkich } \omega \notin E, \quad (1.37)$$

gdzie  $E$  jest podzbiorem zdarzeń przestrzeni  $\Omega$ , które zachodzą z prawdopodobieństwem równym zero (Uzupełnienie 1).



Proces  $\{X_t\}$  określony równaniem (1.36) jest stacjonarny (słabo) (Uzupełnienie 1).

Istotnie: 
$$EX_t = E\left(\sum_{j=0}^{\infty} \phi_1^j Z_{t-j}\right) = \sum_{j=0}^{\infty} \phi_1^j EZ_{t-j} = 0 \quad (1.38)$$

$$\begin{aligned} \text{cov}(X_{t+h}, X_t) &= E(X_{t+h} \cdot X_t) = E\left(\sum_{j=0}^{\infty} \phi_1^j Z_{t+h-j} \sum_{j'=0}^{\infty} \phi_1^{j'} Z_{t-j'}\right) = \sum_{j=0}^{\infty} \sum_{j'=0}^{\infty} \phi_1^j \phi_1^{j'} E(Z_{t+h-j} Z_{t-j'}) \\ &= \lim_{n \rightarrow \infty} \sum_{j=0}^{n+h} \sum_{j'=0}^n \phi_1^j \phi_1^{j'} \text{cov}(Z_{t+h-j}, Z_{t-j'}) = \lim_{n \rightarrow \infty} \sum_{j=0}^{n+h} \sum_{j'=0}^n \phi_1^j \phi_1^{j'} \sigma^2 \delta_{j'+h, j} = \lim_{n \rightarrow \infty} \sum_{j'=0}^n \phi_1^{j'+|h|} \phi_1^{j'} \sigma^2 \\ &= \sigma^2 \phi_1^{|h|} \lim_{n \rightarrow \infty} \sum_{j'=0}^n (\phi_1^{j'})^2 = \sigma^2 \phi_1^{|h|} \lim_{n \rightarrow \infty} \phi_1^0 \frac{1 - \phi_1^{2n}}{1 - \phi_1^2} = \sigma^2 \phi_1^{|h|} \frac{1}{1 - \phi_1^2} \end{aligned} \quad (1.39)$$

gdzie skorzystaliśmy z faktu, że  $|\phi_1| < 1$ .

Z (1.38) i (1.39) wynika więc, że proces (1.36) jest (z dokładnością do momentów drugiego rzędu) procesem stacjonarnym:

a)  $EX_t = 0$  i nie zależy od  $t$  (1.40a)

b)  $\text{cov}(X_{t+h}, X_t) = \sigma^2 \phi_1^{|h|} \frac{1}{1 - \phi_1^2}$  nie zależy od  $t$  (1.40b)

Ponadto proces określony zgodnie z (1.36) jako  $X_t = \sum_{j=0}^{\infty} \phi_1^j Z_{t-j}$  jest jedynym rozwiązaniem równania różnicowego (1.20b), co wynika z procedury, która doprowadziła do jego otrzymania.

Gdyby  $|\phi_1| > 1$ , wtedy szereg po prawej stronie (1.36) nie jest zbieżny w  $L^2$ , tzn.

$\sum_{j=0}^{\infty} \phi_1^j Z_{t-j}$  nie zbiega się wtedy do zmiennej losowej  $X_t$  określonej zależnością (1.23) lub (1.21).

Jednak możemy przekształcić (1.21), tzn.  $X_t = Z_t + \phi_1 X_{t-1}$  do postaci:

$$X_{t-1} = -\phi_1^{-1} Z_t + \phi_1^{-1} X_t \quad (1.41)$$

co można zapisać następująco:

$$X_t = -\phi_1^{-1} Z_{t+1} + \phi_1^{-1} X_{t+1} \quad (1.41')$$

Rozwiążmy (1.41') rekurencyjnie wykorzystując  $X_{t+1} = -\phi_1^{-1} Z_{t+2} + \phi_1^{-1} X_{t+2}$  itd.

Zatem:

$$\begin{aligned} X_t &= -\phi_1^{-1} Z_{t+1} + \phi_1^{-1} (-\phi_1^{-1} Z_{t+2} + \phi_1^{-1} X_{t+2}) = \\ &= -\phi_1^{-1} Z_{t+1} - \phi_1^{-2} Z_{t+2} - \dots - \phi_1^{-k-1} Z_{t+k+1} + \phi_1^{-k-1} X_{t+k+1} \end{aligned} \quad (1.42)$$

zatem: 
$$X_t = -\sum_{j=1}^{k+1} \phi_1^{-j} Z_{t+j} + \phi_1^{-k-1} X_{t+k+1} \quad (1.42')$$

Ponieważ teraz  $|\phi_1| > 1$ , zatem dla stacjonarnego procesu  $\{X_t\}$  ( $EX_t = 0$ ) mamy, że  $\|X_t\|^2 = E(X_t^2) < \infty$  jest stałe i podobnie jak poprzednio w zależności (1.26) otrzymujemy:

$$\left\| X_t - \sum_{j=1}^{k+1} \phi_1^{-j} Z_{t+j} \right\|^2 = \left\| \phi_1^{-k-1} X_{t+k+1} \right\|^2 = \phi_1^{-2k-2} \|X_{t+k+1}\|^2 \xrightarrow{k \rightarrow \infty} 0 \quad (1.43)$$

oraz jak poprzednio (tzn. dla  $|\phi_1| < 1$  i (1.36)), tak i teraz dla  $|\phi_1| > 1$ , szereg

$$X_t = -\sum_{j=0}^{\infty} \phi_1^{-j} Z_{t+j} \quad (1.44)$$

jest jedynym stacjonarnym rozwiązaniem równania (1.20) przy  $|\phi_1| > 1$ .

Stacjonarne rozwiązanie (1.44) jest rozwiązaniem z wyprzedzeniem, tzn. zmienna  $X_t$  w (1.44) jest skorelowana ze zmiennymi  $\{Z_s, s > t\}$ .

Na ogół (w fizyce) jesteśmy zainteresowani rozwiązaniami z opóźnieniem. Zatem modelując stacjonarne szeregi czasowe typu AR(1), ograniczymy się na ogół do przypadku z  $|\phi_1| < 1$ , dla których  $X_t$  ma postać (1.36) i skorelowana ze zmiennymi  $\{Z_s, s \leq t\}$ , tzn. ze zmiennymi w chwili poprzednich. Procesy z opóźnieniami nazywamy również procesami autoregresji niezależnymi od przyszłości lub procesami przyczynowymi.

Można również pokazać, że:

### **Twierdzenie 1.2**

Jeśli  $\{X_t, t = 0, \pm 1, \pm 2, \dots\}$  jest stacjonarnym rozwiązaniem z wyprzedzeniem (tzn. nie przyczynowym) równań AR(1):

$$|\phi_1| > 1, \quad X_t = Z_t + \phi_1 X_{t-1}, \quad \{Z_t\} \sim \text{WN}(0, \sigma^2), \quad (1.45)$$

wtedy  $\{X_t\}$  spełnia również przyczynowe równania AR(1):

$$X_t = \phi_1^{-1} X_{t+1} + \tilde{Z}_{t+1}, \quad \{\tilde{Z}_t\} \sim \text{WN}(0, \tilde{\sigma}^2) \quad (1.46)$$

gdzie  $|\phi_1^{-1}| < 1$ , a  $\{\tilde{Z}_t\}$  odpowiednio dobranym białym szumem.

Zatem możemy pominąć w rozważaniach procesy AR(1) z  $|\phi_1| > 1$ .

**Uwaga:**

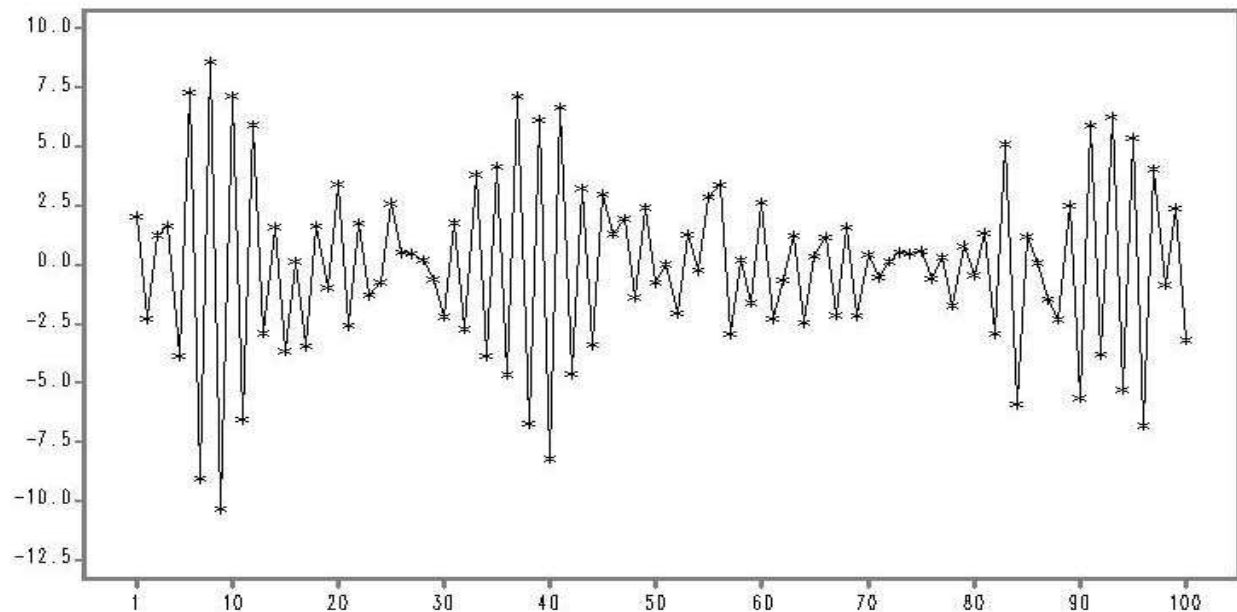
Można również pokazać, że nie ma stacjonarnego rozwiązania równań różnicowych

$$X_t = Z_t + \phi_1 X_{t-1}, \quad \{Z_t\} \sim WN(0, \sigma^2), \text{ dla których } |\phi_1| = 1.$$

Zatem nie ma pojęcia AR(1) z  $|\phi_1| = 1$ .

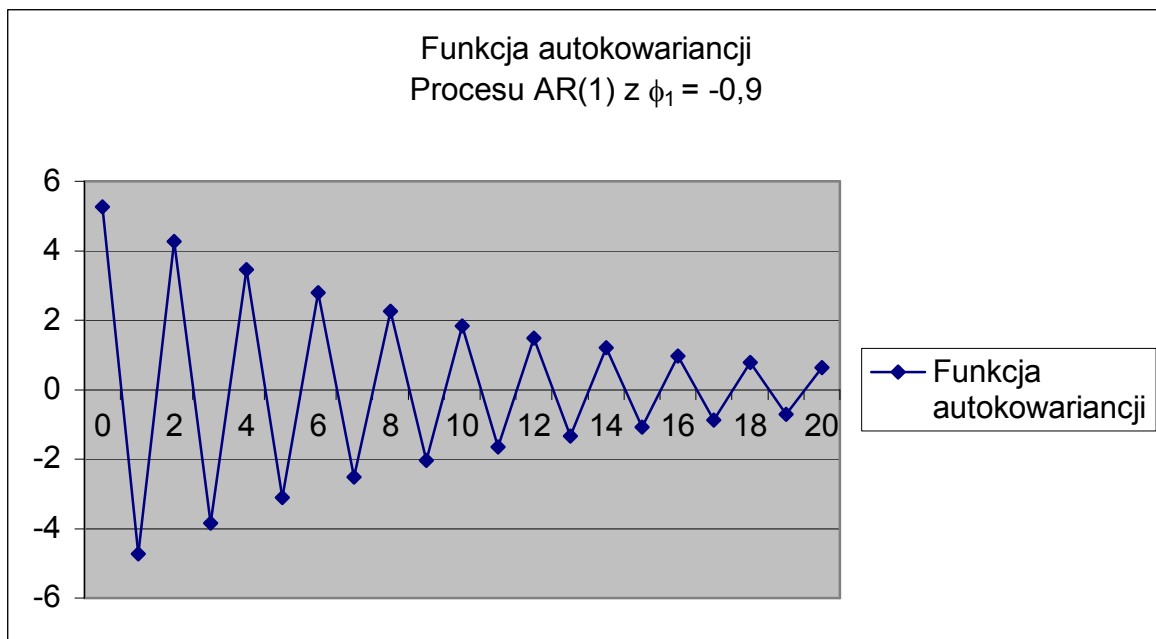
**Przykład 1.2.4.** Poniższy wykres przedstawia realizację procesu AR(1) z  $\phi_1 = -0.9$ , tzn.

$X_t = Z_t - 0.9X_{t-1}$ , gdzie  $Z_t \sim WN(0,1)$ . Wartości szumu  $\{Z_t, t = 1, 2, \dots, 100\}$  wygenerowano generatorem SAS'a.

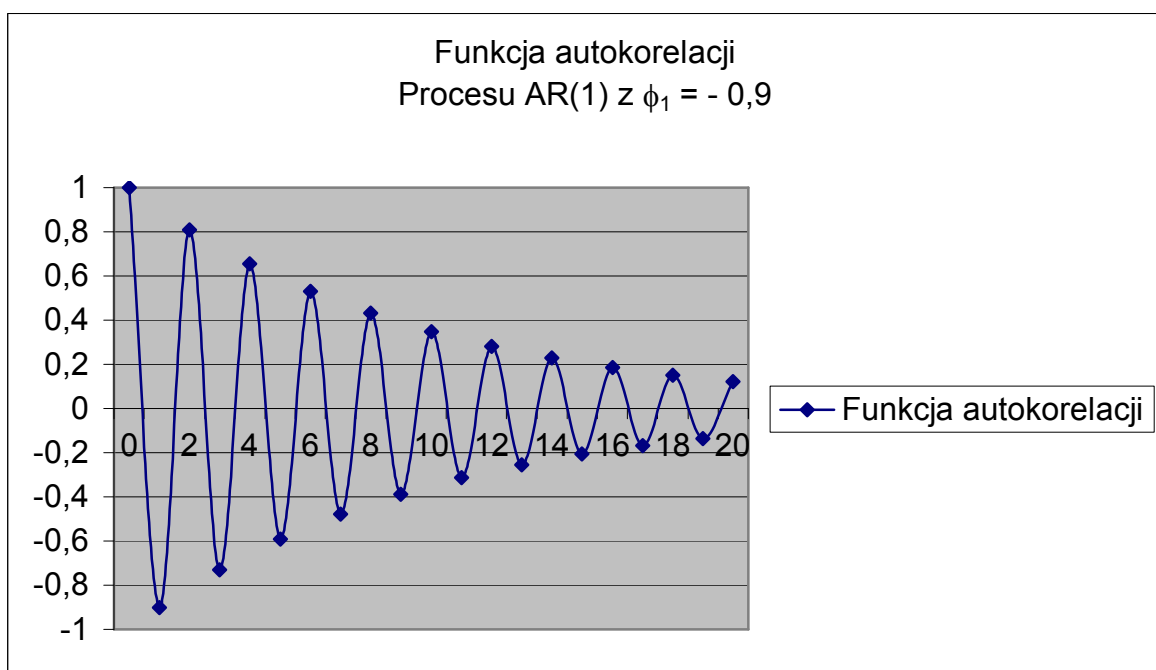


Wykres 3. 100 obserwacji procesu AR(1)  $X_t = Z_t - 0.9X_{t-1}$ , gdzie  $Z_t \sim WN(0,1)$

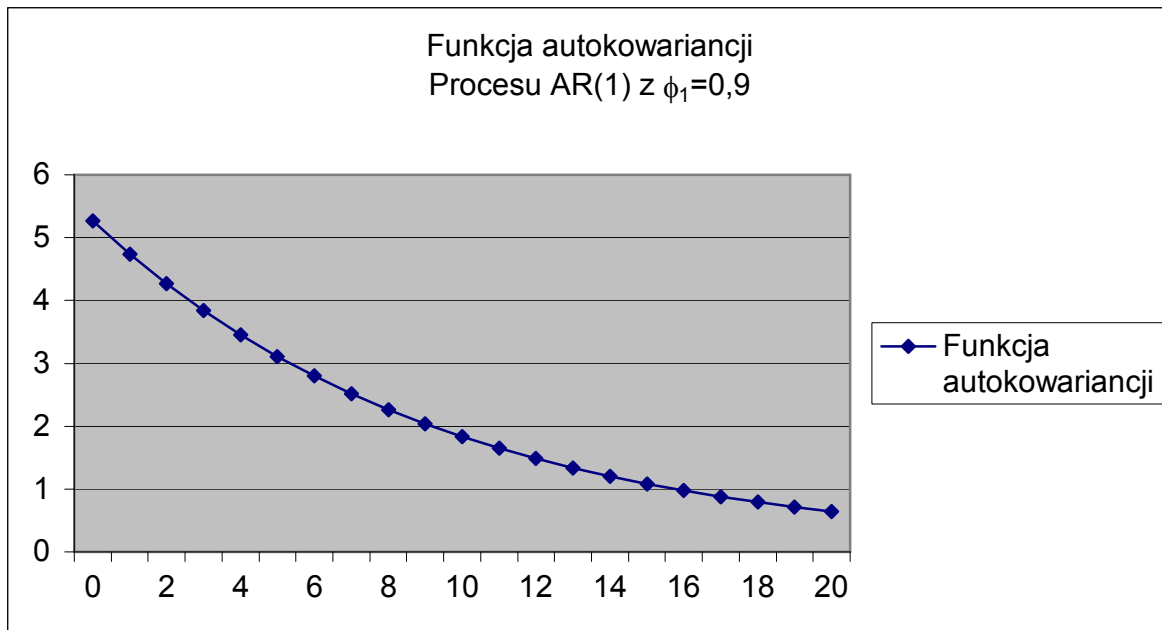
Przedstawiono również wykresy funkcji autokowariancji i autokorelacji wyznaczone z (1.39) przy wartościach parametru  $\phi_1 = -0.9$  i  $\phi_1 = 0.9$ . Wykresy te mogą służyć jako wykresy porównawcze dla procesów rzeczywistych przy ich wstępnej identyfikacji.



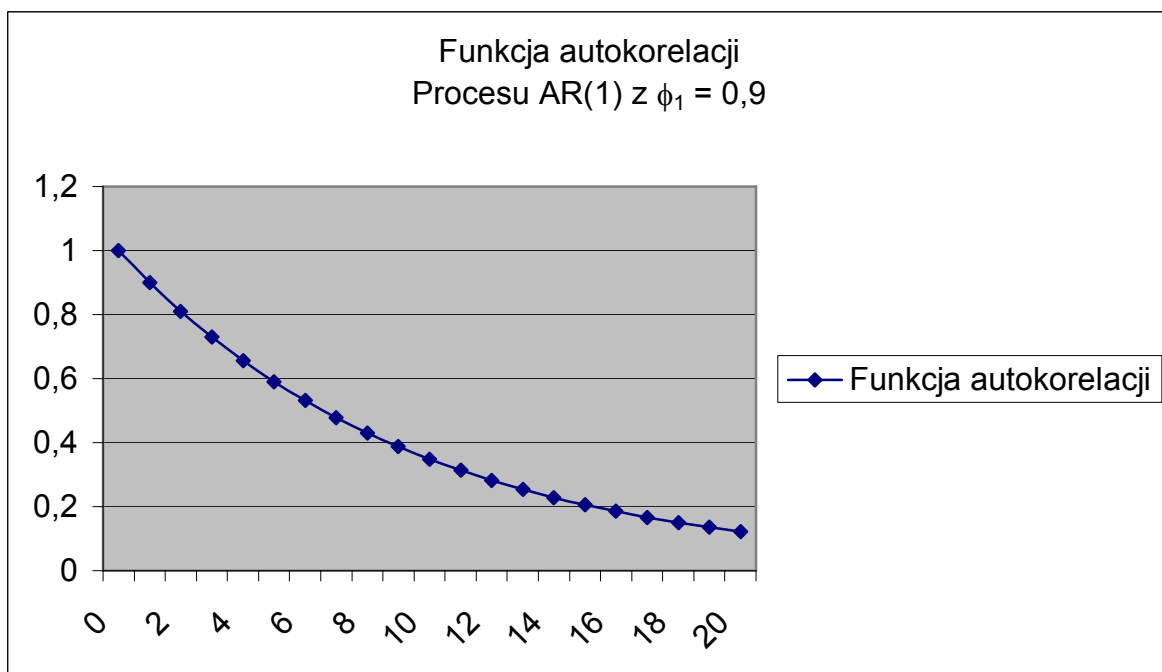
Wykres 4. Funkcja Autokowariancji procesu AR(1)  $X_t = Z_t - 0.9X_{t-1}$ , gdzie  $Z_t \sim WN(0,1)$



Wykres 5. Funkcja Autokorelacji procesu AR(1)  $X_t = Z_t - 0.9X_{t-1}$ , gdzie  $Z_t \sim WN(0,1)$



Wykres 6. Funkcja Autokowariancji procesu AR(1)  $X_t = Z_t + 0.9X_{t-1}$ , gdzie  $Z_t \sim WN(0,1)$



Wykres 7. Funkcja Autokorelacji procesu AR(1)  $X_t = Z_t + 0.9X_{t-1}$ , gdzie  $Z_t \sim WN(0,1)$

### **Definicja 1.2.3 Przyczynowego procesu ARMA**

Proces ARMA(p,q) określony równaniem

$$\phi(B)X_t = \theta(B)Z_t, \quad \{Z_t\} \sim \text{WN}(0, \sigma^2). \quad (1.47)$$

jest *procesem przyczynowym* (lub przyczynową funkcją białego szumu  $\{Z_t\}$ ), wtedy gdy istnieje ciąg stałych  $\{\psi_j\}$

$$\sum_{j=0}^{\infty} |\psi_j| < \infty \quad (1.48)$$

takich, że zachodzi:  $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \quad t = 0, \pm 1, \pm 2, \dots$  (1.49)

Zmienne  $X_t$  i  $Z_t$  są ze sobą skorelowane, zatem przyczynowość jest własnością nie tyle samego procesu  $\{X_t\}$ , ale relacji pomiędzy procesami  $\{X_t\}$  i  $\{Z_t\}$ .

### **Określenie 1.1**

Mówimy, że  $\{X_t\}$  jest procesem przyczynowym, jeśli otrzymujemy go z procesu  $\{Z_t\}$  przez zastosowanie liniowego filtra.

### **Definicja 1.2.4 Procesu odwracalnego**

Proces ARMA (p, q) zdefiniowany równaniem (1.47) jest nazywany procesem odwracalnym,

wtedy gdy istnieje ciąg stałych  $\{\pi_j\}$ , taki że  $\sum_{j=0}^{\infty} |\pi_j| < \infty$  (1.50)

oraz  $Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}, \quad t = 0, \pm 1, \pm 2, \dots$  (1.51)

Podobnie jak dla procesu przyczynowego, odwracalność jest własnością pary procesów  $\{X_t\}$  i  $\{Z_t\}$ . Warunek wystarczający i konieczny odwracalności procesu ARMA(p,q) jest podany dalej.

### 1.3 Stacjonarny proces AR(p)

Rozważmy proces  $\{X_t\}$ , który jest procesem autoregresji rzędu  $p$  o wartościach rzeczywistych, czyli proces AR(p):

$$X_t = \sum_{j=1}^p \phi_j X_{t-j} + Z_t, \text{ gdzie } \{Z_t\} \sim \text{WN}(0, \sigma^2), \text{ gdzie } t = 0, \pm 1, \pm 2, \dots \quad (1.52)$$

Wprowadzając chwilowo stałą  $\phi_0$  (równą 0 dla procesu AR(p) ze średnią równą 0, ale porównaj (1.9)), proces AR(p) możemy zapisać, wykorzystując operator cofnięcia  $B$  (1.13), w następujący sposób:

$$\phi(B)X_t = \phi_0 + Z_t, \quad \{Z_t\} \sim \text{WN}(0, \sigma^2), \quad (1.53)$$

gdzie zgodnie z (1.11):

$$\phi(B) = 1 - \phi_1 B^1 - \phi_2 B^2 - \dots - \phi_p B^p \quad (1.54)$$

jest tzw. wielomianem charakterystycznym omawianego procesu, natomiast

$$\phi(B) = 0 \quad (1.55)$$

jest jego równaniem charakterystycznym.

Korzystając bezpośrednio z (1.53-54), nietrudno pokazać, że dla procesów AR(p) mamy:

$$E(X_t) = \frac{\phi_0}{1 - \phi_1 - \dots - \phi_p}, \quad (1.56)$$

jak również, że

$$\text{Var}(X_t) = \text{Var}\left(\sum_{i=0}^{\infty} \psi_i Z_{t-i}\right) = \sigma^2 \sum_{i=0}^{\infty} \psi_i^2 = \gamma_0, \quad (1.57a)$$

$$\text{Cov}(X_t, X_{t-k}) = E\left[\left(\sum_{i=0}^{\infty} \psi_i Z_{t-i}\right)\left(\sum_{i=0}^{\infty} \psi_i Z_{t-i-k}\right)\right] = \sigma^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+k} = \gamma_k. \quad (1.57b)$$

Szereg czasowy będący realizacją przedstawionego procesu charakteryzuje się tym, że jego bieżąca wartość jest sumą skończonej kombinacji liniowej poprzednich wartości tego szeregu oraz zakłócenia losowego. Nazwa „proces autoregresji” pochodzi stąd, że wzór (1.52) można traktować jako liniowe równanie regresji wielorakiej zmiennej  $X_t$  względem opóźnionych w czasie wartości tej zmiennej.

Można pokazać, że:

Aby zagwarantować stacjonarność procesu  $\{X_t\}$ , współczynniki  $\phi_j$  muszą spełniać pewne ograniczenia, które można określić badając pierwiastki tzw. równania charakterystycznego dla wielomianu charakterystycznego  $\phi(z)$  :

$$\phi(z) = 1 - \phi_1 z^1 - \phi_2 z^2 - \dots - \phi_p z^p = 0 \quad (1.58)$$

Okazuje się, że zachodzi twierdzenie:

### **Twierdzenie 1.3**

Jeśli wszystkie pierwiastki wielomianu charakterystycznego  $\phi(z)$  leżą poza promieniem jednostkowym (tzn. spełniają zależność  $|z| > 1$ ), wtedy proces AR(p) zadany równaniem (1.52) ma jedno stacjonarne (słabe) rozwiązanie:

$$X_t = \sum_{j=1}^{\infty} \psi_j Z_{t-j}, \quad t = 0, \pm 1, \pm 2, \dots, \quad (1.59)$$

które zbiega się średnio – kwadratowo.

Rozwiązanie (1.59) implikuje automatycznie warunek:

$$E(X_t) = 0 \quad \text{dla wszystkich } t. \quad (1.60)$$

### **Uwaga.**

W rozwiązaniu (1.59) pojawiają się zmienne  $Z_t$  jedynie z przeszłości i teraźniejszości, zatem proces  $X_t$  jest przyczynowy. Widać, że (1.59) dla AR(p) jest uogólnieniem procesu AR(1) określonego równaniem (1.36) tzn.:  $\psi_j = \phi_1^j$ . (1.61)

Fakt, że jedynie przeszłe i teraźniejsze zmienne  $Z_t$  wchodzą w rozwiązanie (1.61) podkreślamy mówiąc, że nieskończona ruchoma średnia (1.59) jest jedno – stronna a proces przyczynowy. Zatem proces  $\{X_t\}$  jest liniowy, będąc jednostronnie ruchomą średnią białego szumu  $\{Z_t\}$ .



### Uwaga.

W praktyce wykorzystując do modelowania szeregów czasowych proces (1.52) napotykamy na problem tzw. identyfikacji. Polega on na tym, że znając szereg czasowy prognozowanej zmiennej musimy określić, czy jest on realizacją procesu autoregresji.

Jeżeli jest, to pozostaje pytanie jakiego rzędu  $p$  jest to autoregresja. Identyfikacja najczęściej odbywa się przez porównanie własności teoretycznych funkcji autokorelacji i autokorelacji cząstkowej procesu z zachowaniem się współczynników autokorelacji i autokorelacji cząstkowej, oszacowanych na podstawie szeregu czasowego.

### Uwaga.

Proces (1.52) można zapisać następująco:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + Z_t \quad (1.52')$$

Mnożąc powyższe równanie obustronnie przez  $X_{t-h}$ , gdzie  $h > 0$ , i biorąc wartości oczekiwane (dokładny rachunek jest poniżej (1.121)-(1.126)), a następnie dzieląc przez wartość kowariancji  $\gamma(0) = Cov(X_t, X_t)$ , otrzymujemy równanie Yula – Walker'a:

$$\rho(h) = \phi_1 \rho(h-1) + \phi_2 \rho(h-2) + \dots + \phi_p \rho(h-p), \quad h > 0 \quad (1.62)$$

gdzie  $\rho(i)$  jest funkcją autokorelacji:

$$\rho(i) = \frac{\gamma(i)}{\gamma(0)} \quad (1.63)$$

Dla  $h = 1, 2, \dots, p$  w (1.62) otrzymaliśmy  $p$  liniowych równań nazywanych równaniami

Yule – Walker'a, które pozwalają określić  $p$  parametrów  $\phi_1, \phi_2, \dots, \phi_p$  procesu AR(p).

Jednak aby mogło to miało miejsce należy wpierw oszacować autokowariancję  $\rho(1), \rho(2), \dots, \rho(p)$  na podstawie danych z próby (co pokażemy później).

### Uwaga.

Można również pokazać [1], że ogólne rozwiązanie układu  $p$  równań Yula – Walker'a (1.62) ma postać:

$$\rho(h) = \sum_{i=1}^k \sum_{j=0}^{r_i-1} \beta_{ij} h^j \xi_i^{-h}, \quad h \geq \max(p, q+1) - p \quad (1.64)$$

gdzie  $\xi_i$  są pierwiastkami równania charakterystycznego  $\phi(z) = 0$ , (1.58),  $r_i$  są krotnościami tych pierwiastków ( $\sum_{i=1}^k r_i = p$ ), a  $\beta_{ij}$  są stałymi (jest ich  $p$ ),  $0 \leq j < \max(p, q+1) - p$  (porównaj (1.126)).

Zgodnie z Twierdzeniem 1.3 wiemy, że (1.52) ma stacjonarne rozwiązanie gdy pierwiastki „z” równania charakterystycznego spełniają zależność  $|z| > 1$ . Wtedy  $|\xi_i^{-h}| < 1$  co oznacza, że  $\rho(h)$  maleje eksponentalnie wraz ze wzrostem  $|h|$ .

**Uwaga.**

Fakt, że szereg (1.64) jest w przypadku  $|\xi_i^{-h}| < 1$  bezwzględnie zbieżny, pociąga za sobą istnienie gęstości widmowej dla procesu  $\{X_t\}$  [1].

Poniższe dwa twierdzenia są kluczowe dla zrozumienia stacjonarności procesów przekształconych, będących kombinacją liniową zmiennych  $X_t$ .

**Twierdzenie 1.4**

Jeśli  $\{X_t\}$  jest dowolnym ciągiem zmiennych losowych, takich że:

$$\sup_t E|X_t| < \infty \tag{1.65}$$

i jeśli 
$$\sum_{j=-\infty}^{\infty} |\psi_j| < \infty \tag{1.66}$$

wtedy szereg 
$$\psi(B)X_t = \sum_{j=-\infty}^{\infty} \psi_j B^j X_j = \sum_{j=-\infty}^{\infty} \psi_j X_{t-j} \tag{1.67}$$

zbiega się bezwzględnie z prawdopodobieństwem równym jeden.

Jeśli dodatkowo 
$$\sup_t E|X_t|^2 < \infty \tag{1.68}$$

wtedy szereg (1.67) zbiega się średnio – kwadratowo do tej samej granicy.

### **Twierdzenie 1.5**

Jeśli  $\{X_t\}$  jest stacjonarnym procesem z funkcją autokowariancji  $\gamma(\cdot)$  i jeśli

$$\sum_{j=-\infty}^{\infty} |\psi_j| < \infty, \quad (1.69)$$

wtedy dla każdego  $t \in Z$ , szereg (1.67)

$$\psi(B)X_t = \sum_{j=-\infty}^{\infty} \psi_j B^j X_t = \sum_{j=-\infty}^{\infty} \psi_j X_{t-j} \quad (1.67')$$

zbiega się zarówno bezwzględnie z prawdopodobieństwem równym jeden, jak i średnio – kwadratowo do tej samej granicy.

Jeśli  $Y_t = \psi(B)X_t$  (1.70)

wtedy proces  $\{Y_t\}$  jest stacjonarny z funkcją autokowariancji:

$$\gamma_Y(h) = \sum_{j,k=-\infty}^{\infty} \psi_j \psi_k \gamma(h-j+k). \quad (1.71)$$

### **Dowód:**

Zbieżność  $\psi(B)X_t$  wynika z Twierdzenia 1.4.

Natomiast jeśli  $\sigma^2|X_t| = E|X_t|^2 - (E|X_t|)^2 \geq 0$

zatem  $E|X_t| \leq (E|X_t|^2)^{1/2} = c$  (1.72)

Jeśli  $\{X_t\}$  jest procesem stacjonarnym, wtedy  $c$  jest skończone i niezależne od  $t$ .

Aby sprawdzić stacjonarność procesu  $\{Y_t\}$ , zauważmy (Twierdzenie 1.4), że z średnio – kwadratowej zbieżności szeregu (1.67) i ciągłości iloczynu wewnętrznego, wynika że:

1.  $EY_t = E\left(\sum_{j=-\infty}^{\infty} \psi_j X_{t-j}\right) = E \lim_{n \rightarrow \infty} \sum_{j=-n}^n \psi_j X_{t-j} = \lim_{n \rightarrow \infty} \sum_{j=-n}^n \psi_j EX_{t-j} = \left(\sum_{j=-\infty}^{\infty} \psi_j\right) EX_t$  (1.73)

gdzie w równości  $EX_{t-j} = EX_t$  skorzystano ze stacjonarności procesu  $\{X_t\}$ .

$$\begin{aligned}
2. \quad E(Y_{t+h}Y_t) &= E\left[\left(\sum_{j=-\infty}^{\infty}\psi_jX_{t+h-j}\right)\left(\sum_{k=-\infty}^{\infty}\psi_kX_{t-k}\right)\right] = \lim_{n\rightarrow\infty} E\left[\left(\sum_{j=-n}^n\psi_jX_{t+h-j}\right)\left(\sum_{k=-n}^n\psi_kX_{t-k}\right)\right] = \\
&= \lim_{n\rightarrow\infty} \sum_{j=-n}^n \sum_{k=-n}^n \psi_j\psi_k E(X_{t+h-j}\cdot X_{t-k}) = \lim_{n\rightarrow\infty} \sum_{j=-n}^n \sum_{k=-n}^n \psi_j\psi_k (\gamma(h-j+k) + (EX_t)^2) = \\
&= \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j\psi_k (\gamma(h-j+k) + (EX_t)^2) \tag{1.74}
\end{aligned}$$

gdzie skorzystano z faktu że:

$$\begin{aligned}
&\gamma[(t+h-j)-(t-k)] = \gamma[h-j+k] = \\
&= E(X_{t+h-j}\cdot X_{t-k}) - E(X_{t+h-j})\cdot E(X_{t-k}) = E(X_{t+h-j}\cdot X_{t-k}) - E(X_t)^2
\end{aligned}$$

Ponieważ proces  $\{X_t\}$  jest stacjonarny, co oznacza, że  $E(X_t)$  i  $\gamma[h-j+k]$  są skończone i nie zależą od czasu, zatem zgodnie z (1.73) mamy, że  $E(Y_t)$  jest skończone i niezależne od czasu, a zgodnie z (1.74) skończone i niezależne od czasu jest  $E(Y_{t+h}Y_t)$ . Proces  $\{Y_t\}$  jest więc również stacjonarny.

Korzystając z (1.73) i (1.74) funkcja autokowariancji  $\gamma_Y(h)$  przyjmuje postać:

$$\begin{aligned}
\gamma_Y(h) &= E(Y_{t+h}\cdot Y_t) - E(Y_{t+h})\cdot E(Y_t) = \\
&= \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j\psi_k (\gamma(h-j+k) + (EX_t)^2) - \left(\sum_{j=-\infty}^{\infty} \psi_j\right) E(X_{t+h}) \cdot \left(\sum_{k=-\infty}^{\infty} \psi_k\right) E(X_t) = \\
&= \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j\psi_k \gamma(h-j+k)
\end{aligned}$$

$$\text{zatem} \quad \gamma_Y(h) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j\psi_k \gamma(h-j+k) \tag{1.75}$$

c.n.d.

### Uwaga.

Podsumowanie: Z powyższego twierdzenia wynika, że jeżeli proces  $\{X_t\}$  jest stacjonarny to proces  $\psi(B)X_t$  jest również stacjonarny.

Oznacza to, że operator  $\psi(B) = \sum_{j=-\infty}^{\infty} \psi_j B^j$  z  $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$  zastosowany do procesu stacjonarnego posiada algebraiczne własności szeregów potęgowych, tzn. jeśli np.

$$\alpha(z) = \sum_{j=-\infty}^{\infty} \alpha_j z^j \text{ i } \sum_{j=-\infty}^{\infty} |\alpha_j| < \infty \quad (1.76a)$$

oraz 
$$\beta(z) = \sum_{j=-\infty}^{\infty} \beta_j z^j \text{ i } \sum_{j=-\infty}^{\infty} |\beta_j| < \infty \quad (1.76b)$$

i 
$$\alpha(z) \cdot \beta(z) = \psi(z) \text{ dla } |z| \leq 1 \quad (1.77)$$

wtedy proces 
$$\psi(z) X_t = \alpha(z) \cdot \beta(z) X_t \quad (1.78)$$

jest również stacjonarny i zachodzą własności:

$$\sum_{j=-\infty}^{\infty} |\psi_j| < \infty \quad (1.79a)$$

oraz 
$$\alpha(B) \cdot \beta(B) X_t = \beta(B) \cdot \alpha(B) X_t = \psi(B) X_t \quad (1.79b)$$

## **1.4 Pierwiastki wielomianów $\phi(z)$ i $\theta(z)$ dla procesów ARMA( $p, q$ ) a przyczynowość, odwracalność i stacjonarność procesów ARMA**

### Twierdzenie 1.6

Niech  $\{X_t\}$  będzie procesem ARMA ( $p, q$ ), dla którego  $\phi(\cdot)$  i  $\theta(\cdot)$  nie mają wspólnych zer.

Wtedy  $\{X_t\}$  jest procesem przyczynowym (Definicja 1.2) wtedy i tylko wtedy, gdy

$\phi(z) \neq 0$  dla każdego  $z \in C$ , takiego że  $|z| \leq 1$ .

Współczynniki  $\{\psi_j\}$  w (1.49) w definicji procesu przyczynowego

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \quad t = 0, \pm 1, \pm 2, \dots \quad (1.49')$$

są wtedy określone zależnością:

$$\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \theta(z) / \phi(z), \quad |z| \leq 1 \quad (1.80)$$

### **Twierdzenie 1.7**

Warunkiem wystarczającym dla stacjonarności (słabej) procesu  $\{X_t\}$  jest aby pierwiastki wielomianu  $\phi(z)$  leżały poza okręgiem jednostkowym, tzn.  $|z| > 1$ . Jeśli dodatkowo pierwiastki wielomianu  $\theta(z)$  leżą poza jednostkowym promieniem, to proces  $\{X_t\}$  jest odwracalny (Definicja 1.3), co oznacza, że istnieją takie  $\{\pi_j\}_{j=0}^{\infty}$ , że:

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}, \quad t = 0, \pm 1, \pm 2, \dots \quad (1.81)$$

### **Twierdzenie 1.8**

Niech  $\{X_t\}$  będzie procesem ARMA  $(p, q)$ , dla którego wielomiany  $\phi(z)$  i  $\theta(z)$  nie mają wspólnych zer. Wtedy proces  $\{X_t\}$  jest odwracalny (Definicja 1.3) wtedy i tylko wtedy, gdy  $\theta(z) \neq 0$  dla wszystkich  $z \in C$ , dla których  $|z| \leq 1$ , tzn. zera istnieją jedynie poza okręgiem jednostkowym.

Współczynniki  $\{\pi_j\}$  w (1.51) i (1.81) można wtedy wyznaczyć z relacji:

$$\pi(z) = \sum_{j=0}^{\infty} \pi_j z^j = \frac{\phi(z)}{\theta(z)}, \quad |z| \leq 1 \quad (1.82)$$

Zależność (1.81) można wtedy zapisać następująco:

$$Z_t = \frac{\phi(B)}{\theta(B)} X_t \quad (1.83)$$

### **Podsumowanie.**

W świetle powyższych rozważań podkreślimy, że w dalszej części pracy skoncentrujemy się wyłącznie na przyczynowych i odwracalnych procesach ARMA.

Jak widzieliśmy, okazuje się, że wprowadzone powyżej warunki, narzucane na wielomiany  $\phi(z)$  i  $\theta(z)$  są istotne z punktu widzenia przyczynowości i odwracalności procesu ARMA.

Okazuje się, że z wyjściem poza warunki, aby wielomiany  $\phi(z)$  i  $\theta(z)$  nie miały wspólnych zer i aby  $\phi(z) \neq 0$  oraz  $\theta(z) \neq 0$  dla wszystkich  $z \in C$ , dla których  $|z| \leq 1$ , pojawiają się problemy nie tylko z brakiem przyczynowości lub odwracalności procesu, ale i z brakiem stacjonarności procesu [1].

I tak, np. chociaż  $\phi(z)$  i  $\theta(z)$  nie mają wspólnych zer, ale  $\phi(z) = 0$  dla pewnego  $z \in C$ , dla którego  $|z| = 1$ , wtedy nie ma stacjonarnego rozwiązania równania ARMA  $\phi(B)X_t = \theta(B)Z_t$ .

Z drugiej, strony jeśli  $\phi(z) \neq 0$  dla wszystkich  $z \in C$  dla których  $|z| = 1$ , wtedy z analizy zespolonej wynika, że istnieje promień  $r > 1$ , w którym szereg Laurent'a (1.80):

$$\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \theta(z) / \phi(z), \quad r^{-1} < |z| < r \quad (1.80')$$

jest bezwzględnie zbieżny. Zbieżność powyższego szeregu odgrywa kluczową rolę w dowodzie poniższego twierdzenia o istnieniu stacjonarnego procesu będącego jednoznacznym rozwiązaniem równania ARMA.

### **Twierdzenie 1.9**

Jeśli  $\phi(z) \neq 0$  dla każdego  $z \in C$ , takiego że  $|z| = 1$ , wtedy równanie ARMA  $\phi(B)X_t = \theta(B)Z_t$  ma jednoznaczne stacjonarne rozwiązanie:

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}, \quad t = 0, \pm 1, \pm 2, \dots, \quad (1.84)$$

gdzie współczynniki  $\psi_j$  są określone z równania (1.80).

## **1.5 Procesy średniej ruchomej nieskończonego rzędu $MA(\infty)$**

### **Definicja 1.5 Procesu $MA(\infty)$**

Jeśli  $\{Z_t\} \sim WN(0, \sigma^2)$  wtedy mówimy, że  $\{X_t\}$  jest procesem ze średnią ruchomą  $MA(\infty)$  (nieskończonego rzędu) procesu  $\{Z_t\}$ , jeśli tylko istnieje ciąg  $\{\psi_j\}$ , taki że:

$$\sum_{j=0}^{\infty} |\psi_j| < \infty \quad \text{i} \quad X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \quad t = 0, \pm 1, \pm 2, \dots \quad (1.85)$$

### **Przykład 1.5.1**

Proces  $MA(q)$ :

$$X_t = \theta(B)Z_t, \quad (1.86a)$$

jest średnią ruchomą z  $\{Z_t\}$ , gdzie  $\psi_j = \theta_j$ ,  $j = 0, 1, 2, \dots, q$ , oraz  $\psi_j = 0$ ,  $j > q$ . (1.86b)

### **Przykład 1.5.2**

Proces AR(1) z  $|\phi_1| < 1$  zdefiniowany przez (1.17) jest średnią ruchomą z  $\{Z_t\}$

$$\text{gdzie } \psi_j = \phi_1^j, \quad j = 0, 1, 2, \dots \quad (1.87)$$

### **Przykład 1.5.3**

Zgodnie z Twierdzeniem 1.6, przyczynowy proces ARMA( $p, q$ )  $\phi(B)X_t = \theta(B)Z_t$ , jest średnią ruchomą (1.49) z  $\{Z_t\}$  ze współczynnikami spełniającymi warunek

$$\sum_{j=0}^{\infty} \psi_j z^j = \theta(z) / \phi(z), \quad \text{gdzie } |z| \leq 1. \quad (1.80')$$

### **Uwaga.**

W Definicji 1.4 procesu  $MA(\infty)$  z  $\{Z_t\}$  wymaga się, aby  $X_t$  wyrażało się poprzez  $Z_s$ , gdzie mamy ograniczenie do  $s \leq t$ . Dlatego w przykładzie 1.5.3 założyliśmy, że proces ARMA( $p, q$ ) jest przyczynowy.

Jednakże nawet dla nie przyczynowego procesu ARMA, można znaleźć (ciąg) białego szumu  $\{Z_t^*\}$ , taki że  $\{X_t\}$  spełnia przyczynowe i odwracalne równanie ARMA i jest zatem średnią ruchomą z  $\{Z_t^*\}$  [1].

### **Twierdzenie 1.10**

Niech  $\{X_t\}$  będzie procesem ARMA( $p, q$ ), spełniającym równanie:

$$\phi(B)X_t = \theta(B)Z_t, \quad \{Z_t\} \sim \text{WN}(0, \sigma^2) \quad (1.88)$$

gdzie  $\phi(z) \neq 0$  i  $\theta(z) \neq 0$  dla wszystkich  $z \in C$  takich, że  $|z| = 1$ .

Istnieją wtedy wielomiany  $\tilde{\phi}(z)$  i  $\tilde{\theta}(z)$ , niezerowe dla  $|z| \leq 1$  odpowiednio stopnia  $p$  i  $q$  oraz proces białego szumu  $\{Z_t^*\}$ , taki że proces  $\{X_t\}$  spełnia przyczynowe i odwracalne równania:

$$\tilde{\phi}(B)X_t = \tilde{\theta}(B)Z_t^* \quad (1.89)$$

Dowód [1]. Można również pokazać, że:

$$\tilde{\phi}(z) = \phi(z) \prod_{r < j \leq p} \frac{1 - a_j z}{1 - a_j^{-1} z}, \quad \tilde{\theta}(z) = \theta(z) \prod_{s < j \leq q} \frac{1 - b_j z}{1 - b_j^{-1} z}, \quad (1.90)$$

gdzie  $a_{r+1}, \dots, a_p$  i  $b_{s+1}, \dots, b_q$  są zerami wielomianów  $\phi(z)$  i  $\theta(z)$ , które leżą wewnątrz promienia jednostkowego.



### **Podsumowanie.**

Należy zwrócić uwagę, że duża klasa stacjonarnych procesów ma reprezentację  $MA(\infty)$ .

### **Twierdzenie 1.11**

Jeśli  $\{X_t\}$  jest stacjonarnym procesem z funkcją autokowariancji  $\gamma(\cdot)$ , taką że

$$\gamma(h) = 0 \text{ dla } |h| > q \text{ i } \gamma(q) \neq 0, \quad (1.91)$$

wtedy  $\{X_t\}$  jest procesem  $MA(q)$ , tzn., że istnieje proces białego szumu  $\{Z_t\}$ , taki że:

$$X_t = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q} \quad (1.92)$$

Dowód powyższego, ważnego z punktu widzenia identyfikacji procesów twierdzenia, znajduje się w [1] na str. 90. Drugie i ogólniejsze twierdzenie, istotne z tego samego powodu znajduje się poniżej.

### **Twierdzenie 1.12**

Jeśli  $\{X_t\}$  ma tę samą funkcję autokowariancji co pewien proces  $ARMA(p, q)$   $\{Y_t\}$ , wtedy  $\{X_t\}$  jest również procesem  $ARMA(p, q)$ . Inaczej mówiąc istnieje biały szum  $\{Z_t\}$  oraz współczynniki  $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$  takie, że :

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q} \quad (1.93)$$

**Dowód** dla przypadku gdy  $\{X_t\}$  i  $\{Y_t\}$  są dwoma procesami ze średnią zero, opiera się o poniższy schemat [1]:

Jeśli  $\phi_1, \dots, \phi_p$  są współczynnikami AR dla procesu  $\{Y_t\}$ , wtedy można pokazać, że proces  $\{W_t\}$ , gdzie

$$W_t = X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} \quad (1.94)$$

ma funkcję autokowariancji równą zero, dla lag  $|h| > q$ .

W tej sytuacji, zgodnie z Twierdzeniem 1.11 proces  $\{W_t\}$  jest procesem  $MA(q)$ , tzn.:

$$W_t = Z_t - \theta_1 Z_{t-1} - \dots - \theta_q Z_{t-q}. \quad (1.95)$$

Korzystając z powyższego rozumowania mamy z (1.94) i (1.95):

$$W_t = X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t - \theta_1 Z_{t-1} - \dots - \theta_q Z_{t-q}, \quad (1.96)$$

co oznacza, że  $\{X_t\}$  jest procesem ARMA( $p, q$ ).

c.n.d

Poniżej podane zostanie twierdzenie istotne dla wyznaczania funkcji autokowariancji dla procesu, który może być zapisany w postaci średniej ruchomej  $MA(\infty)$  z  $\{Z_t\}$ .

Z Twierdzenia 1.5 i zależności (1.71) wynika, że:

### **Twierdzenie 1.13**

Proces  $MA(\infty)$  (Definicja 1.4):

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \quad t = 0, \pm 1, \pm 2, \dots, \quad \text{gdzie} \quad \sum_{j=0}^{\infty} |\psi_j| < \infty, \quad (1.84')$$

jest stacjonarny ze średnią równą zero i funkcją autokowariancji:

$$\gamma(k) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|k|}. \quad (1.97)$$

### **Dowód:**

Z zależności (1.71) w Twierdzeniu 1.5 mamy, że jeśli  $\{Z_t\}$  jest stacjonarnym procesem z

funkcją autokowariancji  $\gamma(\cdot)$  i jeśli  $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ , wtedy, dla każdego  $t \in Z$ , proces  $\{X_t\}$

(porównaj (1.67)):

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j} \quad (1.98)$$

jest stacjonarny z funkcją autokowariancji:

$$\gamma_X(h) = \sum_{j,k=-\infty}^{\infty} \psi_j \psi_k \gamma(h - j + k). \quad (1.71')$$

Natomiast z postaci szeregu (1.84):

$$X_t = \sum_{j=1}^{\infty} \psi_j Z_{t-j}, \quad t = 0, \pm 1, \pm 2, \dots \quad (1.84')$$

wynika, że:  $\psi_j = 0$  dla  $j = -1, -2, \dots$ .

(1.99)

Ponieważ proces  $\{Z_t\}$  występujący po prawej stronie (1.84') jest białym szumem, zatem zgodnie z (1.5) :

$$\gamma(h-j+k) = \begin{cases} \sigma^2 & \text{dla } h-j+k=0 \\ 0 & \text{dla } h-j+k \neq 0 \end{cases} \quad (1.100)$$

zatem:  $\gamma(h-j+k) = \sigma^2 \delta_{k, j-h}$  dla  $\{Z_t\} \sim \text{WN}(0, \sigma^2)$  (1.101)

i zgodnie z (1.71') otrzymujemy, że:

$$\gamma_X(h) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k \gamma(h-j+k) = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \psi_j \psi_k \sigma^2 \delta_{k, j-h} = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j-h}, \quad (1.102)$$

co dla dowolnego  $h \in Z$  można ostatecznie zapisać jako:

$$\gamma_X(h) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|h|}.$$

Średnia procesu (1.84') jest oczywiście równa zero, natomiast jego stacjonarność wynika wprost z Twierdzenia 1.5.

c.n.d.

**Uwaga !**

Zauważmy, że powyższe Twierdzenie 1.13 razem z Przykładem 1.5.3 (w którym stwierdzono, że przyczynowy proces  $\text{ARMA}(p, q)$  jest procesem  $\text{MA}(\infty)$ ), całkowicie określa funkcję autokowariancji dowolnego przyczynowego procesu  $\text{ARMA}(p, q)$ .

**Uwaga.**

Pojęcie procesu  $\text{AR}(p)$  (wprowadzone wcześniej) również można rozszerzyć do przypadku gdy  $p$  dąży do nieskończoności.

W szczególności zauważmy, że z Twierdzenia 1.8 i zależności (1.81) i (1.82) wynika, że każdy odwracalny proces  $\text{ARMA}(p, q)$  spełnia równanie:

$$X_t + \sum_{j=1}^{\infty} \psi_j X_{t-j} = Z_t, \quad t = 0, \pm 1, \pm 2, \dots, \quad (1.103)$$

k który ma tę samą postać co proces  $\text{AR}(p)$  (1.52) z  $p \rightarrow \infty$ . Oznacza to, że przyczynowy proces  $\text{ARMA}(p, q)$  jest procesem  $\text{AR}(\infty)$ .

Jak już wspominaliśmy we wstępie wartości parametrów  $p$  i  $q$  procesu, można wywnioskować na podstawie analizy zachowań dwóch funkcji: funkcji autokorelacji procesu ACF (autocorrelation function) i funkcji autokorelacji cząstkowej PACF (Partial Autocorrelation Function). Dlatego też obecnie opiszemy te dwie funkcje oraz podamy twierdzenia oraz metody ich wyznaczania (z przykładami). Estymacją Funkcji ACF i PACF zajmiemy się w rozdziale 2. Metoda Boxa - Jenkinsa [5], [6] opiera się na obserwacji kształtów funkcji ACF i PACF, które zależą od parametrów  $p$  i  $q$  (przykłady omówimy później).

## 1.6 Wyznaczanie funkcji autokowariancji dla procesu ARMA( $p, q$ ).

*Jest to zasadniczy punkt na drodze poprawnego wyboru procesu ARMA. Jest tak, bowiem dobra znajomość kształtu teoretycznych funkcji autokorelacji i częściowej autokorelacji, i uważne porównanie z nimi kształtów empirycznych autokorelogramów, pozwala na w miarę precyzyjne, chociaż ciągle wstępne określenie właściwego rzędu procesu ARMA( $p, q$ ), leżącego być może u podstaw obserwacji.*

Obecnie podamy 2 metody wyznaczania funkcji autokowariancji.

### Metoda pierwsza

Jak wiemy z (1.97) (porównaj Twierdzenie 1.13 i Przykład 1.5.3, z których wynika, że przyczynowy proces ARMA( $p, q$ ) jest procesem MA( $\infty$ )), funkcja autokowariancji  $\gamma$  przyczynowego procesu ARMA( $p, q$ ) dana jest wzorem:

$$\gamma(h) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|h|} \quad (1.97')$$

gdzie 
$$\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \theta(z) / \phi(z) \quad \text{dla } |z| \leq 1 \quad (1.80')$$

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p \quad (1.106a)$$

$$\theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \dots + \theta_q z^q \quad (1.106b)$$

Definiując:

$$\theta_0 = 1, \theta_j = 0 \text{ dla } j > q, \phi_j = 0 \text{ dla } j > p \quad (1.107)$$

i zapisując (1.80) w postaci  $\psi(z) \cdot \phi(z) = \theta(z)$ , otrzymujemy :

$$(\psi_0 + \psi_1 z^1 + \dots + \psi_n z^n + \dots)(1 - \phi_1 z^1 - \dots - \phi_p z^p) = 1 + \theta_1 z^1 + \dots + \theta_q z^q \quad (1.108)$$

skąd:

$$\begin{aligned} & \psi_0 + (-\psi_0 \phi_1 + \psi_1) z^1 + (-\psi_0 \phi_2 - \psi_1 \phi_1 + \psi_2) z^2 + \dots + (-\psi_0 \phi_j - \psi_1 \phi_{j-1} - \psi_2 \phi_{j-2} - \dots + \psi_j) z^j = \\ & = \begin{cases} 1 + \theta_1 z^1 + \dots + \theta_q z^q + 0z^{q+1} + 0z^{q+2} + \dots + 0z^p, & \text{gdy } p > q \\ 1 + \theta_1 z^1 + \dots + \theta_p z^p + \theta_{p+1} z^{p+1} + \theta_{p+2} z^{p+2} + \dots + \theta_q z^q, & \text{gdy } p \leq q \end{cases} \quad (1.109) \end{aligned}$$

Porównując w (1.109) współczynniki przy  $z^j$  i pamiętając że,  $\theta_0 = 1, \theta_j = 0$  dla  $j > q$ ,

$\phi_j = 0$  dla  $j > p$ , otrzymujemy :

$$\psi_0 = \theta_0 = 1 \quad (1.110)$$

a dla  $1 \leq j \leq \max(p-1, q)$  ( lub  $1 \leq j < \max(p, q+1)$  ):

$$\begin{aligned} \psi_1 - \psi_0 \phi_1 &= \theta_1 \\ \psi_2 - \psi_0 \phi_2 - \psi_1 \phi_1 &= \theta_2 \\ \vdots & \\ \psi_j - \psi_0 \phi_j - \psi_1 \phi_{j-1} - \dots - \psi_{j-1} \phi_1 &= \theta_j \end{aligned} \quad (1.112)$$

a dla  $j > \max(p-1, q)$ :

$$\psi_j - \psi_0 \phi_j - \psi_1 \phi_{j-1} - \dots - \psi_{j-1} \phi_1 = 0 \quad (1.114)$$

Zależności (1.110) do (1.114) można zapisać następująco:

$$\psi_0 = \theta_0 = 1 \quad (1.115)$$

$$\psi_j - \sum_{0 < k \leq j} \phi_k \psi_{j-k} = \theta_j, \quad 1 \leq j \leq \max(p-1, q) \quad (1.116)$$

$$\psi_j - \sum_{0 < k \leq p} \phi_k \psi_{j-k} = 0, \quad j \geq \max(p, q+1) \quad (1.117)$$

przy czym ograniczenie  $k \leq p$  w ostatniej sumie wynika z  $\phi_k = 0$  dla  $k > p$ .

Warunki (1.115) i (1.116) są warunkami brzegowymi na  $\psi_j$  dla  $1 \leq j \leq \max(p-1, q)$ ,

a równania (1.117) równaniami swobodnymi.

Rekurencyjne rozwiązanie (1.115) i (1.116) daje

$$\begin{aligned}\psi_0 &= \theta_0 = 1 \\ \psi_1 &= \theta_1 + \psi_0\phi_1 = \theta_1 + \phi_1 \\ \psi_2 &= \theta_2 + \psi_0\phi_2 + \psi_1\phi_1 = \theta_2 + \phi_2 + \theta_1\phi_1 + \phi_1^2 \\ &\dots\dots\dots\text{itd}\dots\dots\dots\end{aligned}\tag{1.118}$$

Natomiast ogólne rozwiązanie równań (1.117) ma postać [1]:

$$\psi_n = \sum_{i=1}^k \sum_{j=0}^{r_i-1} \alpha_{ij} n^j \xi_i^{-n}, \quad n > \max(p-1, q) - p \tag{1.119}$$

gdzie  $\xi_i, i = 1, 2, \dots, k$  są różnymi zerami wielomianu  $\phi(z)$ , a  $r_i$  jest wielokrotnością tych pierwiastków, tak że spełniony jest warunek:

$$\sum_{i=1}^k r_i = p, \tag{1.120}$$

gdzie  $p$  jest stopniem wielomianu  $\phi(z)$ . Liczba  $p$  stałych  $\alpha_{ij}$  oraz współczynników  $\psi_j$ , gdzie  $0 \leq j \leq \max(p-1, q) - p$ , jest wtedy jednoznacznie określona przez  $\max(p-1, q) + 1$  warunków brzegowych (1.115-116).

Oznacza to, że ciąg wartości  $\{\psi_j\}$  można wyznaczyć jednoznacznie, a stąd na podstawie (1.97') można wyznaczyć funkcję autokowariancji  $\gamma$ .

**Metoda druga**

Obecnie podamy alternatywną metodę wyznaczania funkcji autokowariancji przyczynowego procesu ARMA( $p, q$ ) danego równaniem:

$$\phi(B)X_t = \theta(B)Z_t. \tag{1.10'}$$

Oparta jest ona na znalezieniu różnicowych równań dla  $\gamma(k), k = 0, 1, 2, \dots$ , które otrzymuje się z równania (1.10') po przemnożeniu każdej strony tego równania przez  $X_{t-k}$  a następnie wzięciu wartości oczekiwanych:

$$E\{X_{t-k}(X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p})\} = E\{X_{t-k}(Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q})\}. \tag{1.121}$$

Z kolei skorzystajmy z przedstawienia (1.85) procesu ARMA  $(p, q)$  jako:

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \text{ skąd } X_{t-k} = \sum_{j=0}^{\infty} \psi_j Z_{t-j-k}$$

Po prawej stronie równania (1.121) otrzymujemy:

$$\begin{aligned} E \left\{ \sum_{j=0}^{\infty} \psi_j Z_{t-j-k} (Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}) \right\} &= \sum_{j=0}^{\infty} \psi_j E \left\{ Z_{t-k-j} \sum_{k'=0}^q \theta_{k'} Z_{t-k'} \right\} = \\ &= \sum_{j=0}^{\infty} \psi_j \sum_{k'=0}^q \theta_{k'} E \{ Z_{t-k-j} Z_{t-k'} \} = \dots \end{aligned}$$

i korzystając z faktu, że:  $E\{Z_{t-k-j}Z_{t-k'}\} = \sigma^2 \delta_{j,k'-k}$

otrzymujemy :

$$\dots = \sum_{j=0}^{\infty} \psi_j \sum_{k'=0}^q \theta_{k'} Z_{t-k'} \sigma^2 \delta_{j,k'-k} = \sigma^2 \sum_{k'=0}^q \psi_{k'-k} \theta_{k'} = \sigma^2 \sum_{j=0}^q \psi_{j-k} \theta_j, \quad 0 \leq k \leq q. \quad (1.122)$$

Natomiast z lewej strony równania (1.121) otrzymujemy ( $\phi_0 \equiv 1$ ):

$$E \left\{ X_{t-k} (X_t - \sum_{j=1}^p \phi_j X_{t-j}) \right\} = E(X_{t-k} X_t) - \sum_{j=1}^p \phi_j E(X_{t-k} X_{t-j}) = \gamma(k) - \sum_{j=1}^p \phi_j \gamma(k-j), \quad (1.123)$$

gdzie  $0 \leq k \leq p$ .

Ostatecznie z (1.122) i (1.123) otrzymujemy po podstawieniu do (1.121):

$$\gamma(k) - \phi_1 \gamma(k-1) - \dots - \phi_p \gamma(k-p) = \sigma^2 \sum_{k \leq j \leq q} \theta_j \psi_{j-k}, \quad \text{dla } 0 \leq k \leq \max(p, q), \quad (1.124)$$

Natomiast dla  $k > \max(p, q)$  otrzymujemy ze względu na (1.107):

$$\gamma(k) - \phi_1 \gamma(k-1) - \dots - \phi_p \gamma(k-p) = 0 \quad (1.125)$$

Ogólne rozwiązanie równania (1.125) ma postać [1]:

$$\gamma(h) = \sum_{i=1}^k \sum_{j=0}^{r_i-1} \beta_{ij} h^j \xi_i^{-h} \quad \text{dla } h \geq \max(p, q+1) - p \quad (1.126)$$

gdzie  $p$  wartości  $\beta_{ij}$  oraz kowariancje są jednoznacznie wyznaczone z warunków brzegowych (1.124) po tym jak z równań (1.118) wyznaczono  $\psi_0, \psi_1, \dots, \psi_q$ . Natomiast  $\xi_i, i = 1, 2, \dots, k$ , są miejscami zerowymi  $\phi(z)$  tak jak poprzednio w zależnościach (1.119-120).

### Przykład 1.6.1:

Wyznaczyć wartość  $\psi_n$  oraz  $\gamma(k)$  dla procesu  $\{X_t\}$  ARMA(2,1), spełniającego równanie:

$$(1 - \frac{2}{3}B + \frac{1}{9}B^2)X_t = (1 + B)Z_t, \quad \text{gdzie } \{Z_t\} \sim \text{WN}(0, \sigma^2) \quad (1.127)$$

### Metoda pierwsza:

Proces ten można zapisać w postaci:  $X_t - \frac{2}{3}X_{t-1} + \frac{1}{9}X_{t-2} = Z_t + Z_{t-1}$ . (1.127')

Korzystając z (1.118) otrzymujemy, przy parametrach modelu  $\phi_1 = \frac{2}{3}$ ,  $\phi_2 = -\frac{1}{9}$ ,  $\theta_1 = 1$ :

$$\psi_0 = \theta_0 = 1, \quad \psi_1 = \theta_1 + \psi_0\phi_1 = \theta_1 + \phi_1 = 1 + \frac{2}{3} = \frac{5}{3} \quad (1.128)$$

Ponieważ  $\xi_1 = \frac{1}{3}$  (1.129)

jest podwójnym pierwiastkiem równania charakterystycznego:

$$\phi(z) = 1 - \frac{2}{3}z + \frac{1}{9}z^2 = 0, \quad (1.130)$$

stąd ogólne rozwiązanie (1.119) jest postaci:

$$\psi_n = (\alpha_{10} + n\alpha_{11})3^{-n}, \quad n \geq 0 \quad (1.131)$$

gdzie korzystając z wyrażen z (1.128) otrzymujemy:  $\alpha_{10} = 1$ ,  $\alpha_{11} = 4$ . (1.132)

Podstawiając otrzymane wyniki do (1.131) otrzymujemy:

$$\psi_n = (1 + 4n)3^{-n}. \quad (1.133)$$

Wcześniej (1.97) pokazaliśmy, że:

$$\gamma(h) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|h|} \quad (1.97')$$

Ponieważ w (1.133), mamy:

$$\psi_j = (1 + 4j)3^{-j}, \quad \psi_{j+k} = (1 + 4(j+k))3^{-j-k} = (1 + 4j + 4k)3^{-j-k},$$

zatem z (1.97') otrzymujemy dla  $k \geq 0$ :

$$\begin{aligned} \gamma(k) &= \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+k} = \sigma^2 \sum_{j=0}^{\infty} (1 + 4j)3^{-j} (1 + 4j + 4k)3^{-j-k} = \sigma^2 \sum_{j=0}^{\infty} (1 + 4j)(1 + 4j + 4k)3^{-2j-k} = \\ &= \sigma^2 3^{-k} \sum_{j=0}^{\infty} (1 + 4j)(1 + 4j + 4k)9^{-j} = \sigma^2 3^{-k} \sum_{j=0}^{\infty} (1 + 8j + 4k + 16j^2 + 16jk)9^{-j} = \\ &= \sigma^2 3^{-k} \sum_{j=0}^{\infty} (1 + 4k)9^{-j} + (8j + 16jk)9^{-j} + 16j^2 9^{-j} = \\ &= \sigma^2 3^{-k} \sum_{j=0}^{\infty} (1 + 4k)9^{-j} + \sum_{j=0}^{\infty} (8j + 16jk)9^{-j} + \sum_{j=0}^{\infty} 16j^2 9^{-j} = \dots \end{aligned}$$



obliczając poszczególne sumy:

$$\sum_{j=0}^{\infty} (1+4k)9^{-j} = \frac{9}{8}(1+4k), \quad \sum_{j=0}^{\infty} (8j+16jk)9^{-j} = \frac{9}{16}(2+4k), \quad \sum_{j=0}^{\infty} 16j^2 9^{-j} = \frac{45}{16},$$

otrzymujemy:

$$\dots = \sigma^2 3^{-k} \left( \frac{9}{8}(1+4k) + \frac{9}{16}(2+4k) + \frac{45}{16} \right) = \sigma^2 3^{-k} \left( \frac{81}{16} + \frac{27}{4}k \right)$$

Zatem ostatecznie funkcja autokowariancji badanego procesu (1.127), ma postać:

$$\gamma(k) = \sigma^2 3^{-k} \left( \frac{81}{16} + \frac{27}{4}k \right) \quad (1.134)$$

### Metoda druga:

Mamy proces dany równaniem (1.127):

$$X_t - \frac{2}{3}X_{t-1} + \frac{1}{9}X_{t-2} = Z_t + Z_{t-1}.$$

Po przemnożeniu każdej strony tego równania przez  $X_{t-k}$ , i następnie wzięciu wartości oczekiwanych otrzymujemy, zgodnie z (1.124):

$$\gamma(k) - \frac{2}{3}\gamma(k-1) + \frac{1}{9}\gamma(k-2) = \sigma^2 \sum_{k \leq j \leq \infty} \theta_j \psi_{j-k}, \quad k \geq 2 \quad (1.135)$$

Z metody pierwszej otrzymaliśmy współczynniki (1.128):

$$\psi_0 = 1, \quad \psi_1 = \frac{5}{3}. \quad (1.128')$$

Z równania (1.135) dla wartości  $k=0$  otrzymujemy:

$$\gamma(0) - \frac{2}{3}\gamma(1) + \frac{1}{9}\gamma(2) = \frac{8}{3}\sigma^2, \quad (1.136a)$$

natomiast dla  $k=1$  otrzymujemy:

$$\gamma(1) - \frac{2}{3}\gamma(0) + \frac{1}{9}\gamma(1) = \sigma^2. \quad (1.136b)$$

Korzystając z (1.129-130) wiemy, że podwójny pierwiastek  $\phi(z) = 0$  wynosi  $\xi_1 = \frac{1}{3}$ .

Zatem ogólne rozwiązanie funkcji  $\gamma(k)$  (1.126) przyjmuje teraz postać:

$$\gamma(n) = (\beta_{10} + n\beta_{11}) \cdot 3^{-n} \quad n \geq 0 \quad (1.137)$$

i po podstawieniu za  $n=0,1,2$  otrzymujemy:

$$\gamma(0) = \beta_{10}, \quad \gamma(1) = \frac{1}{3}(\beta_{10} + \beta_{11}), \quad \gamma(2) = \frac{1}{9}(\beta_{10} + 2\beta_{11}). \quad (1.138)$$

Podstawiając (1.138) do pary równań (1.136) otrzymujemy układ równań:

$$\begin{cases} \frac{64}{27} \beta_{10} - \frac{16}{27} \beta_{11} = 8\sigma^2 \\ -\frac{8}{27} \beta_{10} + \frac{10}{27} \beta_{11} = \sigma^2 \end{cases} \quad (1.139)$$

Po jego rozwiązaniu otrzymujemy, że:

$$\beta_{10} = \frac{81}{16}, \beta_{11} = \frac{27}{4}, \quad (1.140)$$

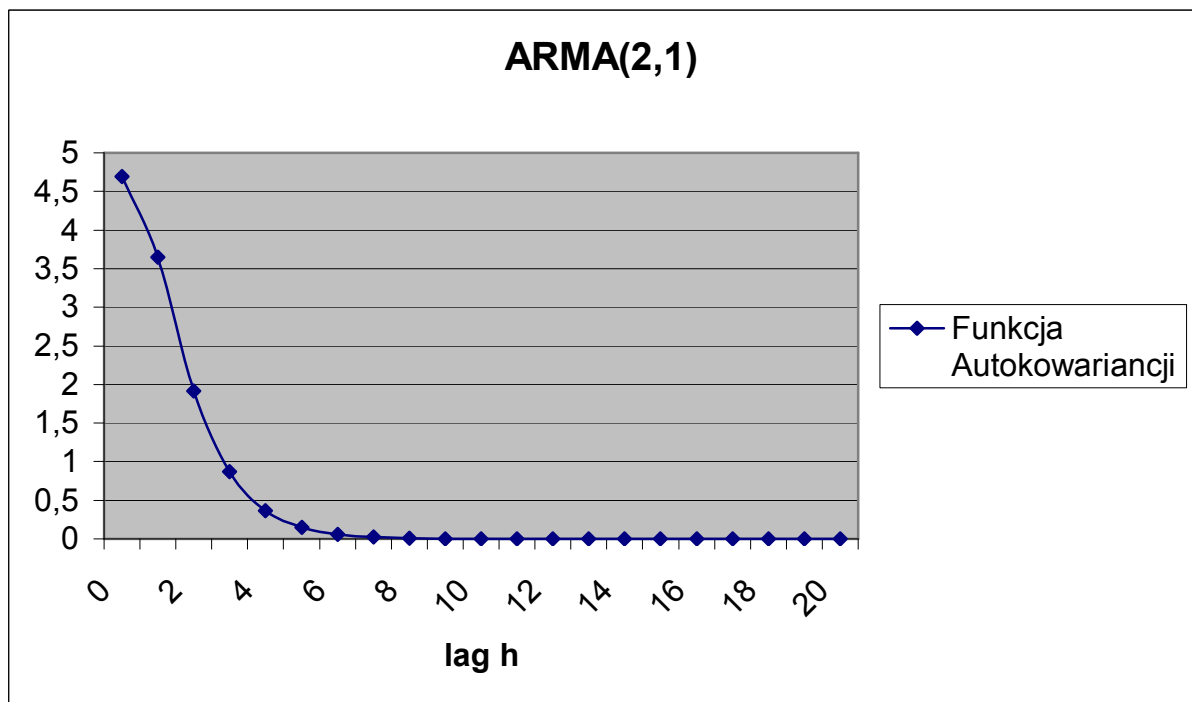
co po podstawieniu do równania (1.137) daje:

$$\gamma(k) = \sigma^2 3^{-k} \left( \frac{81}{16} + \frac{27}{4} k \right). \quad (1.141)$$

Wynik (1.141) można porównać z wynikiem otrzymanym z pierwszej metody (1.134).

Wyniki te są identyczne.

Poniższy wykres przedstawia funkcję autokowariancji naszego procesu.



Wykres 8. Funkcja autokowariancji procesu  $X_t - \frac{2}{3} X_{t-1} + \frac{1}{9} X_{t-2} = Z_t + Z_{t-1}$

## 1.7 Funkcja autokorelacji cząstkowej (PACF)

Podobnie jak funkcja autokorelacji, również funkcja autokorelacji częściowej niesie ze sobą informację o wewnętrznej strukturze zależności stacjonarnego procesu.

Podobnie jak ACF również PACF zależy jedynie od własności drugiego rzędu dla procesu.

Częściową autokorelacją  $\alpha(k)$  dla lag  $k$  można widzieć jako korelację pomiędzy  $X_1$  a  $X_{k+1}$  przy kontrolowanym wpływie zmiennych  $X_2, \dots, X_k$ .

Oznacza to, że zmienna  $X_1$  została dostrojona do zmiennych  $X_2, \dots, X_k$ , (np. na skutek regresji) i pozostałe reszty zmiennej  $X_1$ , nie wytłumaczone już wpływem zmiennych  $X_2, \dots, X_k$  zostały skorelowane z podobnie tworzonymi resztami zmiennej  $X_{k+1}$  dostrojonej do zmiennych  $X_2, \dots, X_k$ .

### Obecnie zdefiniujemy funkcję częściowej autokorelacji (PACF)

Niech  $\{X_t\}$  będzie stacjonarnym procesem ze średnią zero i funkcją autokowariancji  $\gamma(\cdot)$  taką, że  $\gamma(h) \rightarrow 0$  gdy  $h \rightarrow \infty$ , oraz założmy, że  $\phi_{kj}$ ,  $j=1, \dots, k$ ;  $k=1, 2, \dots$  są współczynnikami w:

$$P_{sp\{X_1, \dots, X_k\}} X_{k+1} = \sum_{j=1}^k \phi_{kj} X_{k+1-j}. \quad (1.142)$$

Wtedy z równania predykcji (Uzupełnienie 1):

$$\langle X_{k+1} - P_{sp\{X_1, \dots, X_k\}} X_{k+1}, X_j \rangle = 0, \quad j=1, \dots, k, \quad (1.143)$$

otrzymujemy:

$$\begin{bmatrix} \rho(0) & \rho(1) & \rho(2) & \cdots & \rho(k-1) \\ \rho(1) & \rho(0) & \rho(1) & \cdots & \rho(k-2) \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \rho(k-1) & \rho(k-2) & \rho(k-3) & \cdots & \rho(0) \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(k) \end{bmatrix}, \quad k=1, 2, \dots \quad (1.144)$$

### **Definicja 1.7.1**

Funkcja autokorelacji cząstkowej (PACF)  $\alpha(\cdot)$ , stacjonarnego szeregu czasowego jest zdefiniowana przez:

$$\alpha(1) = \text{Corr}(X_2, X_1) = \rho(1) \quad (1.145a)$$

oraz :

$$\alpha(k) = \text{Corr}(X_{k+1} - P_{sp\{1, X_2, \dots, X_k\}} X_{k+1}, X_1 - P_{sp\{1, X_2, \dots, X_k\}} X_1), \quad k \geq 2 \quad (1.145b)$$

gdzie  $\alpha(k)$  nazywamy autokorelacją cząstkową z opóźnieniem  $k$ .

Rzutowania dla  $X = X_1$  lub  $X = X_{k+1}$  są określone następująco:

$$\hat{X} \equiv P_{sp\{1, X_2, \dots, X_k\}} X = \beta_1 + \sum_{i=2}^k \beta_i X_i \quad (1.146)$$

gdzie  $\hat{X}$  jest najlepszym liniowym przewidywaniem (Uzupełnienie 1) dla  $X$  w przestrzeni wektorowej rozpiętej na  $1, X_2, \dots, X_k$ , natomiast równanie:

$$\langle \hat{X}, W_i \rangle = \langle X, W_i \rangle, \quad W_i = 1, X_2, \dots, X_k, \quad (1.147)$$

jest równaniem przewidywania (Uzupełnienie 1), gdzie  $\langle K, L \rangle \equiv E(K \cdot L)$ .

Zachodzi również (Rozdział 3, (3.42)):

$$\alpha(k) = \phi_{kk}, \quad \text{dla } k = 1, 2, \dots \quad (1.148)$$

gdzie  $\phi_{kk}$  jest określone jednoznacznie przez (1.144). Związek (1.148) można traktować jako definicję  $\alpha(k)$ .

### **Definicja 1.7.2**

Funkcja autokorelacji cząstkowej z próby  $\hat{\alpha}(k)$  z opóźnieniem  $k$  z  $\{x_1, \dots, x_n\}$  jest określona podobnie przez (pod warunkiem, że dla pewnych  $i$  oraz  $j$  zachodzi  $x_i \neq x_j$ ):

$$\hat{\alpha}(k) = \hat{\phi}_{kk}, \quad \text{dla } 1 \leq k \leq n \quad (1.149)$$

gdzie  $\hat{\phi}_{kk}$  jest wyznaczone z (1.144), natomiast każda autokorelacja  $\rho(j)$  zastąpiona jest autokorelacją z próby  $\hat{\rho}(j)$  (Uzupełnienie 1).

### Przykład 1.7.1

Niech  $\{X_t\}$  będzie przyczynowym procesem AR (1):

$$X_t = -0.8X_{t-1} + Z_t \quad (1.150)$$

czyli  $\phi_1 = -0.8$

$$\text{Zatem: } \alpha(1) = \text{Corr}(X_2, X_1) = \text{Corr}(\phi_1 X_1 + Z_2, X_1) = \phi_1 \text{Corr}(X_1, X_1) + \text{Corr}(Z_2, X_1)$$

$$\text{Korzystając z faktu, że: } \text{Corr}(X_1, X_1) = \rho(0) = 1 \quad (1.151)$$

$$\text{oraz, że: } \text{Corr}(Z_2, X_1) = \text{Corr}(Z_2, \sum_{j=0}^{\infty} Z_{1-j}) = 0 \quad (1.152)$$

$$\text{otrzymujemy: } \alpha(1) = \text{Corr}(X_2, X_1) = \phi_1 \quad (1.153)$$

Wyznamy teraz  $\alpha(k)$  dla  $k \geq 2$ .

Można pokazać (Rozdział 3), że jeśli stacjonarny proces ma średnią równą zero, wtedy

$$P_{SP\{1, X_2, \dots, X_n\}}(\cdot) = P_{SP\{X_2, \dots, X_n\}}(\cdot), \quad (3.2')$$

Korzystając z Uzupełnienia 1, otrzymujemy:

$$P_{sp\{X_2, \dots, X_k\}} X_{k+1} = \phi_1 X_k \quad (1.154)$$

gdzie w naszym przykładzie  $\phi_1 = -0.8$ .

Ponieważ proces jest stacjonarny, zatem macierze kowariancji dla układów zmiennych:

$$(X_1, X_2, \dots, X_k)^T, (X_k, \dots, X_2, X_1)^T, (X_{k+1}, X_k, \dots, X_2)^T, (X_2, X_3, \dots, X_{k+1})^T$$

są takie same. Zatem jeśli zachodzi (1.154) to również:

$$P_{sp\{X_2, \dots, X_k\}} X_1 = P_{sp\{X_k, \dots, X_2\}} X_1 = \phi_1 X_2 \quad (1.155)$$

więc dla  $k \geq 2$ :

$$\begin{aligned} \alpha(k) &= \text{Corr}(X_{k+1} - P_{sp\{X_2, \dots, X_k\}} X_{k+1}, X_1 - P_{sp\{X_2, \dots, X_k\}} X_1) = \\ &= \text{Corr}(X_{k+1} - \phi_1 X_k, X_1 - \phi_1 X_2) = \text{Corr}(Z_{k+1}, X_1 - \phi_1 X_2) \\ &= \text{Corr}(Z_{k+1}, X_1) - \phi_1 \text{Corr}(Z_{k+1}, X_2) = 0 \end{aligned} \quad (1.156)$$

gdzie skorzystano z faktu, że ponieważ:

$$X_s = \sum_{j=0}^{\infty} \psi_j Z_{s-j}$$

więc  $\text{Corr}(Z_t, X_s) = 0$  dla  $s < t$

Podsumowując, dla AR (1)

$$\begin{cases} \alpha(1) = \phi_1 \\ \alpha(k) = 0 \quad \text{dla } k \geq 2 \end{cases} \quad (1.157)$$

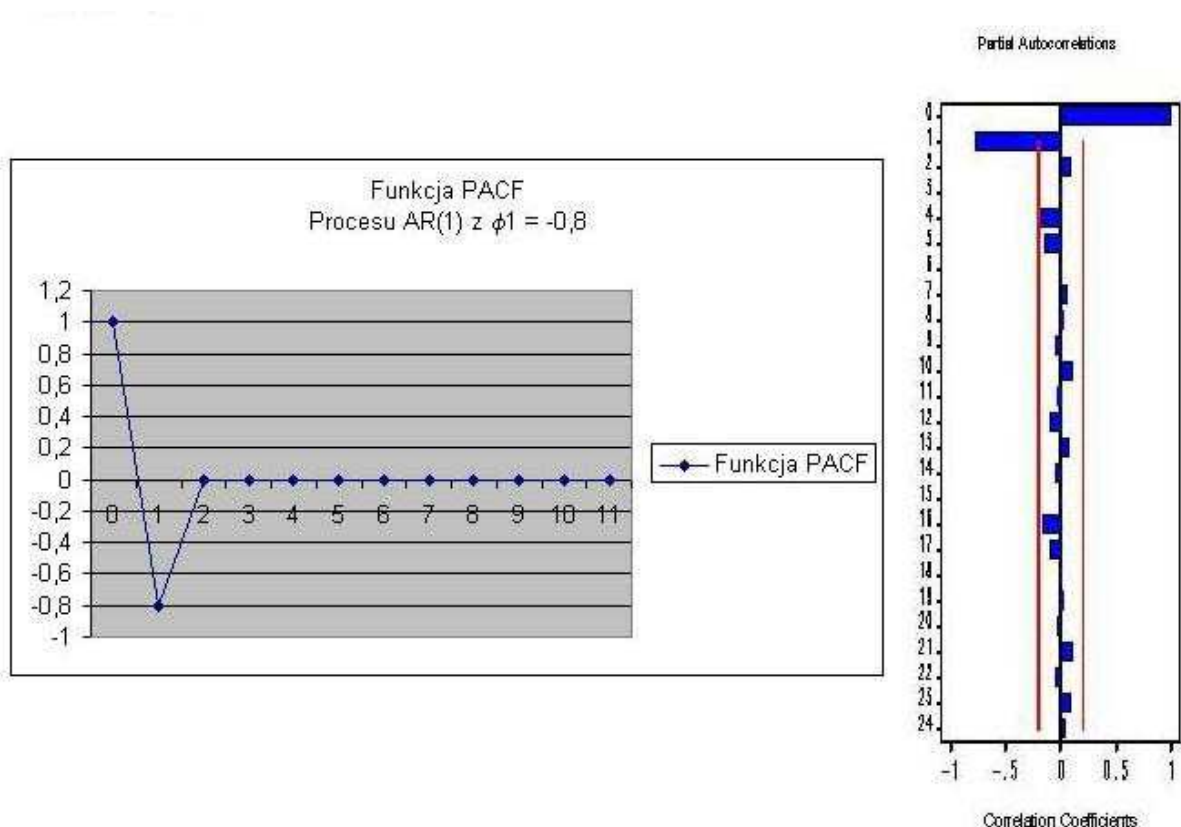
W naszym przypadku  $\phi_1 = -0.8$ .

Wniosek z powyższego przykładu jest taki, że ponieważ dla procesu AR(1):

$$\alpha(k) = 0 \quad \text{dla } k \geq 2$$

zatem korelacja pomiędzy  $X_{t-2}$  a  $X_t$  jest całkowicie wyeliminowana wtedy gdy całkowicie usuniemy z nich informację o zmiennej  $X_{t-1}$ .

Poniższy wykres przedstawia funkcję częściowej autokorelacji (po lewej) oraz funkcję częściowej autokorelacji z próby dla naszego procesu. (Realizację procesu  $\{Z_t\}$  wygenerowano generatorem liczb losowych SAS'a z rozkładu normalnego.)



Wykres 9. Autokorelacja cząstkowa z próby dla procesu danego równaniem  $X_t = -0,8X_{t-1} + Z_t$

Podobny efekt zachodzi dla  $AR(p)$ , tyle, że tym razem należy usunąć informację o  $p$  zmiennych leżących pomiędzy zmienną  $X_{t-(p+1)}$  a  $X_t$ .

### **Przykład 1.7.2**

Niech  $\{X_t\}$  będzie przyczynowym procesem  $AR(p)$ , który dany jest równaniem:

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t \text{ gdzie } \{Z_t\} \sim WN(0, \sigma^2) \quad (1.158)$$

Dla  $k > p$  mamy, najlepszym przewidywaniem średnio-kwadratowym jest (Uzupełnienie 1 (Przykład 1)):

$$P_{sp\{X_2, \dots, X_k\}} X_{k+1} = \sum_{j=1}^p \phi_{kj} X_{k+1-j} \quad (1.142')$$

zatem jeśli  $Y \in \overline{sp}\{X_2, \dots, X_k\}$ , wtedy z warunku przyczynowości (Definicja 1.3)

$Y \in \overline{sp}\{Z_j, j \leq k\}$  i wtedy ponieważ  $X_{k+1} - \sum_{j=1}^p \phi_j X_{k+1-j} \perp \overline{sp}\{Z_j, j \leq k\}$ , otrzymujemy:

$$\langle X_{k+1} - \sum_{j=1}^p \phi_j X_{k+1-j}, Y \rangle = \langle Z_{k+1}, Y \rangle = 0. \quad (1.159)$$

Dla  $k > p$  z (1.142') otrzymujemy:

$$\begin{aligned} \alpha(k) &= \text{Corr}(X_{k+1} - \sum_{j=1}^p \phi_{kj} X_{k+1-j}, X_1 - P_{sp\{X_2, \dots, X_k\}} X_1) = \\ &= \text{Corr}(Z_{k+1}, X_1 - P_{sp\{X_2, \dots, X_k\}} X_1) = 0. \end{aligned} \quad (1.160)$$

### **Wnioski.**

Widzimy, że dla procesów autoregresyjnych wskaźnikiem, który może nam wskazać rząd  $p$  autoregresji jest funkcja autokorelacji częściowej. Gdyż dla  $k > p$ ,  $\alpha(k) = 0$ .

W przypadku procesów  $MA(q)$  wykres funkcji autokorelacji częściowej można wyznaczyć jak w poniższym przykładzie.

### **Przykład 1.7.3**

Niech  $\{X_t\}$  będzie procesem średniej ruchomej  $MA(1)$ .

Proces ten możemy zapisać jako:

$$X_t = Z_t + \theta_1 Z_{t-1}, \text{ gdzie } \{Z_t\} \sim WN(0, \sigma^2) \text{ oraz } |\theta_1| < 1 \quad (1.161)$$

Korzystając z (1.97) i Przykładu 1.5.1 dla MA( $q$ ) gdy  $q = 1$ , można pokazać, że dla  $k = 1$  mamy:

$$\alpha(1) = \rho(1) = \theta_1 / (1 + \theta_1^2) \quad (1.162)$$

Ponieważ [1]:

$$P_{sp\{X_2\}}^- X_3 = [\theta_1 / (1 + \theta_1^2)] X_2 = P_{sp\{X_2\}}^- X_1 \quad (1.163)$$

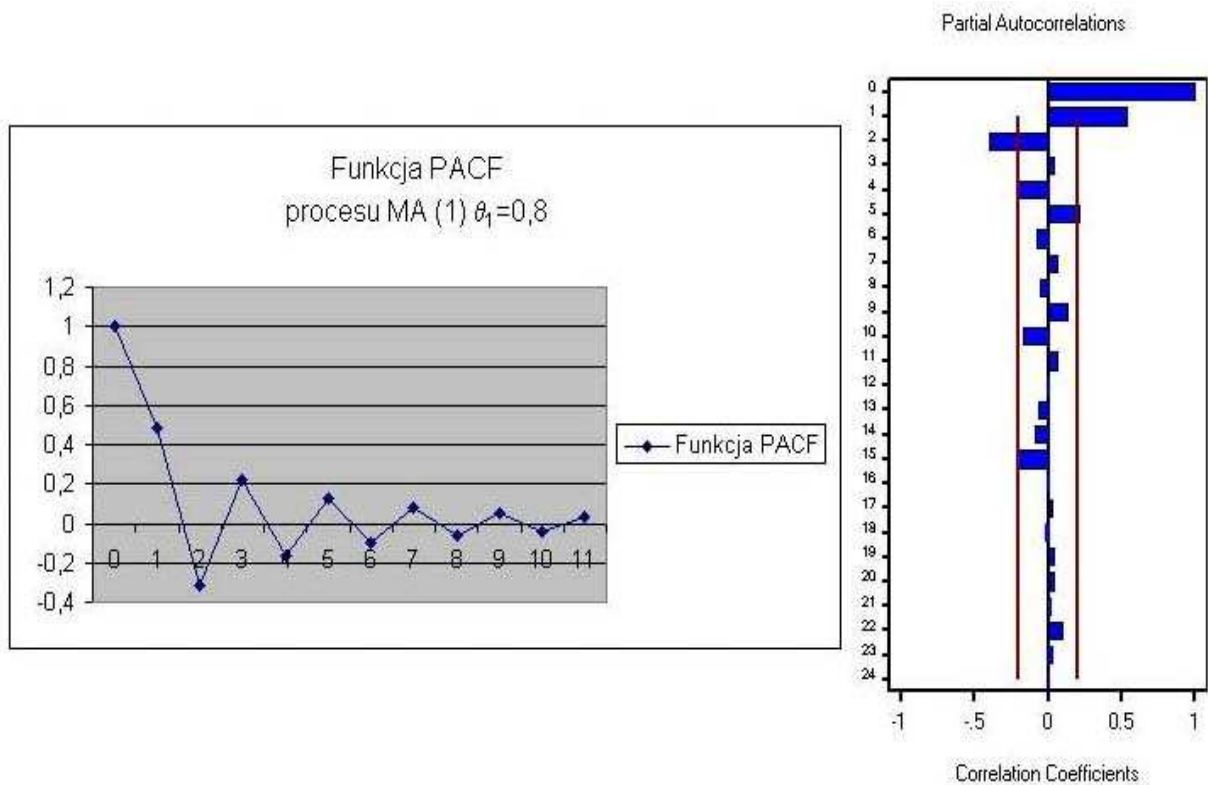
zatem:

$$\alpha(2) = \text{Corr}(X_3 - \theta_1(1 + \theta_1^2)^{-1} X_2, X_1 - \theta_1(1 + \theta_1^2)^{-1} X_2) = -\theta_1^2 / (1 + \theta_1^2 + \theta_1^4) \quad (1.164)$$

i po kolejnych przeliczeniach:

$$\alpha(k) = -\frac{(-\theta_1)^k (1 - \theta_1^2)}{1 - \theta_1^{2(k+1)}} \quad \text{dla } k \geq 1 \quad (1.165)$$

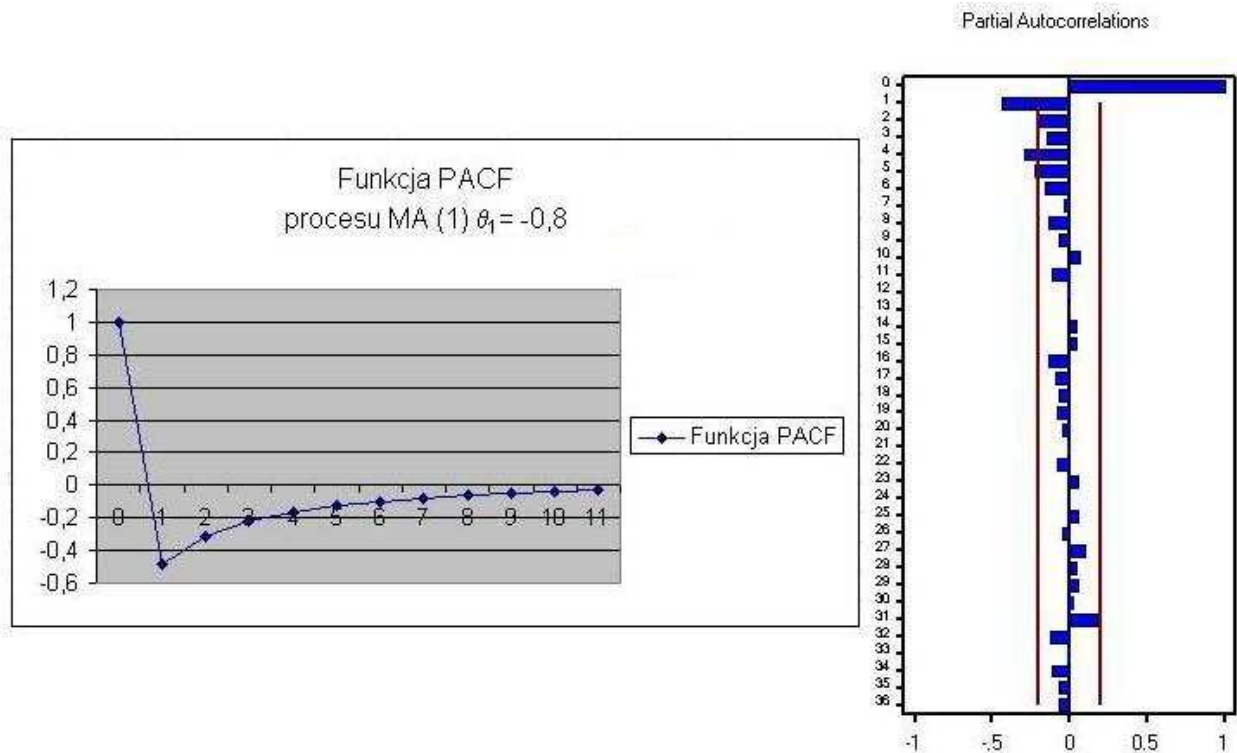
Poniższy wykres przedstawia funkcję częściowej autokorelacji (po lewej), oraz funkcję częściowej autokorelacji z próby dla procesu MA(1) z  $\theta_1 = 0,8$ . (Realizację procesu  $\{Z_t\}$  wygenerowano generatorem liczb losowych SAS'a z rozkładu normalnego.)



Wykres 9. autokorelacji cząstkowej dla procesu danego równaniem  $X_t = Z_t + 0,8Z_{t-1}$



Dla porównania wykres funkcji częściowej autokorelacji (po lewej), oraz funkcji częściowej autokorelacji z próby dla procesu MA(1) z  $\theta_1 = -0,8$ . (Realizację procesu  $\{Z_t\}$  wygenerowano generatorem liczb losowych SAS'a z rozkładu normalnego.)



Wykres 10. autokorelacji cząstkowej dla procesu danego równaniem  $X_t = Z_t - 0,8Z_{t-1}$

## 1.8 Procesy niestacjonarne ARIMA

Przedstawione wyżej procesy pozwalają na modelowanie i prognozowanie stacjonarnych szeregów czasowych. Jeżeli szereg czasowy nie jest stacjonarny, to często możemy przekształcić go w szereg stacjonarny w wyniku operacji różnicowania. Polega ona na d-krotnym obliczeniu różnic (różnicowaniu) sąsiednich wyrazów szeregu.

$$\text{Czyli pierwsza różnica } (d=1): X_t - X_{t-1}, \quad (1.166)$$

$$\text{druga różnica } (d=2): (X_t - X_{t-1}) - (X_{t-1} - X_{t-2}) = X_t - 2X_{t-1} + X_{t-2} \text{ itd.} \quad (1.167)$$

Parametr  $d$  ustala się na takim poziomie, aby otrzymany w wyniku tej operacji szereg czasowy różnic był stacjonarny.

### **Definicja 1.8.1 Procesu zintegrowanego**

Szereg stacjonarny, który można sprowadzić do szeregu stacjonarnego, obliczając przyrosty  $d$  razy nazywamy szeregiem zintegrowanym stopnia  $d$ .

Opisany proces różnicowania można zapisać za pomocą tzw. operatora różnicowania rzędu  $d$  (oznaczonego przez  $\nabla^d$ ). Operator  $\nabla^d$  często wyraża się za pomocą operatora cofnięcia  $B$  zdefiniowanego w (1.13)<sup>1</sup>:

$$\nabla^d = (1 - B)^d . \quad (1.168)$$

Otrzymujemy, zatem:

$$\nabla X_t = (1 - B)X_t = X_t - X_{t-1}$$

$$\nabla^2 X_t = \nabla(\nabla X_t) = \nabla X_t - \nabla X_{t-1} = (X_t - X_{t-1}) - (X_{t-1} - X_{t-2}), \quad (1.169a)$$

lub  $\nabla^2 X_t = (1 - B)^2 X_t = (1 - B)(1 - B)X_t = (1 - 2B + B^2)X_t = X_t - 2X_{t-1} + X_{t-2}$

.....

$$\nabla^d X_t = \nabla \nabla^{d-1} X_t = \nabla^{d-1} \nabla X_t = \nabla^{d-1} X_t - \nabla^{d-1} X_{t-1} = (1 - B)^d X_t = \sum_{l=0}^d \binom{d}{l} (-B)^l X_t \quad (1.170b)$$

skąd wynika, że operacje wielomianami z  $B$  przebiegają tak samo jak na wielomianach zwykłych funkcji rzeczywistych.

Jeżeli na szereg czasowy  $\{X_t, t = 0, 1, 2, \dots, n\}$  o liczbie wyrazów  $n$  zadziałamy operatorem  $\nabla^d$  (inaczej mówiąc obliczymy różnicę rzędu  $d$ ), to otrzymamy szereg czasowy  $\{W_t, t = d + 1, \dots, n\}$  o liczbie wyrazów  $n - d$ . Gdy okaże się on stacjonarny, to możemy go modelować wykorzystując wyżej przedstawione procesy ARMA.

### **Definicja 1.8.2 Procesu ARIMA( $p, d, q$ )**

Jeżeli  $d$  jest nieujemną liczbą całkowitą, wtedy  $\{X_t\}$  nazywamy procesem ARIMA( $p, d, q$ ) gdy  $Y_t = (1 - B)^d X_t$  jest przyczynowym procesem ARMA( $p, q$ ), gdzie  $p$  – parametr autoregresji,  $q$  – parametr średniej kroczącej i  $d$  parametr liczby różnicowań.

---

<sup>1</sup> Nie należy mylić operatora różnicowania  $\nabla^d = (1 - B)^d$  z operatorem cofnięcia z lag  $D$ , czyli z  $\nabla_D = (1 - B^D)$ , który wykorzystuje się do eliminacji sezonowości z pierwotnych danych. Uogólnieniem procesów ARMA na procesy z trendem i sezonowością są procesy SARIMA [1], którymi nie będziemy się zajmowali w tej pracy, chociaż krok ten nie byłby trudny.

Definicja ta oznacza to, że  $\{X_t\}$  spełnia równanie różnicowe o postaci:

$$\phi^*(B)X_t \equiv \phi(B)(1-B)^d X_t = \theta(B)Z_t \quad \{Z_t\} \sim \text{WN}(0, \sigma^2) \quad (1.171)$$

gdzie  $\phi(z)$  i  $\theta(z)$  są kolejno wielomianami stopnia  $p$  i  $q$ , gdzie  $\phi(z) \neq 0$  dla  $|z| \leq 1$ .

Wielomian  $\phi^*(B)$  ma zero rzędu  $d$  w  $z = 1$ .

Proces  $\{X_t\}$  jest stacjonarny, wtedy i tylko wtedy, gdy  $d=0$ , co oznacza, że redukuje się on do procesu ARMA  $(p, q)$ . Możemy zauważyć, że jeśli  $d \geq 1$ , wtedy możemy dodać dowolny wielomian trendu stopnia  $(d-1)$  do  $\{X_t\}$  bez naruszania równania różnicowego (1.171).

Dlatego modele ARIMA przydatne są do reprezentowania danych z trendem, chociaż mogą być również używane do szeregów bez trendu.

Z wyjątkiem przypadku gdy  $d=0$ , średnia  $\{X_t\}$  nie jest wyznaczona przez równanie i w szczególności może wynosić zero.

Ponieważ dla  $d \geq 1$  równanie określa własności drugiego rzędu dla procesu  $\{(1-B)^d X_t\}$  a nie dla  $\{X_t\}$ , estymacja  $\phi$  i  $\theta$  oraz  $\sigma^2$  będzie bazować na obserwowanych wartościach (zróżnicowanych) dla  $(1-B)^d X_t$ .

W celu określenia liczby koniecznych różnicowań, powinno się przeanalizować wykres danych wejściowych i funkcji autokorelacji (ACF). Znaczące zmiany w poziomie (tendencje wzrostowe lub zniżkowe) wymagają zwykle różniczkowania odpowiedniego dla eliminacji trendu (z lag=1) a silne zmiany nachylenia zwykle wymagają dwukrotnego różnicowania. Sezonowe modele wymagają odpowiedniego różnicowania sezonowego. Jeśli szacowane współczynniki autokorelacji zanikają wolno z długimi opóźnieniami  $h$ , wtedy pierwsza operacja różnicowania jest zwykle potrzebna. [3]. W SAS'ie do oceny stacjonarności (a więc do oceny czy, i ile razy potrzeba zróżnicować szereg aby był on stacjonarny) posługujemy się testem Dickey'a – Fuller'a. (Uzupełnienie 3).

## Rozdział 2. Estymacja parametrów $\mu$ oraz $\rho(\cdot)$

Jeśli  $\{X_t\}$  jest rzeczywistym stacjonarnym procesem, wtedy uwzględniając jedynie momenty do drugiego rzędu włącznie, proces ten jest scharakteryzowany jedynie przez średnią  $\mu$  i funkcję autokowariancji  $\gamma(\cdot)$ .

Zatem estymacja  $\mu$ ,  $\gamma(\cdot)$  oraz funkcji autokorelacji  $\rho(\cdot) = \frac{\gamma(\cdot)}{\gamma(0)}$  na podstawie próby  $X_1, X_2, \dots, X_n$  odgrywa najistotniejszą rolę w problemie wnioskowania, a w tym, w konstrukcji właściwego modelu dla opisu danych.

### 2.1 Estymacja średniej $\mu$

Nieobciążonym estymatorem średniej  $\mu$  dla stacjonarnego procesu  $\{X_t\}$  jest średnia z próby:

$$\bar{X}_n = \frac{1}{n}(X_1 + \dots + X_n) = \frac{1}{n} \sum_{i=1}^n X_i \quad (2.1)$$

W pierwszej kolejności rozważymy własności średnio – kwadratowego błędu  $E(\bar{X}_n - \mu)^2$  (tzn. wariancji średniej  $\bar{X}_n$ ) dla dużych  $n$ .

#### **Twierdzenie 2.1**

Jeśli  $\{X_t\}$  jest stacjonarnym procesem ze średnią równą  $\mu$  i funkcją autokowariancji  $\gamma(\cdot)$ , wtedy dla  $n \rightarrow \infty$ :

$$1. \quad \sigma^2(\bar{X}_n) \equiv E(\bar{X}_n - \mu)^2 \rightarrow 0, \text{ o ile } \gamma(n) \rightarrow 0 \quad (2.2)$$

$$2. \quad nE(\bar{X}_n - \mu)^2 \rightarrow \sum_{h=-\infty}^{\infty} \gamma(h), \text{ o ile } \sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty \quad (2.3)$$

Dowód twierdzenia [1]. Istotnym elementem dowodu jest pokazanie, że:

$$\sigma^2(\bar{X}_n) = \frac{1}{n} \sum_{|h|=0}^{n-1} \left(1 - \frac{|h|}{n}\right) \gamma(h) \leq \quad (2.4)$$

$$\leq \frac{1}{n} \sum_{|h|=0}^{n-1} |\gamma(h)| \quad (2.5)$$

Uwaga: Wprowadźmy oznaczenie  $AN$  na asymptotyczną ( $n \rightarrow \infty$ ) normalność rozkładu.

## **Twierdzenie 2.2**

Jeśli  $\{X_t\}$  jest stacjonarnym procesem

$$X_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}, \quad \{Z_t\} \sim \text{iid}(0, \sigma^2) \quad (2.6)$$

gdzie 
$$\sum_{j=-\infty}^{\infty} |\psi_j| < \infty \quad (2.7)$$

oraz 
$$\sum_{j=-\infty}^{\infty} \psi_j \neq 0 \quad (2.8)$$

wtedy:

$$\bar{X}_n \text{ jest } AN\left(\mu, \frac{v}{n}\right), \quad (2.9)$$

gdzie z (2.3) w Twierdzeniu 2.1 oraz z zależności (U1.40) w Uzupełnieniu 1:

$$\sum_{h=-\infty}^{\infty} \gamma(h) = \sigma^2 \left( \sum_{j=-\infty}^{\infty} \psi_j \right)^2, \quad (2.10)$$

mamy, że

$$v \equiv \sum_{h=-\infty}^{\infty} \gamma(h) = \sigma^2 \left( \sum_{j=-\infty}^{\infty} \psi_j \right)^2, \quad (2.11)$$

gdzie  $\gamma(\cdot)$  jest funkcją autokowariancji procesu  $\{X_t\}$ .

## **Uwaga.**

Powyższe twierdzenie jest użyteczne przy znajdowaniu przybliżonej postaci przedziału ufności dla  $\mu$  przy wystarczająco dużej próbie.

## **Uwaga.**

Jeśli proces  $\{X_t\}$  jest nie tylko stacjonarny ale także Gaussowski, wtedy możemy zapisać dokładną postać rozkładu dla  $\bar{X}_n$ .

Wykorzystując twierdzenie o addytywności rozkładu normalnego, otrzymujemy z Twierdzenia 2.1 i wzoru (2.4):

$$\bar{X}_n \sim N\left(\mu, \frac{1}{n} \sum_{|h|=0}^{n-1} \left(1 - \frac{|h|}{n}\right) \gamma(h)\right) \quad (2.12)$$

lub 
$$n^{1/2}(\bar{X}_n - \mu) \sim N\left(0, \sum_{|h|=0}^{n-1} \left(1 - \frac{|h|}{n}\right) \gamma(h)\right). \quad (2.13)$$

zależność (2.13) pozwala na wyznaczenie dokładnej postaci przedziału ufności dla  $\mu$  (o ile znamy  $\gamma(\cdot)$ ):

$$(\bar{X}_n - u_\alpha \sigma(\bar{X}_n), \bar{X}_n + u_\alpha \sigma(\bar{X}_n)) \quad (2.14)$$

gdzie  $\sigma(\bar{X}_n) = \sqrt{\sigma^2(\bar{X}_n)}$ , a  $\sigma^2(\bar{X}_n)$  dana jest wzorem (2.4), natomiast  $u_\alpha = u(1 - \frac{\alpha}{2})$ , jest kwantylem rzędu  $(1 - \frac{\alpha}{2})$  rozkładu  $N(0, 1)$ .

## 2.2 Estymacja funkcji autokowariancji $\gamma(\cdot)$ i autokorelacji $\rho(\cdot)$

Estymatorami dla  $\gamma(h)$  i  $\rho(h)$ , którymi będziemy się posługiwać są statystyki:

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-h} (X_t - \bar{X}_n)(X_{t+h} - \bar{X}_n), \quad 0 \leq h \leq n-1 \quad (2.15)$$

i 
$$\hat{\rho}(h) = \hat{\gamma}(h) / \hat{\gamma}(0). \quad (2.16)$$

gdzie 
$$\bar{X}_n = \frac{1}{n}(X_1 + \dots + X_n) = \frac{1}{n} \sum_{i=1}^n X_i. \quad (2.1')$$

Estymator  $\hat{\gamma}(h)$  funkcji autokowariancji  $\gamma(h)$ , zadany przez (2.15) jest obciążony.

Jednak jego rozkład asymptotyczny (dla  $n \rightarrow \infty$ ) ma średnią  $\gamma(h)$ .

Estymatory  $\hat{\gamma}(h)$ ,  $0 \leq h \leq n-1$  mają również, dla każdego  $n \geq 1$  ważną własność związaną z nieujemną określonością macierzy :

$$\hat{\Gamma}_n = \begin{pmatrix} \hat{\gamma}(0) & \hat{\gamma}(1) & \dots & \hat{\gamma}(n-1) \\ \hat{\gamma}(1) & \hat{\gamma}(0) & \dots & \hat{\gamma}(n-2) \\ \vdots & \vdots & \dots & \vdots \\ \hat{\gamma}(n-1) & \hat{\gamma}(n-2) & \dots & \hat{\gamma}(0) \end{pmatrix} \quad (2.17)$$

Okazuje się, że macierz  $\hat{\Gamma}_n$  można zapisać następująco:

$$\hat{\Gamma}_n = n^{-1} T T^T \quad (2.18)$$

gdzie  $T$  jest macierzą stopnia  $n \times 2n$ :

$$T = \begin{pmatrix} 0 & \dots & 0 & 0 & Y_1 & Y_2 & \dots & Y_n \\ 0 & \dots & 0 & Y_1 & Y_2 & \dots & Y_n & 0 \\ \vdots & \dots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & Y_1 & Y_2 & \dots & Y_n & 0 & \dots & 0 \end{pmatrix} \quad (2.19)$$

gdzie:

$$Y_i = X_i - \bar{X}_n, \quad i = 1, 2, \dots, n \quad (2.20)$$

Dowód zależności (jako bardzo prosty) pomijamy.

Nieujemna określoność macierzy estymatorów funkcji autokowariancji wynika z tego, że dla dowolnego rzeczywistego,  $n \times 1$  wymiarowego wektora  $\vec{a}$ , zachodzi:

$$\vec{a}^T \Gamma_n \vec{a} = \vec{a}^T \frac{1}{n} T T^T \vec{a} = \frac{1}{n} (\vec{a}^T T) (\vec{a}^T T)^T \geq 0 \quad (2.21)$$

gdzie  $(\vec{a}^T T) (\vec{a}^T T)^T$  jest długością wektora  $(\vec{a}^T T)$ .

Tak więc macierz funkcji autokowariancji z próby jest nieujemnie określona.

Wynika z tego, że macierz funkcji autokorelacji z próby:

$$\hat{R}_n = \frac{\hat{\Gamma}_n}{\hat{\gamma}(0)} \quad (2.22)$$

jest również nieujemnie określona.

### **Uwaga.**

Czynnik  $\frac{1}{n}$ , który pojawił się w definicji estymatora  $\hat{\gamma}(h)$  jest czasami zastępowany przez  $\frac{1}{n-h}$ . Okazuje się jednak [1], że wtedy macierze  $\hat{\Gamma}_n$  oraz  $\hat{R}_n$  mogą nie być nieujemnie określone.

Będziemy się, więc zawsze posługiwać definicjami (2.15) i (2.16). Można zauważyć również, że z powodu symetryczności macierzy  $\hat{\Gamma}_n$ , otrzymujemy (por. Twierdzenie 3.1):

$$\det \hat{\Gamma}_n > 0, \text{ jeśli tylko } \hat{\gamma}(h) > 0. \quad (2.23)$$

### **Uwaga.**

Oszacowywanie na podstawie próby  $X_1, X_2, \dots, X_n$  parametrów  $\gamma(h)$  oraz  $\rho(h)$  na podstawie (2.15) i (2.16) jest mało solidne gdy  $h < n$  ( $h$  jest niewiele mniejsze od  $n$ ), gdyż mamy wtedy niewiele par  $(X_t, X_{t+h})$  z których liczymy  $\hat{\gamma}(h)$  i  $\hat{\rho}(h)$ . Dla  $h = n - 1$  istnieje tylko jedna taka para. Natomiast oszacowanie dla  $h \geq n$  jest niemożliwe.

Box i Jenkins sugerują, że użyteczny estymator dla  $\rho(h)$  można otrzymać wtedy, gdy  $n > 50$  i  $h \leq n/4$ .

### Uwaga.

Przy wyborze dla danego zbioru danych, właściwego modelu ARMA, istotną rzeczą jest możliwość rozpoznania, czy wartości funkcji autokorelacji z próbki są istotnie statystycznie różne od zera.

Poniżej podamy twierdzenie, które wypowiada się co do asymptotycznego rozkładu łącznego dla  $\hat{\rho}(1), \hat{\rho}(2), \dots, \hat{\rho}(h)$  przy  $n \rightarrow \infty$ . Twierdzenie to jest użyteczne przy stwierdzeniu istotności statystycznej wartości estymatorów  $\hat{\rho}(k), k=1,2,\dots,h$  w próbce.

### Twierdzenie 2.3.

Jeśli  $\{X_t\}$  jest procesem stacjonarnym:

$$X_t - \mu = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}, \{Z_t\} \sim \text{iid}(0, \sigma^2) \quad (2.24)$$

gdzie :

$$\sum_{j=-\infty}^{\infty} |\psi_j| < \infty \quad (2.25)$$

$$EZ_t^4 < \infty \quad (2.26)$$

wtedy, dla każdego  $h \in \{1,2,\dots\}$  mamy:

$$\hat{\rho}(h) \text{ jest } AN(\bar{\rho}(h), n^{-1}W) \quad (2.27)$$

gdzie:  $\hat{\rho}(h)^T = (\hat{\rho}(1), \hat{\rho}(2), \dots, \hat{\rho}(h))$  (2.28)

$$\bar{\rho}(h)^T = (\rho(1), \rho(2), \dots, \rho(h)) \quad (2.29)$$

a  $W$  jest macierzą kowariancji, której element  $(i, j)$ -ty zadany jest formułą Bartlett'a:

$$w_{ij} = \sum_{k=-\infty}^{\infty} \rho(k+i)\rho(k+j) + \rho(k-i)\rho(k+j) + 2\rho(i)\rho(j)\rho^2(k) - 2\rho(i)\rho(k)\rho(k+j) - 2\rho(j)\rho(k)\rho(k+i) \quad (2.30)$$

lub równoważnie:

$$w_{ij} = \sum_{k=1}^{\infty} \{\rho(k+i) + \rho(k-i) - 2\rho(i)\rho(k)\} * \{\rho(k+j) + \rho(k-j) - 2\rho(j)\rho(k)\} \quad (2.31)$$

Dowód [1].



### **Uwaga.**

Wprowadzając zmienne:

$$Y_i = \sum_{k=1}^{\infty} \{\rho(k+i) + \rho(k-i) - 2\rho(i)\rho(k)\} N_k, \quad i=1,2,\dots,h \quad (2.32)$$

gdzie  $N_1, N_2, \dots$  są zmiennymi losowymi (niezależnymi z takim samym rozkładem normalnym), tzn. iid  $N(0,1)$  i odwołując się do Twierdzenia 2.3, można pokazać (dowód nie jest trudny), że łączny rozkład asymptotyczny ( $n \rightarrow \infty$ ) rozkład wektora losowego  $(\hat{\rho}(h) - \bar{\rho}(h)) \cdot n^{1/2}$  jest taki sam jak rozkład wektora losowego  $\bar{Y} = (Y_1, Y_2, \dots)^T$ , tzn.

$$\bar{Y} = (Y_1, Y_2, \dots)^T \sim N(0, W) \quad (2.33)$$

oraz :

$$(\hat{\rho}(h) - \bar{\rho}(h)) \cdot n^{1/2} \sim AN(0, W) \quad (2.34)$$

### **Uwaga.**

Można pokazać [1], że zmienne:

$$Y_i \equiv \sigma^{-2} n^{-1/2} \sum_{t=1}^n Z_t Z_{t+i} \quad (2.35)$$

gdzie  $\{Z_t\} \sim \text{iid}(0, \sigma^2)$ , są iid i mają asymptotycznie ( $n \rightarrow \infty$ ) rozkład  $N(0,1)$ , oraz że łącznie określają graniczny rozkład wektora losowego  $(\hat{\rho}(h) - \bar{\rho}(h)) \cdot n^{1/2}$ .

### **Twierdzenie 2.4**

W Twierdzeniu 2.3 można zastąpić warunek (2.26)  $EZ_t^4 < \infty$  nieco silniejszym:

$$\sum_{j=-\infty}^{\infty} \psi_j^2 |j| < \infty, \quad (2.36)$$

narzuconym na ciąg  $\{\psi_j\}$  [1].

### **Uwaga.**

Założenia Twierdzenia 2.3 oraz Twierdzenia 2.4 są spełnione przez każdy proces ARMA ( $p, q$ ) z ciągiem  $\{Z_t\} \sim \text{iid}(0, \sigma^2)$ , gdzie  $\sigma^2 < \infty$ .

Warto zauważyć, że założenie  $\{Z_t\} \sim \text{iid}(0, \sigma^2)$  może być zastąpione w przypadku nie posiadających identycznych rozkładów zmiennych  $Z_t$ , warunkami ograniczenia dla  $E|Z_t^3|$  i  $E|Z_t^6|$ . Tzn. warunki  $E|Z_t^3| < \infty$  oraz  $E|Z_t^6| < \infty$  również pozwalają na wyprowadzenie powyższego centralnego twierdzenia granicznego (Twierdzenie 2.3, Twierdzenie 2.4).

**Przykład** (Niezależny biały szum)

Niech  $\{X_t\} \sim \text{iid}(0, \sigma^2)$ . (2.37)

Wtedy  $\rho(l) = 0$  dla  $|l| > 0$ . (2.38)

Z zależności (2.31) otrzymujemy, że:

$$w_{ij} = \sum_{k=1}^{\infty} \{\rho(k+i) + \rho(k-i) - 2\rho(i)\rho(k)\} \cdot \{\rho(k+j) + \rho(k-j) - 2\rho(j)\rho(k)\} = \dots$$

podstawiając  $\rho(l) = \rho(0)\delta_{l,0}$

otrzymujemy :

$$\dots = \sum_{k=1}^{\infty} \{\rho(0)(\delta_{0,k+i} + \delta_{0,k-i} - 2\rho(0)\delta_{0,i}\delta_{0,k})\} \{\rho(0)(\delta_{0,k+j} + \delta_{0,k-j} - 2\rho(0)\delta_{0,j}\delta_{0,k})\} = \dots$$

z powyższego równania widać, że tylko dla  $k = i = j$ ,  $\delta_{0,k-i}$  oraz  $\delta_{0,k-j}$  są równe 1 ,

$$\dots = \{\rho(0)(0+1-0)\} \cdot \{\rho(0)(0+1-0)\} = \rho^2(0) \tag{2.39}$$

Zatem:

$$w_{ij} = \begin{cases} \rho^2(0) & \text{dla } i = j \\ 0 & \text{dla } i \neq j \end{cases} \tag{2.40}$$

i ponieważ :  $\rho(l) = \frac{\gamma(l)}{\gamma(0)}$  (2.41)

więc:

$$w_{ij} = \begin{cases} 1 & \text{dla } i = j \\ 0 & \text{dla } i \neq j \end{cases} \tag{2.42}$$

Ponieważ dla dużego  $n$  (a dokładnie dla  $n \rightarrow \infty$ ), mamy (2.27):

$$\vec{\hat{\rho}}(h) = (\hat{\rho}(1), \hat{\rho}(2), \dots, \hat{\rho}(h)) \text{ jest } AN(\rho(h), n^{-1}W), \tag{2.43}$$

więc z (2.42) widać, że dla białego niezależnego szumu  $\{X_t\} \sim \text{iid}(0, \sigma^2)$ , macierz kowariancji

$W = (w_{ij}) = (\text{Cov}(\hat{\rho}(i), \hat{\rho}(j)))$  ma postać:

$$W = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} \tag{2.44}$$

stąd wynika, że statystyki:  $\hat{\rho}(1), \hat{\rho}(2), \dots, \hat{\rho}(h)$  są w przybliżeniu niezależne i posiadają w przybliżeniu takie same rozkłady  $N(0, n^{-1})$ .

### Przykład (Średnia ruchoma rzędu $q$ )

$$\text{Jeśli } X_t = \theta_0 Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}, \quad \text{gdzie } \{Z_t\} \sim iid(0, \sigma^2) \quad (1.14')$$

wtedy z formuły Bartlett'a (2.31), otrzymujemy:

$$w_{ii} = n^{-1}[1 + 2\rho^2(1) + \dots + 2\rho^2(q)], \quad \text{dla } j > q \quad (2.45)$$

Oznacza to, że jeśli narysujemy wykres funkcji autokorelacji  $\hat{\rho}(k)$  w próbce, w funkcji  $k$ , wtedy zbiór krytyczny dla hipotezy zerowej:  $H_0 : \rho(k) = 0$  dla  $k = 1, 2, \dots, h$  ma dla każdego  $k = 1, 2, \dots, h$  w przybliżeniu postać:

$$W = (-\infty, -u_\alpha \frac{1}{\sqrt{n}}) \cup (u_\alpha \frac{1}{\sqrt{n}}, \infty) \quad (2.46)$$

co przy poziomie istotności  $\alpha = 0,05$  daje:

$$W = (-\infty, -1,96 \frac{1}{\sqrt{n}}) \cup (1,96 \frac{1}{\sqrt{n}}, \infty) \quad (2.47)$$

Oznacza to, że jeśli wartości  $\hat{\rho}(k)$ ,  $k = 1, 2, \dots, h$  w próbce mają być statystycznie nieistotnie różne od zera, to powinny przy poziomie istotności  $\alpha$ , mieścić się w przedziale ufności:

$$\left(-u_\alpha \frac{1}{\sqrt{n}}, u_\alpha \frac{1}{\sqrt{n}}\right) \text{ dla poziomu ufności } (1 - \alpha), \quad (2.48)$$

tzn. np. dla  $\alpha = 0,05$ , w przedziale:

$$\left(-1,96 \frac{1}{\sqrt{n}}, 1,96 \frac{1}{\sqrt{n}}\right) \text{ dla } 1 - \alpha = 0,95. \quad (2.49)$$

Zależności (2.46) do (2.47) są dokładne asymptotycznie, tzn. dla  $n \rightarrow \infty$ .

Zatem jeśli narysujemy wykres funkcji autokorelacji  $\hat{\rho}(k)$  w próbce, wtedy przy założeniu prawdziwości hipotezy  $H_0 : \rho(k) = 0$ ,  $k = 1, 2, \dots, h$  i np. na poziomie istotności  $\alpha = 0,05$ , wiemy, że około 95% wartości  $\hat{\rho}(k)$  w próbce, powinno leżeć pomiędzy granicami:

$$\pm 1,96 \frac{1}{\sqrt{n}} \quad (2.50)$$

Spełnienie tego warunku oznacza, że nie mamy podstaw na poziomie istotności  $\alpha = 0,05$  do odrzucenia hipotezy zerowej o tym, że obserwacje istotnie pochodzą z procesu, w którym zmienne są niezależne i mają identyczne rozkłady, tzn. że  $\{X_t\} \sim iid(0, \sigma^2)$ , i proces jest niezależnym białym szumem.

## Rozdział 3. Predykcja modeli ARMA i ARIMA

### 3.1 Przewidywanie dla procesów stacjonarnych

Tematem tego rozdziału jest przewidywanie wartości  $\{X_t, t \geq n+1\}$  stacjonarnego procesu wykorzystując wartości tego procesu dla  $\{X_1, \dots, X_n\}$ .

Mając dowolną domkniętą podprzestrzeń  $M$  przestrzeni  $L^2(\Omega, \mathfrak{F}, P)$  definiujemy najlepsze przewidywanie dla  $X_{n+h}$  w  $M$  jako element podprzestrzeni  $M$ , który posiada najmniejszą średnio – kwadratową odległość od elementu  $X_{n+h}$ .

W Uzupełnieniu 1 wspomnieliśmy (Twierdzenie 6), że rzutowanie  $P_{M(x_1, \dots, x_n)} X_{n+h}$  jest najlepszą funkcją  $X_1, \dots, X_n$ , a  $P_{sp(x_1, \dots, x_n)} X_{n+h}$  jest najlepszą liniową kombinacją zmiennych  $1, X_1, \dots, X_n$ , która daje przewidywanie wartości zmiennych  $X_{n+h}$ .

### 3.2 Równanie predykcji

Niech  $\{X_t\}$  jest procesem stacjonarnym ze średnią  $\mu$  i funkcją autokowariancji  $\gamma(\cdot)$ . Wtedy proces  $\{Y_t\} = \{X_t - \mu\}$  jest procesem stacjonarnym ze średnią równą zero i funkcją autokowariancji  $\gamma(\cdot)$ .

Można pokazać (Przykład 3.1), że :

$$P_{sp(1, X_1, \dots, X_n)} X_{n+h} = \mu + P_{sp(Y_1, \dots, Y_n)} Y_{n+h} \quad (3.1)$$

Poniżej założymy, że  $\mu = 0$ .

Wtedy:

$$P_{sp(1, X_1, \dots, X_n)} X_{n+h} = P_{sp(X_1, \dots, X_n)} X_{n+h} \quad (3.2)$$

#### Przykład 3.1

Niech  $\{X_t\}$  jest stacjonarnym procesem ze średnią  $\mu$ .

Pokazać, że:  $P_{sp(1, X_1, \dots, X_n)} X_{n+h} = \mu + P_{sp(Y_1, \dots, Y_n)} Y_{n+h}$ , gdzie  $\{Y_t\} = \{X_t - \mu\}$ .

**Rozwiązanie:**

$$\hat{X}_{n+h} = P_{sp(1, X_1, \dots, X_n)} X_{n+h} = \alpha_0 + \sum_{i=1}^n \alpha_i X_i \quad (3.3)$$

oraz :

$$\hat{Y}_{n+h} = \sum_{i=1}^n \alpha_i Y_i = \sum_{i=1}^n \alpha_i (X_i - \mu) = \sum_{i=1}^n \alpha_i X_i - \mu \sum_{i=1}^n \alpha_i$$

korzystając z 3.3 otrzymujemy:

$$\sum_{i=1}^n \alpha_i X_i = \hat{X}_{n+h} - \alpha_0$$

po podstawieniu otrzymujemy:

$$\hat{Y}_{n+h} = \hat{X}_{n+h} - \alpha_0 - \mu \sum_{i=1}^n \alpha_i \quad (3.4)$$

gdzie wprowadziliśmy oznaczenie:

$$\hat{Y}_{n+h} = P_{sp(Y_1, \dots, Y_n)} Y_{n+h}$$

Mamy  $E(\hat{X}_{n+h}) = \mu$  podstawiając tę zależność do (3.3) otrzymujemy:

$$\mu = E(\hat{X}_{n+h}) = \alpha_0 + \sum_{i=1}^n \alpha_i E(X_i) = \alpha_0 + \sum_{i=1}^n \alpha_i \mu = \alpha_0 + \mu \sum_{i=1}^n \alpha_i$$

zatem:

$$\alpha_0 + \mu \sum_{i=1}^n \alpha_i = \mu \quad (3.5)$$

Z (3.4) i (3.5) otrzymujemy:

$$\hat{Y}_{n+h} = \hat{X}_{n+h} - \mu \quad \text{c.n.d.}$$

### 3.3 Równanie dla predykcji z jednym krokiem

Niech  $M \equiv H_n$  będzie domkniętą liniową podprzestrzenią  $\overline{sp}(X_1, \dots, X_n)$ ,  $n \geq 1$ ,

i niech  $\hat{X}_{n+1}$  oznacza przewidywanie w jednym kroku, określone następująco:

$$\hat{X}_{n+1} = \begin{cases} 0 & \text{dla } n = 0 \\ P_{H_n} X_{n+1} & \text{dla } n \geq 1 \end{cases} \quad (3.6)$$

Ponieważ  $\hat{X}_{n+1} \in H_n, n \geq 1$ , zatem:

$$\hat{X}_{n+1} = \phi_{n1}X_n + \dots + \phi_{nm}X_1 \quad (3.7)$$

gdzie  $\phi_{n1}, \dots, \phi_{nm}$  spełniają równanie predykcji (Uzupełnienie 1) dla  $x \in H_n$

(tzn.  $\langle x - \hat{x}, y \rangle = 0$ , dla każdego  $y \in M$ ), które przyjmuje teraz postać:

$$\langle X_{n+1} - \hat{X}_{n+1}, X_{n+1-j} \rangle = 0, \quad j = 1, 2, \dots, n \quad (3.8)$$

lub 
$$\langle \sum_{i=1}^n \phi_{ni}X_{n+1-i}, X_{n+1-j} \rangle = \langle X_{n+1}, X_{n+1-j} \rangle, \quad j = 1, 2, \dots, n \quad (3.9)$$

gdzie  $\hat{x} \equiv \hat{X}_{n+1} = \sum_{i=1}^n \phi_{ni}X_{n+1-i} \in H_n$ ,  $x \equiv X_{n+1} \in H_n$ ,  $y \equiv X_{n+1-j} \in H_n$ , dla  $j = 1, 2, \dots, n$ .

Zgodnie z uwagami w Uzupełnieniu 1, iloczyn wewnętrzny  $\langle X, Y \rangle$  spełnia rolę wartości oczekiwanej  $E(X, Y)$ :

$$\langle X, Y \rangle = E(X, Y) \quad (3.10)$$

która w przypadku zmiennych z  $\mu = 0$  jest równa  $Cov(X, Y)$ .

Z powodu liniowości iloczynu wewnętrznego równanie (3.9) można zapisać następująco:

$$\sum_{i=1}^n \phi_{ni}\gamma(i-j) = \gamma(j), \quad j = 1, 2, \dots, n \quad (3.11)$$

gdzie  $\gamma(i-j) \equiv \langle X_{n+1-i}, X_{n+1-j} \rangle$ .

Układ równań (3.11) można zapisać równoznacznie jako:

$$\Gamma_n \vec{\phi}_n = \vec{\gamma}_n \quad (3.12)$$

gdzie  $\Gamma_n = [\gamma(i-j)]_{i,j=1}^n$ ,  $\vec{\gamma}_n = (\gamma(1), \dots, \gamma(n))^T$  i  $\vec{\phi}_n = (\phi_{n1}, \dots, \phi_{nm})^T$ . (3.13)

### Dowód (przejścia z równania (3.11) do (3.12))

Wprowadźmy oznaczenia:

$$\vec{\phi}_n = (\phi_{n1}, \phi_{n2}, \dots, \phi_{n,n-1}, \phi_{nn})^T, \quad \vec{\gamma}_n = (\gamma(1), \dots, \gamma(n))^T \quad (3.14)$$

$$c_{ij} \equiv \gamma(i-j) = Cov(X_i, X_j) = Cov(X_{i-j}, X_{j-j}) = Cov(X_{i-j}, X_0) \quad (3.15)$$

gdzie ostatnie dwa przejścia wynikają z założenia stacjonarności procesu.

Zachodzi :

$$c_{ij} = c_{ji} \quad (3.16)$$

$$c_{nn} = \gamma(0), \quad c_{n-1,n} = c_{n,n-1} = \gamma(n - (n-1)) = \gamma(1), \dots, \text{ itd.} \quad (3.17)$$

Zatem równanie (3.11) można zapisać jako:

$$\begin{pmatrix} c_{n,n} & c_{n-1,n} & \cdots & c_{2,n} & c_{1,n} \\ c_{n,n-1} & c_{n-1,n-1} & \cdots & c_{2,n-1} & c_{1,n-1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ c_{n1} & c_{n-1,1} & \cdots & c_{2,1} & c_{1,1} \end{pmatrix} \begin{pmatrix} \phi_{n,1} \\ \phi_{n,2} \\ \vdots \\ \phi_{n,n} \end{pmatrix} = \begin{pmatrix} c_{n+1,n} \\ c_{n+1,n-1} \\ \vdots \\ c_{n+1,1} \end{pmatrix} \quad (3.18)$$

lub

$$\begin{pmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(n-2) & \gamma(n-1) \\ \gamma(1) & \gamma(0) & \cdots & \gamma(n-3) & \gamma(n-2) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma(n-1) & \gamma(n-2) & \cdots & \gamma(1) & \gamma(0) \end{pmatrix} \begin{pmatrix} \phi_{n,1} \\ \phi_{n,2} \\ \vdots \\ \phi_{n,n} \end{pmatrix} = \begin{pmatrix} \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(n) \end{pmatrix} \quad (3.19)$$

Oznaczając:  $\Gamma_n = [\gamma(i-j)]_{i,j=1}^n \in M_{n \times n}$  (3.20)

Otrzymujemy:  $\Gamma_n \vec{\phi}_n = \vec{\gamma}_n$  (3.12') c.n.d (3.21)

Z twierdzenia o rzutowaniu (Uzupełnienie 1) wynika, że równanie (3.12) ma przynajmniej jedno rozwiązanie, co jest związane z tym, że przewidywanie  $\hat{X}_{n+1}$  można wyrazić w postaci (3.7),

tzn.:  $\hat{X}_{n+1} = \sum_{i=1}^n \phi_{ni} X_{n+1-i}$ , dla pewnego  $\vec{\phi}_n \in R^n$ . (3.22)

Równania (3.7) i (3.12) :

$$\begin{cases} \hat{X}_{n+1} = \phi_{n1} X_n + \dots + \phi_{nn} X_1, & (n \geq 1) \\ \Gamma_n \vec{\phi}_n = \vec{\gamma}_n \end{cases} \quad (3.23)$$

są nazywane równaniami predykcji z krokiem równym jeden.

Chociaż może istnieć wiele rozwiązań równania (3.12) na  $\vec{\phi}$ , jednak każde z nich podstawione do równania (3.7) musi, ze względu na fakt, że  $\hat{X}_{n+1}$  jest określone jednoznacznie (Uzupełnienie 1, Twierdzenie o rzutowaniu), dać to samo przewidywanie  $\hat{X}_{n+1}$ . Ponadto, gdy macierz  $\Gamma_n$  jest nieosobliwa, wtedy istnieje dokładnie jedno rozwiązanie równania (3.12) i na odwrót.

W przypadku tym mamy:

$$\vec{\phi}_n = \Gamma^{-1} \vec{\gamma}_n \quad (3.24)$$

Warunki określone w poniższym twierdzeniu w sposób wystarczający gwarantują nieosobliwość macierzy  $\Gamma_n$  dla każdego  $n$ .

### **Twierdzenie 3.1**

Jeśli  $\gamma(0) > 0$  i  $\gamma(h) \rightarrow 0$  dla  $h \rightarrow \infty$ , (3.25)

wtedy macierz kowariancji  $\Gamma_n = [\gamma(i-j)]_{i,j=1}^n$  dla  $\vec{X}_{(n)} = (X_1, \dots, X_n)^T$  jest nieosobliwa dla każdego  $n$ .

Dowód ([1] str.167)

### **Wnioski z twierdzenia 3.1**

Przy zachowaniu warunków twierdzenia 3.1, najlepszym liniowym przewidywaniem  $\hat{X}_{n+1}$  dla  $X_{n+1}$ , wyrażonym poprzez  $X_1, \dots, X_n$  jest:

$$\hat{X}_{n+1} = \sum_{i=1}^n \phi_{ni} X_{n+1-i}, \quad n = 1, 2, \dots \quad (3.26)$$

gdzie:  $\vec{\phi}_n = (\phi_1, \dots, \phi_n)^T = \Gamma_n^{-1} \vec{\gamma}_n$  (3.24')

$$\vec{\gamma}_n = (\gamma(1), \dots, \gamma(n))^T$$

i  $\Gamma_n = [\gamma(i-j)]_{i,j=1}^n$

średnio – kwadratowy błąd przewidywania wynosi:

$$v_n \equiv E(X_{n+1} - \hat{X}_{n+1})^2 = \gamma(0) - \vec{\gamma}_n^T \Gamma_n^{-1} \vec{\gamma}_n \quad (3.27)$$



### 3.4 Równania dla przewidywania z krokiem $h$ ( $h \geq 1$ )

Najlepsze liniowe przewidywanie dla  $X_{n+h}$  wyrażone poprzez  $X_1, \dots, X_n$  dla  $h \geq 1$  można znaleźć w sposób analogiczny do  $\hat{X}_{n+1}$ .

Tak więc :

$$P_{H_n} X_{n+h} = \phi_{n1}^{(h)} X_n + \dots + \phi_{nm}^{(h)} X_1, \quad n, h \geq 1 \quad (3.28)$$

gdzie 
$$\vec{\phi}_n^{(h)} = (\phi_{n1}^{(h)}, \dots, \phi_{nm}^{(h)})^T \quad (3.29)$$

jest jakimkolwiek rozwiązaniem (jedynym gdy  $\Gamma_n$  jest nie osobliwa) równania:

$$\Gamma_n \vec{\phi}_n^{(h)} = \vec{\gamma}_n^{(h)} \quad (3.30)$$

gdzie

$$\vec{\gamma}_n^{(h)} = (\gamma(h), \gamma(h+1), \dots, \gamma(n+h-1))^T \quad (3.31)$$

Istotnie:

$\phi_{n1}^{(h)}, \dots, \phi_{nm}^{(h)}$  spełniają równanie predykcji (Uzupełnienie 1), tzn.:

$$\left\langle \sum_{i=1}^n \phi_{ni}^{(h)} X_{n+1-i}, X_{n+1-j} \right\rangle = \left\langle X_{n+h}, X_{n+1-j} \right\rangle, \quad j = 1, 2, \dots, n$$

skąd otrzymujemy:

$$\sum_{i=1}^n \phi_{ni}^{(h)} \gamma(i-j) = \gamma(h+j-1), \quad j = 1, 2, \dots, n \quad (3.30')$$

tzn. zależność (3.30).

### 3.5 Metody rekurencyjne dla wyznaczania najlepszych liniowych przewidywań.

Poniżej przedstawimy dwa rekurencyjne algorytmy służące do określania przewidywań jednego kroku  $\hat{X}_{n+1}$ ,  $n \geq 1$ , zdefiniowanych równaniem (3.6). Następnie pokażemy jak można je wykorzystać do wyznaczenia przewidywań  $h$  – tego kroku  $P_{H_n} X_{n+h}$ ,  $h \geq 1$ .

Procedury rekurencyjne mają dlatego tak wielkie praktyczne zastosowanie, bowiem bezpośrednio wyznaczanie  $P_{H_n} X_{n+h}$  z równań (3.28) i (3.30) wymagałoby rozwiązania układu równań liniowych dla dużych  $n$ . Na dodatek za każdym razem gdy wzrasta liczba obserwacji, cała procedura musiałaby być powtarzana od nowa. Algorytmy przedstawione poniżej pozwalają na wyznaczenie najlepszych przewidywań bez konieczności liczenia macierzy odwrotnych.

Ponadto pozwalają one na wyznaczanie w oparciu o przewidywania otrzymane dla  $n$  – obserwacji, na wyznaczenie przewidywań opartych na  $n+1$  – obserwacjach,  $n=1,2,\dots$  .

W dalszej części pokażemy również, jak drugi z przedstawionych algorytmów, ułatwia obliczenie dokładnej wiarygodności dla  $\{X_1, \dots, X_n\}$  wtedy, gdy proces  $\{X_t\}$  jest gaussowski.

### Przewidywanie rekurencyjne z wykorzystaniem algorytmu Durbin'a – Levinson'a

Ponieważ zgodnie z (3.6):  $\hat{X}_{n+1} = P_{H_n} X_{n+1} \in H_n, n \geq 1$

zatem  $\hat{X}_{n+1}$  można wyrazić w postaci:

$$\hat{X}_{n+1} = \phi_{n1} X_n + \dots + \phi_{nn} X_1, n \geq 1 \quad (3.32)$$

średnio – kwadratowy błąd przewidywania, oznaczony  $v_n$ , wynosi:

$$v_n = E(X_{n+1} - \hat{X}_{n+1})^2, n \geq 1 \quad (3.33)$$

a z (3.6) dla  $n=0$  i faktu, że  $E(X_i) = 0$  otrzymujemy, że:

$$v_0 = \gamma(0) \quad (3.34)$$

Poniższy algorytm Durbin'a – Levinson'a jest rekurencyjnym schematem dla wyznaczenia  $\vec{\phi}_n = (\phi_{n1}, \dots, \phi_{nn})^T$  oraz  $v_n$  dla  $n = 1, 2, \dots$

#### **Twierdzenie 3.2 (Algorytm Durbin'a – Levinson'a)**

Jeśli  $\{X_t\}$  jest stacjonarnym procesem ze średnią równą zero i funkcją autokowariancji  $\gamma(\cdot)$  taką, że  $\gamma(0) > 0$  i  $\gamma(h) \rightarrow 0$  dla  $h \rightarrow \infty$ , wtedy współczynniki  $\phi_{nj}$  i średnio - kwadratowe błędy  $v_n$  zdefiniowane równaniami (3.33) i (3.34) spełniają zależności:

$$\phi_{11} = \frac{\gamma(1)}{\gamma(0)}, v_0 = \gamma(0) \quad (3.35)$$

$$\phi_{nn} = \left[ \gamma(n) - \sum_{j=1}^{n-1} \phi_{n-1,j} \gamma(n-j) \right] v_{n-1}^{-1} \quad (3.36)$$

$$\begin{bmatrix} \phi_{n1} \\ \vdots \\ \phi_{n,n-1} \end{bmatrix} = \begin{bmatrix} \phi_{n-1,n} \\ \vdots \\ \phi_{n-1,n-1} \end{bmatrix} - \phi_{nn} \begin{bmatrix} \phi_{n-1,n-1} \\ \vdots \\ \phi_{n-1,1} \end{bmatrix} \quad (3.37)$$

$$v_n = v_{n-1} [1 - \phi_{nn}^2] \quad (3.38)$$

$$\phi_{nj} = \phi_{n-1,j} - \phi_{nn} \phi_{n-1,n-j} \quad (3.39)$$

W Rozdziale 1 zostały podane dwie definicje funkcji częściowej autokorelacji dla procesu  $\{X_t\}$  przy lag  $n$ , a mianowicie:

$$\alpha(n) = \text{Corr}(X_{n+1} - P_{sp\{X_2, \dots, X_n\}} X_{n+1}, X_1 - P_{sp\{X_2, \dots, X_n\}} X_1) \quad (3.40)$$

oraz: 
$$\alpha(n) = \phi_{nn} \quad (3.41)$$

Dowód [1]

Centralną sprawą w dowodzie tego twierdzenia jest rozkład przestrzeni  $H_n$  na dwie ortogonalne podprzestrzenie  $H_1 = \overline{sp\{X_2, \dots, X_n\}}$  i  $H_2 = \overline{sp\{X_1 - P_{H_1} X_1\}}$  gdzie  $P_{H_1}$  jest rzutowaniem na  $H_1$ .

**Uwaga. O równości zależności (3.40) i (3.41)**

Przy zachowaniu założeń Twierdzenia 3.2, zachodzi związek:

$$\phi_{nn} = \text{Corr}(X_{n+1} - P_{sp\{X_2, \dots, X_n\}} X_{n+1}, X_1 - P_{sp\{X_2, \dots, X_n\}} X_1) \quad (3.42)$$

patrz rozdział pierwszy (funkcja PACF, Definicja 1.6 , 1.7).

### **Przewidywanie rekurencyjne z wykorzystaniem algorytmu innowacyjnego**

Centralną sprawą w dowodzie twierdzenia 3.2 był rozkład przestrzeni  $H_n$  na dwie ortogonalne podprzestrzenie  $H_1$  i  $H_2$ . Druga procedura rekurencyjna przedstawiona poniżej w twierdzeniu 3.4, jest związana z rozkładem  $H_n$  na  $n$  ortogonalnych podprzestrzeni w oparciu o procedurę Gram'a - Schmidt'a.

Poniżej przedstawione twierdzenie 3.4 ma szerszy zakres zastosowań niż poprzednio przedstawione twierdzenie 3.2. Przyczyną jest fakt, że dopuszcza ono procesy  $\{X_t\}$ , które mogą być procesami niestacjonarnymi ze średnią zero i funkcją autokowariancji:

$$\kappa(i, j) = \langle X_i, X_j \rangle = E(X_i X_j) \quad (3.43)$$

Tak jak poprzednio, definiujemy  $H_n = \overline{sp\{X_1, \dots, X_n\}}$  i  $\hat{X}_{n+1}$  jak we wzorze (3.6),

a 
$$v_n = \left\| X_{n+1} - \hat{X}_{n+1} \right\|^2. \quad (3.44)$$

Przyjmując  $\hat{X}_1 := 0 \quad (3.45)$

i zauważając, że zgodnie z (3.7)

$$\hat{X}_{n+1} = \phi_{n1} X_n + \dots + \phi_{nm} X_1, \quad n \geq 1 \quad (3.46)$$

co oznacza, że  $\hat{X}_{n+1}$  jest liniową kombinacją  $\{X_1, \dots, X_n\}$ ,

mamy, że  $H_n = \overline{sp}\{X_1, \dots, X_n\}$  można również rozpisać na  $\{X_1 - \hat{X}_1, X_2 - \hat{X}_2, \dots, X_n - \hat{X}_n\}$ ,

$$\text{tzn.} \quad H_n = \overline{sp}\{X_1 - \hat{X}_1, X_2 - \hat{X}_2, \dots, X_n - \hat{X}_n\}, \quad n \geq 1 \quad (3.47)$$

$$\text{zatem:} \quad \hat{X}_{n+1} = \sum_{j=1}^n \theta_{nj} (X_{n+1-j} - \hat{X}_{n+1-j}). \quad (3.48)$$

Poniżej podamy rekurencyjną procedurę wyznaczania  $\{\theta_{nj}, j = 1, \dots, n; v_n\}$ , gdzie  $n \geq 1$ .

### Twierdzenie 3.4 (Algorytm innowacyjny)

Jeśli  $\{X_i\}$  ma średnią zero oraz  $E(X_i X_j) = \kappa(i, j)$ , gdzie macierz  $[\kappa(i, j)]_{i,j=1}^n$  jest nieosobliwa dla każdego  $n = 1, 2, \dots$ , wtedy przewidywanie jednego kroku  $\hat{X}_{n+1}$ ,  $n \geq 0$  oraz ich średnio – kwadratowe błędy  $v_n$ ,  $n \geq 1$ , są zadane związkami :

$$\hat{X}_{n+1} = \begin{cases} 0 & n = 0 \\ \sum_{j=1}^n \theta_{nj} (X_{n+1-j} - \hat{X}_{n+1-j}) & n \geq 1 \end{cases} \quad (3.49)$$

oraz:

$$\begin{cases} v_0 = \kappa(1,1) \\ \theta_{n,n-k} = v_k^{-1} \left( \kappa(n+1, k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j} \theta_{n,n-j} v_j \right) = 0, 1, \dots, n-1 \\ v_n = \kappa(n+1, n+1) - \sum_{j=0}^{n-1} \theta_{n,n-j}^2 v_j \end{cases} \quad (3.50)$$

gdzie dla  $k=0$  za drugi składnik w zależności na  $\theta_{n,n-k}$  wstawiamy zero. Rekurencyjnie rozwiązanie równań (3.50) jest proste i przebiegające w kolejności:

$$v_0; \quad \theta_{11}, v_1; \quad \theta_{22}, \theta_{21}, v_2; \quad \theta_{33}, \theta_{32}, \theta_{31}, v_3; \dots \quad (3.51)$$

### Dowód:

Zbiór  $\{X_1 - \hat{X}_1, X_2 - \hat{X}_2, \dots, X_n - \hat{X}_n\}$  jest ortogonalny. Wynika to z faktu, że  $(X_i - \hat{X}_i) \in H_{j-1}$  dla  $i < j$  oraz z tego, że zgodnie z definicją  $\hat{X}_j$ ,  $(X_j - \hat{X}_j) \perp H_{j-1}$ .

Licząc iloczyn wewnętrzny obydwu stron w (3.49) z  $(X_{k+1} - \hat{X}_{k+1})$  dla  $0 \leq k < n$ , otrzymujemy:

$$\begin{aligned} \langle \hat{X}_{n+1}, X_{k+1} - \hat{X}_{k+1} \rangle &= \langle \sum_{j=0}^n \theta_{nj} (X_{n+1-j} - \hat{X}_{n+1-j}), X_{k+1} - \hat{X}_{k+1} \rangle = \\ &= \sum_{j=0}^n \theta_{nj} \langle X_{n+1-j} - \hat{X}_{n+1-j}, X_{k+1} - \hat{X}_{k+1} \rangle = \sum_{j=0}^n \theta_{nj} \delta_{n+1-j, k+1} \langle X_{n+1-j} - \hat{X}_{n+1-j}, X_{k+1} - \hat{X}_{k+1} \rangle = \\ &= \dots \{ \delta_{n+1-j, k+1} = \delta_{j, n-k} \} \dots = \theta_{n, n-k} \langle X_{n+1-(n-k)} - \hat{X}_{n+1-(n-k)}, X_{k+1} - \hat{X}_{k+1} \rangle = \\ &= \theta_{n, n-k} \langle X_{k+1} - \hat{X}_{k+1}, X_{k+1} - \hat{X}_{k+1} \rangle = \theta_{n, n-k} v_k \end{aligned}$$

$$\text{tzn. :} \quad \langle \hat{X}_{k+1}, X_{k+1} - \hat{X}_{k+1} \rangle = \theta_{n, n-k} v_k \quad (3.52)$$

$$\text{gdzie} \quad v_k = \langle X_{k+1} - \hat{X}_{k+1}, X_{k+1} - \hat{X}_{k+1} \rangle = \|X_{k+1} - \hat{X}_{k+1}\|^2 \quad (3.53)$$

jest średnio – kwadratowym błędem.

Dla  $n = 0$ ,  $k = 0$  otrzymujemy:

$$\langle \hat{X}_{n+1}, X_{k+1} - \hat{X}_{k+1} \rangle = \langle \hat{X}_1, X_1 - \hat{X}_1 \rangle = 0 \quad (3.54)$$

Ponieważ dla  $k = 0, \dots, n-1$  zachodzi  $(X_{n+1} - \hat{X}_{n+1}) \perp (X_{k+1} - \hat{X}_{k+1})$  zatem :

$$\begin{aligned} \langle X_{n+1} - \hat{X}_{n+1}, X_{k+1} - \hat{X}_{k+1} \rangle &= \langle X_{n+1}, X_{k+1} - \hat{X}_{k+1} \rangle - \langle \hat{X}_{n+1}, X_{k+1} - \hat{X}_{k+1} \rangle = \\ &= \langle X_{n+1}, X_{k+1} - \hat{X}_{k+1} \rangle - \theta_{n, n-k} v_k = 0 \end{aligned} \quad (3.55)$$

Otrzymujemy więc, że:

$$\langle X_{n+1}, X_{k+1} - \hat{X}_{k+1} \rangle = \theta_{n, n-k} v_k \quad (3.56)$$

$$\text{lub} \quad \theta_{n, n-k} = v_k^{-1} \langle X_{n+1}, X_{k+1} - \hat{X}_{k+1} \rangle \quad (3.57)$$

Wstawiając do powyższej zależności  $\hat{X}_{k+1}$  zapisane zgodnie z (5.49) jako:

$$\hat{X}_{k+1} = \begin{cases} 0 & k = 0 \\ \sum_{j=1}^k \theta_{kj} (X_{k+1-j} - \hat{X}_{k+1-j}) & k \geq 1 \end{cases}$$

otrzymujemy dla  $k = 1, \dots, n-1$ :

$$\begin{aligned} \theta_{n,n-k} &= v_k^{-1} \langle X_{n+1}, X_{k+1} - \sum_{j=1}^k \theta_{kj} (X_{k+1-j} - \hat{X}_{k+1-j}) \rangle = \\ &= v_k^{-1} \left[ \langle X_{n+1}, X_{k+1} \rangle - \sum_{j=1}^k \theta_{kj} \langle X_{n+1}, X_{k+1-j} - \hat{X}_{k+1-j} \rangle \right] = \end{aligned}$$

podstawiając:

$$j' = k - j, \quad j = k - j'$$

$$\text{otrzymujemy: } = v_k^{-1} \left[ \kappa(n+1, k+1) - \sum_{j'=k-1}^0 \theta_{k,k-j'} \langle X_{n+1}, X_{j'+1} - \hat{X}_{j'+1} \rangle \right] =$$

natomiast przyjmując:

$$j' = j$$

$$\begin{aligned} \text{otrzymujemy: } &= v_k^{-1} \left[ \kappa(n+1, k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j} \langle X_{n+1}, X_{j+1} - \hat{X}_{j+1} \rangle \right] = \\ &= v_k^{-1} \left[ \kappa(n+1, k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j} v_j \theta_{n,n-j} \right] \end{aligned} \quad (3.58)$$

gdzie w ostatnim przejściu skorzystano z zależności (3.56) otrzymując środkową zależność w (3.50), a  $\kappa(n+1, k+1) \equiv \langle X_{n+1}, X_{k+1} \rangle$ .

Dla  $k=0$  otrzymujemy z samego określenia  $\hat{X}_{n+1}$  w (3.49), że  $\theta_{0,0} = \theta_{0,-1} = 0$  oraz również  $\theta_{1,0} = 0$ . (3.59)

Z twierdzenia o rzutowaniu (Uzupełnienie 1), mamy, dla dowolnego  $x \in H$  oraz domkniętej podprzestrzeni  $M \subset H$  i rzutowania  $P_M$  na podprzestrzeń  $M$ :

$$\|x\|^2 = \|P_M x\|^2 + \|(I - P_M)x\|^2 \quad (3.60)$$

$$\text{przyjmując } x \equiv X_{n+1}, \quad M \equiv H_n, \quad P_M x \equiv P_{H_n} X_{n+1} = \hat{X}_{n+1} \quad (3.61)$$

$$\text{otrzymujemy: } \|X_{n+1}\|^2 = \|\hat{X}_{n+1}\|^2 + \|(I - P_{H_n})X_{n+1}\|^2 = \|\hat{X}_{n+1}\|^2 + \|X_{n+1} - \hat{X}_{n+1}\|^2 \quad (3.62)$$

Pokażemy teraz ostatnią zależność w (3.50). Z (3.62) mamy:

$$v_n = \|X_{n+1} - \hat{X}_{n+1}\|^2 = \|X_{n+1}\|^2 - \|\hat{X}_{n+1}\|^2 \quad (3.63)$$

$$\text{gdzie } \|X_{n+1}\|^2 \equiv \kappa(n+1, n+1). \quad (3.64)$$

Wyznaczmy  $\|\hat{X}_{n+1}\|^2$ . Z zależności (3.48), mamy:

dla  $n = 0$

$$\|\hat{X}_{n+1}\|^2 = \|\hat{X}_1\|^2 = \langle \hat{X}_1, \hat{X}_1 \rangle = \langle 0, 0 \rangle = 0 \quad (3.65)$$

dla  $n \geq 1$

$$\begin{aligned} \|\hat{X}_{n+1}\|^2 &= \langle \hat{X}_{n+1}, \hat{X}_{n+1} \rangle = \sum_{j=1}^n \sum_{j'=1}^n \theta_{nj} \theta_{nj'} \langle X_{n+1-j} - \hat{X}_{n+1-j}, X_{n+1-j'} - \hat{X}_{n+1-j'} \rangle = \\ &= \sum_{j=1}^n \sum_{j'=1}^n \theta_{nj} \theta_{nj'} \delta_{jj'} \langle X_{n+1-j} - \hat{X}_{n+1-j}, X_{n+1-j'} - \hat{X}_{n+1-j'} \rangle = \sum_{j=1}^n \theta_{nj}^2 v_{n-j} = \end{aligned}$$

podstawiając:  $k = n - j, \quad j = n - k$

$$\text{otrzymujemy: } = \sum_{k=n-1}^0 \theta_{n,n-k}^2 v_k = \sum_{k=0}^{n-1} \theta_{n,n-k}^2 v_k \quad (3.66)$$

Zatem ostatecznie:

$$v_n = \|X_{n+1}\|^2 - \|\hat{X}_{n+1}\|^2 = \kappa(n+1, n+1) - \sum_{k=0}^{n-1} \theta_{n,n-k}^2 v_k \quad (3.67)$$

gdzie dla  $n = 0$  drugi składnik jest równy tożsamościowo zero. To kończy dowód.

### 3.6 Obliczenia rekurencyjne dla przewidywania $h$ – tego kroku, $h \geq 1$

Wprowadzimy oznaczenie  $P_n$  dla operatora rzutowego  $P_{H_n}$ . Wtedy przewidywanie

$h$  – tego kroku, czyli  $P_n X_{n+h}$  można wyznaczyć w oparciu o twierdzenie (3.4).

Ponieważ dla  $h \geq 1, n+h-1 \geq n$ , więc dla każdego  $x \in H$ ,  $P_n P_{n+h-1} x = P_n x$ , (3.68)

zatem:  $P_n X_{n+h} = P_n P_{n+h-1} X_{n+h} = P_n \hat{X}_{n+h}$  (3.69)

gdzie  $P_{n+h-1}$  jest rzutowaniem na  $(n+h-1)$  - wymiarową podprzestrzeń dając przewidywanie

$\hat{X}_{n+h}$ .

Korzystając z (3.48), mamy:

$$\hat{X}_{n+h} = \sum_{j=1}^{n+h-1} \theta_{n+h-1,j} (X_{n+h-j} - \hat{X}_{n+h-j}) \quad n+h-1 \geq 1 \quad . \quad (3.70)$$

Ponieważ dla  $j < h$   $(X_{n+h-j} - \hat{X}_{n+h-j}) \perp H_n$ , zatem dla  $j < h$ ,

$$P_n (X_{n+h-j} - \hat{X}_{n+h-j}) = 0 \quad (3.71)$$

$$P_n \hat{X}_{n+h} = \sum_{j=1}^{n+h-1} \theta_{n+h-1,j} P_n (X_{n+h-j} - \hat{X}_{n+h-j}) = \sum_{j=h}^{n+h-1} \theta_{n+h-1,j} P_n (X_{n+h-j} - \hat{X}_{n+h-j}) \quad (3.72)$$

Ponieważ dla  $j \geq h$ ,  $n+h-j \leq n$ , więc  $X_{n+h-j} \in H_n$  oraz  $\hat{X}_{n+h-j} \in H_n$ , zatem:

$$P_n X_{n+h-j} = X_{n+h-j} \quad (3.73)$$

$$P_n \hat{X}_{n+h-j} = \hat{X}_{n+h-j} \quad (3.74)$$

Zbierając związki (3.69) – (3.74), otrzymujemy:

$$P_n X_{n+h} = P_n P_{n+h-1} X_{n+h} = P_n \hat{X}_{n+h} = \sum_{j=h}^{n+h-1} \theta_{n+h-1,j} (X_{n+h-j} - \hat{X}_{n+h-j}) \quad (3.75)$$

gdzie  $\theta_{nj}$  wyznacza się z (3.50),

Natomiast średnio – kwadratowy błąd  $v_{n+h}$  wynosi:

$$\begin{aligned} v_{n+h} &= E\{(X_{n+h} - P_n X_{n+h})^2\} = \|X_{n+h} - P_n X_{n+h}\|^2 = \|(I - P_n)X_{n+h}\|^2 = \|X_{n+h}\|^2 - \|P_n X_{n+h}\|^2 = \\ &= \kappa(n+h, n+h) - (P_n X_{n+h}, P_n X_{n+h}) = \\ &= \kappa(n+h, n+h) - \sum_{j=h}^{n+h-1} \sum_{j'=h}^{n+h-1} \theta_{n+h-1,j} \theta_{n+h-1,j'} (X_{n+h-j} - \hat{X}_{n+h-j}, X_{n+h-j'} - \hat{X}_{n+h-j'}) = \\ &= \kappa(n+h, n+h) - \sum_{j=h}^{n+h-1} \sum_{j'=h}^{n+h-1} \theta_{n+h-1,j} \theta_{n+h-1,j'} \delta_{j,j'} (X_{n+h-j} - \hat{X}_{n+h-j}, X_{n+h-j'} - \hat{X}_{n+h-j'}) = \\ &= \kappa(n+h, n+h) - \sum_{j=h}^{n+h-1} \theta_{n+h-1,j}^2 v_{n+h-j-1} \end{aligned} \quad (3.76)$$

$$\text{gdzie} \quad v_{n+h-j-1} = \|X_{n+h-j} - \hat{X}_{n+h-j}\|^2 \quad (3.77)$$

Otrzymaliśmy więc, że średnio – kwadratowy błąd wynosi:

$$v_{n+h} = E\{(X_{n+h} - P_n X_{n+h})^2\} = \kappa(n+h, n+h) - \sum_{j=h}^{n+h-1} \theta_{n+h-1,j}^2 v_{n+h-j-1} \quad (3.78)$$



### 3.7 Przewidywanie rekurencyjne dla procesów ARMA( $p, q$ ) z jednym krokiem

Obecnie przedstawimy metodę, która umożliwi nam szybsze otrzymanie wyrażen na predykcję. Bazować będzie ona na przekształceniach opisanych poniżej.

Twierdzenie 3.4 można zastosować do przewidywań dotyczących przyczynowych procesów ARMA. (1.10):  $\phi(B)X_t = \theta(B)Z_t$ ,  $Z_t \sim WN(0, \sigma^2)$  (3.79)

Gdzie zgodnie z (1.11 – 1.13) :

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p, \quad (3.80)$$

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q, \quad (3.81)$$

Okazuje się, że można uzyskać znaczne uproszczenie rachunków wtedy, gdy zamiast analizować bezpośrednio proces  $\{X_t\}$  przejdzie się do procesu przetransformowanego  $\{W_t\}$  [1], określonego w następujący sposób:

$$W_t = \sigma^{-1} X_t, \quad t = 1, 2, \dots, m \quad (3.82)$$

$$W_t = \sigma^{-1} \phi(B) X_t, \quad t > m \quad (3.83)$$

$$\text{gdzie } m = \max(p, q). \quad (3.84)$$

Wprowadźmy oznaczenie  $\theta_0 = 1$  i załóżmy, że  $p \geq 1$  i  $q \geq 1$ .

Według potrzeby, w konkretnej analizie niektóre  $\phi_i$  oraz  $\theta_i$  mogą być równe zero, np. może się zdarzyć, że  $\phi_1 = 0$ , itd. Ponieważ zmienne  $W_t$  są liniowymi kombinacjami zmiennych  $X_t$ , przy czym w rozwinięciu zmiennej  $W_t$  zawsze pojawia się zmienna  $X_t$  z największym indeksem czasowym, tzn. zmienna  $X_t$  zatem, zgodnie z notacją wprowadzoną wcześniej:

$$H_n = \overline{sp}\{X_1, \dots, X_n\} = \overline{sp}\{W_1, \dots, W_n\}, \quad n \geq 1. \quad (3.85)$$

Dla  $n \geq 1$ ,  $\hat{X}_{n+1}$  i  $\hat{W}_{n+1}$  oznaczają rzutowania odpowiednio  $X_{n+1}$  i  $W_{n+1}$  na przestrzeń  $H_n$ .

Tak jak w (3.6) przyjmijmy, że z definicji:

$$\hat{X}_1 = \hat{W}_1 = 0 \quad (3.86)$$

Funkcję autokowariancji  $\gamma_x(\cdot)$  dla procesu  $\{X_t\}$  można wyznaczyć metodą 1 lub 2 podaną w Rozdziale 1.

Mając  $\gamma_x(\cdot)$  możemy wyrazić autokowariancję

$$\kappa(i, j) = E(W_i W_j) \quad (3.86)$$

w następujący sposób [1]:

$$\kappa(i, j) = \begin{cases} \sigma^{-2} \gamma_x(i-j) & 1 \leq i, j \leq m \\ \sigma^{-2} [\gamma_x(i-j) - \sum_{r=1}^p \phi_r \gamma_x(r - |i-j|)] & \text{dla } \min(i, j) \leq m < \max(i, j) \leq 2m \\ \sum_{r=0}^q \theta_r \theta_{r+|i-j|} & \min(i, j) > m \\ 0 & \text{dla pozostałych } i, j \end{cases} \quad (3.87)$$

gdzie  $\theta_j = 0$  dla  $j > q$ . (3.88)

Możemy teraz zastosować Twierdzenie 3.4 innowacyjnego algorytmu do procesu  $\{W_t\}$ .

Otrzymujemy:

$$\begin{cases} \hat{W}_{n+1} = \sum_{j=1}^n \theta_{nj} (W_{n+1-j} - \hat{W}_{n+1-j}) & \text{dla } 1 \leq n < m \\ \hat{W}_{n+1} = \sum_{j=1}^q \theta_{nj} (W_{n+1-j} - \hat{W}_{n+1-j}) & \text{dla } n \geq m \end{cases} \quad (3.88)$$

gdzie współczynniki  $\theta_{nj}$  oraz średnio – kwadratowe błędy

$$r_n \equiv E(W_{n+1} - \hat{W}_{n+1})^2 \quad (3.89)$$

znajdujemy rekurencyjnie z zależności (3.50) gdzie  $\kappa(i, j)$  są wyznaczone zgodnie z (3.87) na podstawie wcześniej znalezionej  $\gamma$  metodą 1 lub 2 (Rozdział 1).

W celu wyznaczenia przewidywania  $\hat{X}_n$  z  $\hat{W}_n$  dokonujemy rzutowania obydwu stron (3.82), (3.83) na podprzestrzeń  $H_{t-1}$ :

$$P_{H_{t-1}} W_t = \sigma^{-1} P_{H_{t-1}} X_t, \quad t = 1, 2, \dots, m \quad (3.90)$$

$$P_{H_{t-1}} W_t = \sigma^{-1} P_{H_{t-1}} (X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p}), \quad t > m \quad (3.91)$$

skąd otrzymujemy po skorzystaniu z  $P_{H_{t-1}} W_t = \hat{W}_t$ ,  $P_{H_{t-1}} X_t = \hat{X}_t$ , (3.92)

oraz  $P_{H_{t-1}} X_{t-i} = \hat{X}_{t-i}$ , dla  $i \geq 1$

związki:

$$\hat{W}_t = \sigma^{-1} \hat{X}_t, \quad t = 1, 2, \dots, m \quad (3.93)$$

$$\hat{W}_t = \sigma^{-1} (\hat{X}_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p}), \quad t > m \quad (3.94)$$

Powyższe związki, łącznie (3.82), (3.83) dają :

$$W_t - \hat{W}_t = \sigma^{-1} (X_t - \hat{X}_t) \quad \text{dla } t = 1, 2, \dots, m \quad (3.95)$$

natomiast dla  $t > m$

$$\begin{aligned} W_t - \hat{W}_t &= \sigma^{-1} (X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p}) - \sigma^{-1} (\hat{X}_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p}) \\ W_t - \hat{W}_t &= \sigma^{-1} (X_t - \hat{X}_t) \end{aligned} \quad (3.96)$$

Zatem otrzymujemy, że:

$$X_t - \hat{X}_{t_t} = \sigma (W_t - \hat{W}_t) \quad \text{dla } t \geq 1 \quad (3.97)$$

Zastępując  $(W_t - \hat{W}_t)$  przez  $\sigma^{-1} (X_t - \hat{X}_t)$  w równaniach (3.88) i korzystając z (3.93) i (3.94) otrzymujemy:

$$\begin{cases} \sigma^{-1} \hat{X}_{n+1} = \sum_{j=1}^n \theta_{nj} \sigma^{-1} (X_{n+1-j} - \hat{X}_{n+1-j}) & \text{dla } 1 \leq n < m \\ \sigma^{-1} (X_{n+1} - \phi_1 X_{n+1-1} - \dots - \phi_p X_{n+1-p}) = \sum_{j=1}^q \theta_{nj} \sigma^{-1} (X_{n+1-j} - \hat{X}_{n+1-j}) & \text{dla } n > m \end{cases} \quad (3.98)$$

stąd

$$\begin{cases} \hat{X}_{n+1} = \sum_{j=1}^n \theta_{nj} (X_{n+1-j} - \hat{X}_{n+1-j}) & \text{dla } 1 \leq n < m \\ \hat{X}_{n+1} = \phi_1 X_{n+1-1} + \dots + \phi_p X_{n+1-p} + \sum_{j=1}^q \theta_{nj} (X_{n+1-j} - \hat{X}_{n+1-j}) & \text{dla } n > m \end{cases} \quad (3.99)$$

Otrzymujemy również średnio – kwadratowy błąd:

$$E(X_{n+1} - \hat{X}_{n+1})^2 = \sigma^2 E(W_{n+1} - \hat{W}_{n+1})^2 \equiv \sigma^2 r_n \quad (3.100)$$

gdzie  $\theta_{nj}$  i  $r_n$  znajdujemy z (3.50) dla  $\kappa$  wyznaczonych zgodnie z (3.87).

Równania (3.99) pozwalają na wyznaczenie metodą rekurencyjną przewidywań jednego kroku.

**Uwaga:** Kowariancje  $\kappa(i, j)$  przetransformowanego procesu  $\{W_t\}$  zależą jedynie od  $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$  a nie zależą od  $\sigma^2$ . To samo dotyczy zatem  $\theta_{nj}$  i  $r_n$  [1].

**Uwaga:** Można pokazać, że jeśli proces  $\{X_t\}$  jest odwracalny, wtedy dla  $n \rightarrow \infty$  mamy:  $r_n \rightarrow 1$  i  $\theta_{nj} \rightarrow \theta_j, j=1, \dots, q$  [1].

### 3.8 Przewidywanie rekurencyjne dla procesów ARMA( $p, q$ ) $h$ – tego kroku, $h \geq 1$

Wykorzystując notacje z metod rekurencyjnych dla predykcji  $h$  – tego kroku,  $h \geq 1$  z równania (3.75) otrzymujemy:

$$P_n W_{n+h} = \sum_{j=h}^{n+h-1} \theta_{n+h-1,j} (W_{n+h-j} - \hat{W}_{n+h-j}) = \sigma^{-1} \sum_{j=h}^{n+h-1} \theta_{n+h-1,j} (X_{n+h-j} - \hat{X}_{n+h-j})$$

Stosując  $P_n$  do obu stron równań (3.82) i (3.83) otrzymujemy:

$$P_n X_{n+h} = \begin{cases} \sum_{j=1}^{n+h-1} \theta_{n+h-1,j} (X_{n+h-j} - \hat{X}_{n+h-j}) & \text{dla } 1 \leq h \leq m-n \\ \sum_{i=1}^p \phi_i P_n X_{n+h-i} + \sum_{h \leq j \leq q} \theta_{n+h-1,j} (X_{n+h-j} - \hat{X}_{n+h-j}) & \text{dla } h > m-n \end{cases} \quad (3.101)$$

Jak już wyznaczyliśmy przewidywania  $\hat{X}_1, \dots, \hat{X}_n$  z zależności (3.99), wtedy zwykle rachunki z ustaloną wartością  $n$ , pozwalają nam z zależności (3.101), wyznaczyć rekurencyjnie przewidywania  $P_n X_{n+1}, P_n X_{n+2}, P_n X_{n+3}, \dots$ .

Przyjmując, że  $n > m$  (co jest normalne w praktycznych problemach predykcji), dla  $h \geq 1$  otrzymujemy:

$$P_n X_{n+h} = \sum_{i=1}^p \phi_i P_n X_{n+h-i} + \sum_{h \leq j \leq q} \theta_{n+h-1,j} (X_{n+h-j} - \hat{X}_{n+h-j}) \quad (3.102)$$

gdzie druga suma jest równa zero dla  $h > q$ .

Jeśli  $X_{n+h}$  zapiszemy jako:  $\hat{X}_{n+h} + (X_{n+h} - \hat{X}_{n+h})$ , mamy również (po skorzystaniu z (3.99)) równanie:

$$X_{n+h} = \sum_{i=1}^p \phi_i X_{n+h-i} + \sum_{j=0}^q \theta_{n+h-1,j} (X_{n+h-j} - \hat{X}_{n+h-j}) \quad (3.103)$$

gdzie dla wszystkich  $n$ ,  $\theta_{n0} := 1$ .

Odejmując (3.102) od (3.103) otrzymujemy:

$$X_{n+h} - P_n X_{n+h} - \sum_{i=1}^p \phi_i (X_{n+h-i} - P_n X_{n+h-i}) = \sum_{j=0}^{h-1} \theta_{n+h-1,j} (X_{n+h-j} - \hat{X}_{n+h-j}) \quad (3.104)$$

i w związku z tym: 
$$\bar{\Phi} \begin{bmatrix} X_{n+1} - P_n X_{n+1} \\ \vdots \\ X_{n+h} - P_n X_{n+h} \end{bmatrix} = \bar{\Theta} \begin{bmatrix} X_{n+1} - \hat{X}_{n+1} \\ \vdots \\ X_{n+h} - \hat{X}_{n+h} \end{bmatrix} \quad (3.105)$$

gdzie  $\bar{\Phi}$  oraz  $\bar{\Theta}$  są dolnymi macierzami trójkątnymi,

$$\bar{\Phi} = -[\phi_{i-j}]_{i,j=1}^h \quad (\phi_0 := -1, \phi_j := 0 \text{ jeśli: } j > p \text{ lub } j < 0) \quad (3.106)$$

$$\bar{\Theta} = [\theta_{n+i-1,i-j}]_{i,j=1}^h \quad (\theta_{n0} := 1, \theta_{nj} := 1 \text{ jeśli: } j > q \text{ lub } j < 0) \quad (3.107)$$

Z równania (3.105) można znaleźć macierz kowariancji wektora:

$$(X_{n+1} - P_n X_{n+1}, \dots, X_{n+h} - P_n X_{n+h})^T$$

błędów predykcji, jako równą:

$$C = \Phi^{-1} \Theta V \Theta^T (\Phi^T)^{-1} \quad (3.108)$$

gdzie  $V = \text{diag}(v_n, v_{n+1}, \dots, v_{n+h-1})$ .

Można pokazać [1]), że :

$$\Phi^{-1} = [\chi_{i-j}]_{i,j=1}^h \quad (\chi_0 := 1, \chi_j := 0 \text{ dla } j < 0) \quad (3.109)$$

jest dolną macierzą trójkątną, której współczynniki  $\chi_j$  dla  $j \geq 1$  można łatwo wyznaczyć z rekurencyjnej zależności:

$$\chi_j = \sum_{k=1}^{\min(p,j)} \phi_k \chi_{j-k}, \quad j = 1, 2, \dots \quad (3.110)$$

.....

### Wtrącenie

Pisząc rekurencyjne relacje na współczynniki w rozwinięciu w szereg potęgowy dla  $1/\phi(z)$  (porównaj Rozdziału 1), mamy, że:

$$\sum_{j=0}^{\infty} \chi_j z^j = (1 - \phi_1 z - \dots - \phi_p z^p)^{-1}, \text{ dla } |z| \leq 1.$$

.....

Średnio – kwadratowy błąd predykcji  $h$  – tego kroku  $P_n X_{n+h}$ , znajdujemy wtedy z równania (3.108). Ma on następującą postać:

$$\sigma_n^2(h) := E(X_{n+h} - P_n X_{n+h})^2 = \sum_{j=0}^{h-1} \left( \sum_{r=0}^j \chi_r \theta_{n+h-r-1, j-r} \right)^2 v_{n+h-j-1} \quad (3.111)$$

Przyjmując odwracalność procesów ARMA, dla  $n \rightarrow \infty$ , równania (3.102) i (3.111) można zapisać w postaci słusznej w przybliżeniu dużej próby:

$$P_n X_{n+h} \approx \sum_{i=1}^p \phi_i P_n X_{n+h-i} + \sum_{j=h}^q \theta_j (X_{n+h-j} - \hat{X}_{n+h-j}) \quad (3.112)$$

oraz

$$\sigma_n^2(h) \approx \sigma^2 \sum_{j=0}^{h-1} \left( \sum_{r=0}^j \chi_r \theta_{j-r} \right)^2 = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2 \quad (3.113)$$

gdzie 
$$\sum_{j=0}^{\infty} \psi_j z^j = \left( \sum_{j=0}^{\infty} \chi_j z^j \right) \left( \sum_{j=0}^{\infty} \theta_j z^j \right) = \theta(z) / \phi(z), \quad |z| \leq 1.$$

### 3.9 Predykcja niestacjonarnych procesów ARIMA

Obecnie pokażemy jak za pomocą metod opisanych wcześniej można przewidywać wartości niestacjonarnych procesów ARIMA  $(p, d, q)$ .

Jeśli  $d \geq 1$ , wtedy momenty, pierwszy  $EX_t$  i drugi  $E(X_{t+h}X_t)$  nie są określone przez równanie różnicowe (1.171):

$$\phi^*(B)X_t \equiv \phi(B)(1-B)^d X_t = \theta(B)Z_t$$

gdzie  $\{Z_t\} \sim \text{WN}(0, \sigma^2)$ .

Nie możemy zatem oczekiwać, że określimy najlepsze liniowe przewidywanie dla  $\{X_t\}$  bez dalszych założeń.

Przypuśćmy, że  $\{Y_t\}$  jest przyczynowym procesem ARMA  $(p, q)$ , i że  $X_0$  jest dowolną zmienną losową.

Zdefiniujemy: 
$$X_t = X_0 + \sum_{j=1}^t Y_j \quad \text{dla } t = 1, 2, \dots \quad (3.114)$$

Wtedy  $\{X_t, t \geq 0\}$  jest procesem ARIMA  $(p, 1, q)$ , z wartością oczekiwaną:  $EX_t = EX_0$

oraz autokowariancją  $E(X_{t+h}X_t) - (EX_0)^2$  zależną od  $\text{Var}(X_0)$  i  $\text{Cov}(X_0, Y_j)$  gdzie  $j = 1, 2, \dots$

Najlepszą liniową predykcją dla  $X_{n+1}$  bazującą na  $X_0, X_1, \dots, X_n$  jest rzutowanie  $P_{S_n} X_{n+1}$  gdzie:

$$S_n = \overline{sp}\{X_0, X_1, \dots, X_n\} = \overline{sp}\{X_0, Y_1, \dots, Y_n\}. \quad (3.115)$$

Tak więc otrzymujemy, że:

$$P_{S_n} X_{n+1} = P_{S_n} (X_0 + Y_1 + \dots + Y_n) = X_n + P_{S_n} Y_{n+1} \quad (3.116)$$

Aby wyznaczyć to rzutowanie, należy w ogólności znać  $E(X_0 Y_j)$  dla  $j = 1, \dots, n+1$ , oraz  $EX_0^2$ .

Jednakże jeśli założymy, że  $X_0$  jest nie skorelowane z  $Y_j$ ,  $j = 1, 2, \dots$  wtedy  $P_{S_n} X_{n+1}$  jest po prostu rzutowaniem  $Y_{n+1}$  na  $\overline{sp}\{Y_1, \dots, Y_n\}$  tak jak opisano to wcześniej (Przewidywanie rekurencyjne dla procesów ARMA(p,q)) przy omawianiu rzutowania stacjonarnych procesów ARMA (p,q). Założenie, że  $X_0$  jest nie skorelowane z  $Y_j$ ,  $j = 1, 2, \dots$  wystarcza w tym przypadku, by określić najlepsze liniowe przewidywanie  $P_{S_n} X_{n+1}$ .

Wróćmy teraz do przypadku ogólnego, przyjmijmy, że nasz obserwowany proces  $\{X_t\}$  spełnia równanie różnicowe:

$$(1 - B)^d X_t = Y_t, \text{ dla } t = 1, 2, \dots \quad (3.117)$$

gdzie  $\{Y_t\}$  jest przyczynowym procesem ARMA (p, q), i że wektor  $(X_{1-d}, \dots, X_0)$  nie jest skorelowany z  $Y_t, t > 0$ . Równanie różnicowe może zostać napisane w formie:

$$X_t = Y_t - \sum_{j=1}^d \binom{d}{j} (-1)^j X_{t-j}, \text{ dla } t = 1, 2, \dots \quad (3.118)$$

Wygodnie jest poprzez zmianę indeksów czasowych zakładać, że obserwujemy:

$$X_{1-d}, X_{2-d}, \dots, X_n. \quad (3.119)$$

(Obserwowanymi wartościami  $\{Y_t\}$  są wtedy:  $Y_1, \dots, Y_n$ ).

Naszym celem jest wyznaczenie najlepszego liniowego przewidywania dla  $X_{n+h}$ , bazując na  $X_{1-d}, X_{2-d}, \dots, X_n$ :

$$P_{S_n} X_{n+h} := P_{\overline{sp}\{X_{1-d}, \dots, X_n\}} X_{n+h}. \quad (3.120)$$

W zapisie poprzedniego podrozdziału, możemy napisać:

$$P_n Y_{n+h} = P_{\overline{sp}\{Y_1, \dots, Y_n\}} Y_{n+h} \quad (3.121)$$

i 
$$\hat{Y}_{n+1} = P_n Y_{n+1}.$$

Ponieważ:  $S_n = \overline{sp}\{X_{1-d}, \dots, X_0, Y_1, \dots, Y_n\}$  (3.122)

oraz ponieważ z założenia:  $\overline{sp}\{X_{1-d}, \dots, X_0\} \perp \overline{sp}\{Y_1, \dots, Y_n\}$  (3.123)

zatem otrzymujemy:  $P_{S_n} Y_{n+h} = P_{S_0} Y_{n+h} + P_n Y_{n+h} = P_n Y_{n+h}$ . (3.124)

Jeśli zastosujemy operator  $P_{S_n}$  do równania (3.118) (dla  $t = n + h$ ), to otrzymamy:

$$P_{S_n} X_{n+h} = P_{S_n} Y_{n+h} - \sum_{j=1}^d \binom{d}{j} (-1)^j P_{S_n} X_{n+h-j}. \quad (3.125)$$

Ponieważ przewidywania  $P_n Y_{n+1}, P_n Y_{n+2}, \dots$  mogą być znalezione z (3.102), zatem przewidywania  $P_{S_n} X_{n+1}, P_{S_n} X_{n+2}, \dots$  mogą zostać wyznaczone rekurencyjnie z (3.125).

W celu znalezienia średnio – kwadratowy błąd przewidywania, wygodnie jest wyrazić  $P_n Y_{n+h}$  w zależności od  $\{X_j\}$ .

Zdefiniujmy dla  $t \geq 0$ :

$$X_{t+1}^* = P_{S_t} X_{t+1}. \quad (3.126)$$

Wtedy z (3.118) oraz z (3.125) dla  $n = t$  otrzymujemy:

$$X_{t+1} - X_{t+1}^* = Y_{t+1} - \hat{Y}_{t+1}, \quad \text{dla } t \geq 0, \quad (3.127)$$

co w konsekwencji dla  $n > m = \max(p, q)$  oraz dla  $h \geq 1$ , daje nam:

$$P_n Y_{n+h} = \sum_{i=1}^p \phi_i P_n Y_{n+h-i} + \sum_{j=h}^q \theta_{n+h-1, j} (X_{n+h-j} - X_{n+h-j}^*). \quad (3.128)$$

Podstawiając  $\phi^*(z) = (1-z)^d \phi(z) = 1 - \phi_1^* z - \dots - \phi_{p+d}^* z^{p+d}$ , znajdujemy z równań (3.124), (3.125) oraz z (3.128), że dla  $n > m$  i  $h \geq 1$ :

$$P_{S_n} X_{n+h} = \sum_{j=1}^{p+d} \phi_j^* P_{S_n} X_{n+h-j} + \sum_{j=h}^q \theta_{n+h-1, j} (X_{n+h-j} - X_{n+h-j}^*), \quad (3.129)$$

który wygląda analogicznie jak formuła przewidywania  $h$  – tego kroku (3.102) dla procesów ARMA.



Ten sam argument, który doprowadził do (3.111), pokazuje, że średnio – kwadratowy błąd predykcji  $h$  – tego kroku wyraża się przez [1]:

$$\sigma_n^2(h) = E(X_{n+h} - P_{S_n} X_{n+h})^2 = \sum_{j=0}^{h-1} \left( \sum_{r=0}^j \chi_r \theta_{n+h-r-1, j-r} \right)^2 v_{n+h-j-1}, \quad (3.130)$$

gdzie  $\theta_{n0} = 1$ ,

$$\chi(z) = \sum_{r=0}^{\infty} \chi_r z^r = (1 - \phi_1^* z - \dots - \phi_{p+d}^* z^{p+d})^{-1}, \text{ dla } |z| < 1 \quad (3.131)$$

oraz: 
$$v_{n+h-j-1} = E(X_{n+h-j} - X_{n+h-j}^*)^2 = E(Y_{n+h-j} - \hat{Y}_{n+h-j})^2. \quad (3.132)$$

Współczynniki  $\chi_j$  mogą zostać znalezione rekurencyjnie z (3.110) gdzie zamiast  $\phi_j$  podstawiamy  $\phi_j^*$ .

Dla dużych  $n$ , możemy przybliżyć (3.130) (zastrzegając, że  $\theta(\cdot)$  jest odwracalny), przez :

$$\sigma_n^2(h) = \sum_{j=0}^{h-1} \psi_j^2 \sigma^2 \quad (3.133)$$

gdzie: 
$$\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\theta(z)}{\phi^*(z)}, \text{ dla } |z| < 1. \quad (3.134)$$

## Rozdział 4. Estymacja parametrów $\phi$ , $\theta$ modeli ARMA( $p$ , $q$ )

Poniższy rozdział poświęcony jest najbardziej oczywistej części modelowania, a mianowicie estymacji (przy ustalonych wartościach  $p$  i  $q$ ), wartości  $\vec{\phi} = (\phi_1, \dots, \phi_p)^T$ ,  $\vec{\theta} = (\theta_1, \dots, \theta_q)^T$ , oraz  $\sigma^2$ .

Dane zostaną skorygowane o średnią, tak, aby problem był sprowadzony do dopasowania określonego modelu ARMA ze średnią równą zero do danych  $x_1, \dots, x_n$ . Zatem, o ile model dopasowany do danych ma postać:

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q} \quad (4.1)$$

gdzie  $\{Z_t\} \sim WN(0, \sigma^2)$ , wtedy właściwy model dla pierwotnego stacjonarnego szeregu  $\{Y_t\}$

znajdujemy, gdy zastąpimy  $X_j$  przez  $Y_j - \bar{y}$ ,  $j = t, \dots, t-p$ , gdzie  $\bar{y} = \frac{1}{n} \sum_{j=1}^n y_j$  jest średnią z

próby dla oryginalnych danych, którą traktujemy jako stałą (co wynika to ze stacjonarności procesu).

## 4.1 Estymacja procesów autoregresyjnych AR(p)

### 4.1.1 Równanie Yule'a - Walker'a

Niech  $\{X_t\}$  będzie przyczynowym procesem AR(p) (z wartością oczekiwaną równą zero), danym równaniem:

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t, \text{ gdzie } Z_t \sim \text{WN}(0, \sigma^2). \quad (4.2)$$

Bazując na obserwacjach  $X_1, \dots, X_n$  będziemy chcieli znaleźć estymatory współczynników wektora  $\vec{\phi} = (\phi_1, \dots, \phi_p)^T$  oraz wariancji białego szumu  $\sigma^2$ .

Założenie przyczynowości procesu (Definicja 1.3) pozwala nam zapisać  $X_t$  w formie:

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} \quad t = 0, 1, 2, \dots \quad (4.3)$$

gdzie zgodnie z Twierdzeniem 1.6:

$$\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{1}{\phi(z)}, \quad |z| \leq 1. \quad (4.4)$$

Mnożąc każdą ze stron (4.2) przez  $X_{t-j}$   $j = 0, \dots, p$ , biorąc wartość oczekiwaną, i korzystając z (4.3) by wyznaczyć prawą stronę równania, otrzymujemy równanie Yule-Walker'a ((1.62), por.(1.121-126)):

$$\Gamma_p \vec{\phi} = \vec{\gamma}_p \quad (4.5)$$

$$\text{oraz} \quad \sigma^2 = \gamma(0) - \vec{\phi}^T \vec{\gamma}_p \quad (4.6)$$

gdzie  $\Gamma_p$  jest macierzą kowariancji  $[\gamma(t_i - t_j)]_{i,j=1}^p$ , a  $\vec{\gamma}_p = (\gamma(1), \gamma(2), \dots, \gamma(p))^T$ .

Równania te mogą zostać użyte by wyznaczyć  $\gamma(0), \dots, \gamma(p)$  z  $\sigma^2$  oraz  $\vec{\phi}$ . Z drugiej strony jeśli zastąpimy w równaniach (4.5) i (4.6) kowariancje  $\gamma(j)$ ,  $j = 0, \dots, p$  przez odpowiadające im w próbie estymatory  $\hat{\gamma}(j)$  to otrzymamy układ  $p+1$  równań dla tzw. estymatorów Yule - Walker'a  $\hat{\phi}$  i  $\hat{\sigma}^2$  parametrów  $\vec{\phi}$  oraz  $\sigma^2$ :

$$\hat{\Gamma}_p \hat{\phi} = \hat{\gamma}_p \quad (4.7)$$

$$\text{oraz} \quad \hat{\sigma}^2 = \hat{\gamma}(0) - \hat{\phi}^T \hat{\gamma}_p \quad (4.8)$$

gdzie:  $\hat{\Gamma}_p = [\hat{\gamma}(t_i - t_j)]_{i,j=1}^p$  i  $\hat{\gamma}_p = (\hat{\gamma}(1), \hat{\gamma}(2), \dots, \hat{\gamma}(p))^T$ .

Jeśli  $\hat{\gamma}(0) > 0$  wtedy  $\hat{\Gamma}_p$  jest nieosobliwa.

Dzieląc każdą ze stron równania (4.7) przez  $\hat{\gamma}(0)$  otrzymujemy :

$$\hat{\phi} = \hat{R}_p^{-1} \hat{\rho}_p \quad (4.9)$$

oraz : 
$$\hat{\sigma}^2 = \hat{\gamma}(0)[1 - \hat{\rho}_p^T \hat{R}_p^{-1} \hat{\rho}_p] , \quad (4.10)$$

gdzie 
$$\hat{\rho}_p = (\hat{\rho}(1), \dots, \hat{\rho}(p))^T = \hat{\gamma}_p / \hat{\gamma}(0).$$

Uwaga: Definiując  $\hat{\phi}$  tak jak w równaniu (4.9), można pokazać, że [1]:

$$1 - \hat{\phi}_1 z - \dots - \hat{\phi}_p z^p \neq 0 \text{ dla } |z| \leq 1.$$

Oznacza to, że dopasowywany model  $X_t - \hat{\phi}_1 X_{t-1} - \dots - \hat{\phi}_p X_{t-p} = Z_t$  gdzie  $Z_t \sim \text{WN}(0, \sigma^2)$  jest przyczynowy. Autokowariancja  $\gamma_F(h)$ ,  $h = 0, \dots, p$  dopasowanego modelu musi więc spełniać  $p+1$  równań liniowych:

$$\gamma_F(h) - \hat{\phi}_1 \gamma_F(h-1) - \dots - \hat{\phi}_p \gamma_F(h-p) = \begin{cases} 0 & \text{dla } h = 1, \dots, p \\ \hat{\sigma}^2 & \text{dla } h = 0 \end{cases} \quad (4.11)$$

otrzymanych podobnie jak ich analogi (4.5) i (4.6) dla autokowariancji i wariancji .

Postać macierzowa dla (4.11) jest taka jak (4.7)- (4.8) tyle, że zamiast  $\hat{\gamma}(h)$  występują w niej  $\gamma_F(h)$  (por. również wyprowadzenie przejścia od (3.11) do (3.12)). Z porównania (4.11) z (4.7) i (4.8) widać więc, że ponieważ mają one takie same współczynniki  $\hat{\phi}_1, \dots, \hat{\phi}_p$ , zatem rozwiązaniem tych równań jest  $\gamma_F(h) = \hat{\gamma}(h)$  dla  $h = 0, \dots, p$ , skąd wnioskujemy, że autokowariancja z dopasowanego modelu z opóźnieniami  $0, \dots, p$  zgadza się z odpowiednią autokowariancją z próby.

Powyższe rozważania pokazują, że dla każdej nieosobliwej macierzy kowariancji  $\Gamma_{p+1} = [\gamma(t_i - t_j)]_{i,j=1}^{p+1}$  istnieje proces AR(p) którego autokowariancje w opóźnieniach  $0, \dots, p$  wynoszą  $\gamma(0), \dots, \gamma(p)$ , gdzie wymagane współczynniki i wariancję białego szumu znajdziemy z równania (4.9) i (4.10) poprzez (przyrównanie) zastąpienie  $\hat{\rho}(j)$  przez  $\gamma(j)/\gamma(0)$ ,  $j = 0, \dots, p$ , oraz  $\hat{\gamma}(0)$  przez  $\gamma(0)$ .

**Uwaga:**

Może się jednak okazać, że nie ma procesu MA( $p$ ), który miałby taką własność. Na przykład jeśli:  $\gamma(0) = 1$  i  $\gamma(1) = \gamma(-1) = \beta$ , wtedy macierz

$$\Gamma_2 = \begin{pmatrix} \gamma(0) & \gamma(-1) \\ \gamma(1) & \gamma(0) \end{pmatrix}$$

jest nieosobliwa dla wszystkich  $\beta \in (-1, 1)$ . W konsekwencji istnieje dla wszystkich  $\beta \in (-1, 1)$ , proces AR(1) z autokowariancją 1 i  $\beta$  w opóźnieniach kolejno 0 i 1. Można pokazać, że co prawda, istnieje również proces MA(1) z autokowariancją 1 i  $\beta$  w opóźnieniach odpowiednio 0 i 1, ale wtedy i tylko wtedy gdy  $|\beta| \leq \frac{1}{2}$ .

**Uwaga:**

Powyzsza metoda przyrównywania teoretycznych momentów populacji do ich estymatorów w próbie jest nazwana metodą momentów. Zazwyczaj estymatory (np.  $\hat{\phi}$  lub  $\hat{\theta}$ ) parametrów modelu, otrzymane tą metodą są mniej efektywne niż estymatory otrzymane metodą największej wiarygodności czy metodą najmniejszych kwadratów.

Np. estymacja parametrów  $\vec{\theta}$  procesu MA(1) przez porównanie teoretycznych i empirycznych funkcji autokorelacji dla lag 1, okazuje się wyjątkowo nieefektywna.

Jednakże, w przypadku procesów AR( $p$ ) można pokazać, że estymator Yule'a - Walker'a  $\vec{\hat{\phi}}$  ma asymptotycznie ( $n \rightarrow \infty$ ) taki sam rozkład jak estymator parametru  $\vec{\phi}$  otrzymany metodą największej wiarygodności.

Obecnie podamy twierdzenie, które umożliwia podanie przedziałów ufności dla  $\vec{\phi}$  w przypadku dużych próbek.

**Twierdzenie 4.1**

Jeśli  $\{X_t\}$  jest przyczynowym procesem AR( $p$ ) (4.2), dla którego  $Z_t \sim \text{iid}(0, \sigma^2)$ ,

a  $\vec{\hat{\phi}}$  jest estymatorem Yule'a - Walker'a dla  $\vec{\phi}$ , wtedy :

$$n^{1/2}(\vec{\hat{\phi}} - \vec{\phi}) \Rightarrow N(\vec{0}, \sigma^2 \Gamma_p^{-1}) \quad (4.12)$$

gdzie  $\Gamma_p^{-1}$  jest macierzą odwrotną do macierzy kowariancji  $\hat{\Gamma}_p = [\hat{\gamma}(t_i - t_j)]_{i,j=1}^p$ .

Ponadto estymator  $\hat{\sigma}^2$  jest zbieżny stochastycznie do wariancji  $\sigma^2$  (tzn. dla każdego  $\varepsilon > 0$ ,  $P(|\hat{\sigma}^2 - \sigma^2| < \varepsilon) = 1$  dla  $n \rightarrow \infty$ ), tzn.:

$$\hat{\sigma}^2 \xrightarrow{P} \sigma^2. \quad (4.13)$$

**Uwaga:** W dopasowaniu modelu AR( $p$ ) do danych wielkość  $p$  będzie na ogół nieznana. Jeśli prawdziwy rząd wynosi  $p$ , a my usiłujemy dopasować proces o rzędzie  $m$ , to powinniśmy się spodziewać, że oszacowane współczynniki  $\hat{\phi}_{mm}$  wektora  $\vec{\hat{\phi}}_m = (\hat{\phi}_{m1}, \dots, \hat{\phi}_{mm})^T$  będą miały małą wartość dla każdego  $m > p$ . Ponadto, chociaż nawet w Gaussowskim przypadku, dokładny rozkład  $\hat{\phi}_{mm}$  dla  $m > p$  nie jest znany, jednak poniższe Twierdzenie 4.2 podaje asymptotyczny rezultat, który jest niezwykle pożyteczny w identyfikacji odpowiedniego rzędu dopasowywanego procesu.

#### **Twierdzenie 4.2**

Jeśli  $\{X_t\}$  jest przyczynowym procesem AR( $p$ ):

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t, \text{ gdzie } Z_t \sim \text{iid}(0, \sigma^2) \quad (4.14)$$

oraz jeśli 
$$\vec{\hat{\phi}}_m = (\hat{\phi}_{m1}, \dots, \hat{\phi}_{mm})^T = \hat{R}_m^{-1} \hat{\rho}_m, \quad m > p \quad (4.15)$$

gdzie 
$$R_m := [\hat{\rho}(i-j)]_{i,j=1}^m, \quad (4.16)$$

wtedy :

$$n^{1/2}(\vec{\hat{\phi}}_m - \vec{\phi}_m) \Rightarrow N(\vec{0}, \sigma^2 \Gamma_m^{-1}) \quad (4.17)$$

gdzie  $\vec{\phi}_m$  jest wektorem współczynników najlepszego liniowego przewidywania  $\vec{\phi}_m^T \vec{X}_m$  dla  $X_{m+1}$  na podstawie  $\vec{X}_m = (X_m, \dots, X_1)^T$ ,

gdzie 
$$\vec{\phi}_m = R_m^{-1} \vec{\rho}_m. \quad (4.18)$$

W szczególnym przypadku dla  $m > p$  zachodzi:

$$n^{1/2} \hat{\phi}_{mm} \Rightarrow N(0,1) \quad (4.19)$$

Uwaga: Symbol  $\Rightarrow$  oznacza zbieżność w sensie zbieżności dystrybuant.

Dowód powyższego twierdzenia znajduje się w [1].

### **Uwaga:**

Zastosowanie Twierdzenia 4.1 do identyfikacji odpowiedniego rzędu dopasowywanego procesu będzie omówione później.

### **4.1.2 Wstępna estymacja procesu autoregresji z wykorzystaniem algorytmu Durbin'a – Levinson'a**

Załóżmy, że mamy obserwacje  $x_1, \dots, x_n$  stacjonarnego szeregu czasowego ze średnią zero.

Zakładając, że:

$$\hat{\gamma}(0) > 0 \quad (4.20)$$

możemy dopasować proces autoregresji rzędu  $m < n$  do danych wykorzystując równania Yule'a – Walker'a.

Dopasowany proces  $AR(m)$  ma postać :

$$X_t - \hat{\phi}_{m1}X_{t-1} - \dots - \hat{\phi}_{mm}X_{t-m} = Z_t, \quad \text{gdzie } Z_t \sim \text{WN}(0, \hat{v}_m) \quad (4.21)$$

gdzie z (4.9) i (4.10) mamy:

$$\hat{\phi}_m = (\hat{\phi}_{m1}, \dots, \hat{\phi}_{mm})^T = \hat{R}_m^{-1} \hat{\rho}_m \quad (4.22)$$

$$\text{oraz } \hat{v}_m \equiv \hat{\sigma}^2 = \hat{\gamma}(0)[1 - \hat{\rho}_m^T \hat{R}_m^{-1} \hat{\rho}_m] , \quad (4.23)$$

Porównując (4.22) i (4.23) z Twierdzeniem 3.1 i Wnioskiem z tego twierdzenia (por. (3.24'), (3.27)), można zauważyć, że  $\hat{\phi}_m$  i  $\hat{v}_m$  są związane z autokowariancjami w próbie, tak jak  $\bar{\phi}_m$  i  $v_m$  są związane z autokowariancjami procesu  $\{X_t\}$ . Na dodatek z porównania (4.23) i (3.27), okazało się, że estymator średnio – kwadratowego błędu przewidywania (3.27),  $v_m \equiv E(X_{m+1} - \hat{X}_{m+1})^2$ , jest równocześnie estymatorem wariancji białego szumu występującego w procesie  $AR(m)$ .

### **Podsumujmy:**

Tak jak jest to w twierdzeniu 4.2,  $\bar{\phi}_m$  jest zdefiniowany jako wektor współczynników najlepszego liniowego przewidywania  $\bar{\phi}_m^T \bar{X}_m$  dla  $X_{m+1}$  na podstawie  $\bar{X}_m = (X_m, \dots, X_1)^T$ , natomiast  $v_m$  jest odpowiednim średnim błędem kwadratowym przewidywania  $\hat{X}_{m+1}$ , przy czym estymatorem  $v_m$  jest zgodnie z (4.23) estymator  $\hat{\sigma}^2$  wariancji białego szumu występującego w procesie AR( $m$ ).

### **Uwaga:**

Jeśli więc  $\hat{\gamma}(0) > 0$ , tak że  $\hat{R}_1, \hat{R}_2, \dots$  są nieosobliwe, wtedy możemy wykorzystać algorytm Durbin'a – Levinson'a (Rozdział 3) do dopasowania do empirycznych danych modeli autoregresji z sukcesywnie rosnącymi rzędami 1, 2, ... . Oszacowane wektory współczynników  $\hat{\phi}_1, \hat{\phi}_2, \dots$ , oraz wariancje białego szumu  $\hat{v}_1, \hat{v}_2$  są wyznaczone rekurencyjnie z próby, podobnie jak  $\bar{\phi}_1, \bar{\phi}_2, \dots$  i  $v_1, v_2, \dots$  są wyznaczone w Rozdziale 3 ((3.35) – (3.39)) na podstawie kowariancji.

Otrzymujemy zatem twierdzenie:

### **Twierdzenie 4.3 (Algorytm Durbin'a – Levinson'a dla dopasowanych modeli autoregresji).**

Jeśli  $\hat{\gamma}(0) > 0$ , wtedy dopasowane modele autoregresji AR( $m$ ) dla  $m = 1, 2, \dots, n-1$  można określić rekurencyjnie z relacji:

$$\hat{\phi}_{11} = \hat{\rho}(1), \quad \hat{v}_1 = \hat{\gamma}(0)(1 - \hat{\rho}^2(1)) \quad (4.24)$$

$$\hat{\phi}_{mm} = [\hat{\gamma}(m) - \sum_{j=1}^{m-1} \hat{\phi}_{m-1,j} \hat{\gamma}(m-j)] / \hat{v}_{m-1} \quad (4.25)$$

$$\begin{bmatrix} \hat{\phi}_{m1} \\ \vdots \\ \hat{\phi}_{m,m-1} \end{bmatrix} = \begin{bmatrix} \hat{\phi}_{m-1,1} \\ \vdots \\ \hat{\phi}_{m-1,m-1} \end{bmatrix} - \hat{\phi}_{mm} \begin{bmatrix} \hat{\phi}_{m-1,m-1} \\ \vdots \\ \hat{\phi}_{m-1,1} \end{bmatrix} \quad (4.26)$$

$$\text{tzn.} \quad \hat{\phi}_{mj} = \hat{\phi}_{m-1,j} - \hat{\phi}_{mm} \hat{\phi}_{m-1,m-j}, \quad j=1, \dots, m-1 \quad (4.27)$$

$$\text{oraz} \quad \hat{v}_n = \hat{v}_{n-1} [1 - \hat{\phi}_{nn}^2] \quad (4.28)$$

Użycie powyższych rekurencyjnych zależności pozwala na ominięcie wyznaczania macierzy odwrotnej, co byłoby konieczne gdyby wyznaczać  $\hat{\phi}_m$  i  $\hat{v}_m$  bezpośrednio w oparciu o (4.7) i (4.8). Twierdzenie 4.3 podaje również postać estymatorów  $\hat{\phi}_{11}, \hat{\phi}_{22}, \dots$  częściowych funkcji autokorelacji ((3.40), (3.41)) dla opóźnień lag równych  $1, 2, \dots$ . Estymatory te są szczególnie użyteczne: po pierwsze w podjęciu decyzji o własności modelu autoregresji i po drugie w wyborze właściwego rzędu dopasowywanego modelu.

Z Rozdziału 1 wiemy, że dla procesu  $AR(p)$  wszystkie współczynniki funkcji PACF  $\hat{\phi}_{mm}$  dla  $m > p$  są równe zero (1.160). Można powiedzieć, że funkcja ta urywa się na odstępie  $p$ . Własność tą wykorzystuje się do określenia rzędu autoregresji, czyli parametru  $p$ .

Ponadto z Twierdzenia 4.2 (4.19) wiemy, że  $\hat{\phi}_{mm}$  ma dla dużych  $n$  i  $m > p$ , rozkład będący w przybliżeniu rozkładem normalnym ze średnią zero i wariancją równą  $\frac{1}{n}$ . Zatem jeśli model autoregresji jest właściwym modelem dla danych, wtedy powinien istnieć skończony lag, poza którym obserwowane wartości estymatora  $\hat{\phi}_{mm}$  są zgodne z rozkładem  $N(0, \frac{1}{n})$ .

W szczególności, jeśli rząd procesu wynosi  $p$ , wtedy dla  $m > p$  i poziomu ufności  $1 - \alpha = 0,95$ , estymator  $\hat{\phi}_{mm}$  przyjmuje, z prawdopodobieństwem w przybliżeniu równym 0,95, wartości z przedziału  $\pm 1,96 \frac{1}{\sqrt{n}}$ . Sugeruje to przyjęcie, jako wstępnego oszacowania rzędu  $p$ , takiej wartości  $r$ , dla której wartości  $|\hat{\phi}_{mm}| < 1,96 \frac{1}{\sqrt{n}}$  dla  $m > r$ . W końcu gdy tylko wartość  $p$  zostanie ostatecznie wybrana, dopasowany proces jest określony przez (4.21) do (4.23) dla  $m=p$ .

**Uwaga:**

Bardziej systematyczne podejście do wyboru rzędu  $p$  jest oparte o kryterium AICC (Rozdział 5).



Asymptotyczne przedziały ufności dla prawdziwego wektora współczynników  $\vec{\phi}_p$  i jego składowych mogą być wyznaczone w oparciu o Twierdzenie 4.1. Tak więc jeśli  $\chi^2_{1-\alpha}(p)$  oznacza kwantyl rzędu  $(1-\alpha)$  rozkładu chi – kwadrat z  $p$  stopniami swobody, wtedy dla dużej wielkości próby  $n$ , obszar ufności (zapis symboliczny):

$$\left\{ \vec{\phi} \in \mathbf{R}^p : (\vec{\phi} - \hat{\vec{\phi}}_p)^T \hat{\Gamma}_p (\vec{\phi} - \hat{\vec{\phi}}_p) \leq n^{-1} \hat{v}_p \chi^2_{1-\alpha}(p) \right\} \quad (4.29)$$

pokrywa z prawdopodobieństwem  $1-\alpha$ , wartości wektora współczynników  $\vec{\phi}_p = (\phi_{p1}, \dots, \phi_{pp})^T$ .

Podobnie, wykorzystując Twierdzenie 4.1 można pokazać, że dla dużego  $n$ , przedział ufności (zapis symboliczny):

$$\left\{ \phi \in \mathbf{R} : |\phi - \hat{\phi}_{pj}| \leq n^{-1/2} u_\alpha \hat{v}_{jj}^{1/2} \right\} \quad (4.30)$$

pokrywa wartość  $\phi_{pj}$  na poziomie ufności  $(1-\alpha)$ , gdzie  $u_\alpha$  jest kwantylem rzędu  $(1-\alpha/2)$  rozkładu normalnego  $N(0,1)$  a  $\hat{v}_{jj}$  jest  $j$ -tym diagonalnym wyrazem macierzy  $\hat{v}_p \hat{\Gamma}_p^{-1}$ .

#### **Uwaga.**

Zależności (4.29) oraz (4.30) można wyprowadzić posługując się Twierdzeniem 4.1 oraz poniższym Twierdzenie 4.4.

#### **Twierdzenie 4.4**

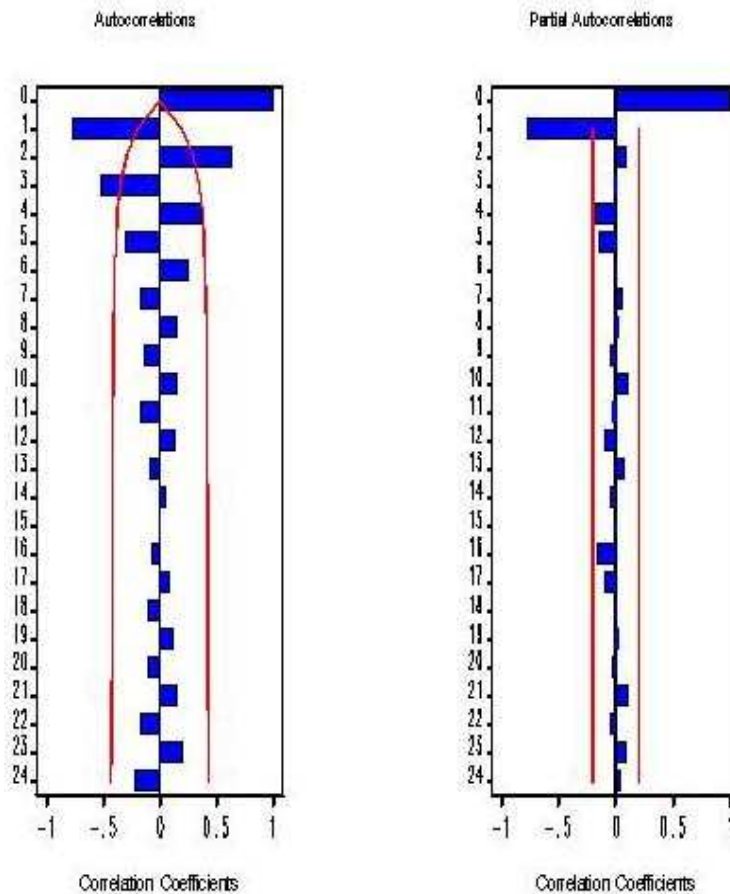
Jeśli  $\vec{X}_n = (X_{n1}, \dots, X_{nm})^T \Rightarrow N(\vec{0}, \Sigma)$  (tzn. dla  $n \rightarrow \infty$  istnieje zbieżność w sensie dystrybucji do rozkładu  $N(\vec{0}, \Sigma)$ ) i jeśli  $\Sigma_n \xrightarrow{P} \Sigma$  gdzie  $\Sigma$  jest nieosobliwą macierzą, to :

$$\vec{X}_n^T \Sigma_n^{-1} \vec{X}_n \Rightarrow \chi^2(m) \quad (4.31)$$

#### **Przykład 4.1**

Dla procesu AR(1) teoretyczna wartość funkcji ACF wynosi zgodnie z (1.40b)  $\rho(h) = \phi_1^h$ . Możemy również zauważyć, że proces ten ma skończoną wariancję tylko dla  $|\phi_1| < 1$ .

Wynika z tego że ACF maleje dla tego przykładu wykładniczo. Natomiast zgodnie z (1.157) PACF jest równa zero dla  $h \geq 1$ . Poniżej zamieszczono wykres ACF i PACF dla wygenerowanej (z wykorzystaniem generatora SAS'a) realizacji procesu  $X_t = -0,8X_{t-1} + Z_t$  w próbie. Na wykresach widać 95%-owe przedziały ufności.



Wykres 4.1 Autokorelacja (ACF) oraz autokorelacja cząstkowa (PACF) procesu  $X_t = -0,8X_{t-1} + Z_t$  gdzie  $Z_t \sim WN(0,1)$ .

Na podstawie powyższych autokorelogramów w próbie, można wyciągnąć przypuszczenie, że pochodzą one z realizacji procesu AR(1).

#### **Przykład 4.2**

Tysiąc obserwacji  $x_1, \dots, x_{1000}$ , stacjonarnego procesu ze średnią zero, dało autokowariancję z próby:

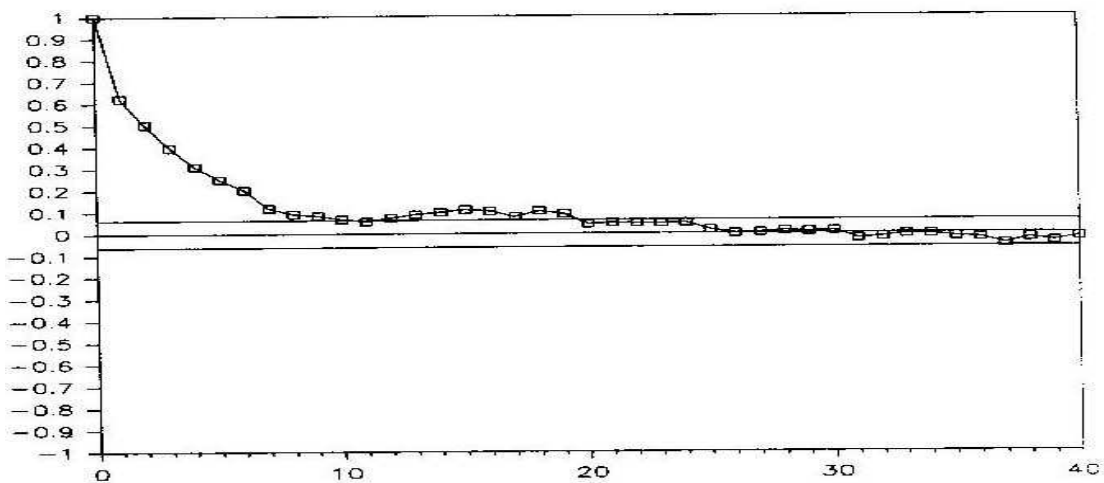
$$\hat{\gamma}(0) = 3,6840, \hat{\gamma}(1) = 2,2948, \hat{\gamma}(2) = 1,8491.$$

Zakładamy, że dane pochodzą z procesu AR( $p$ ) i stosując Algorytm Durbin'a – Levinson'a dopasowujemy rekurencyjnie proces autoregresji do danych (Twierdzenie 4.3), otrzymując:

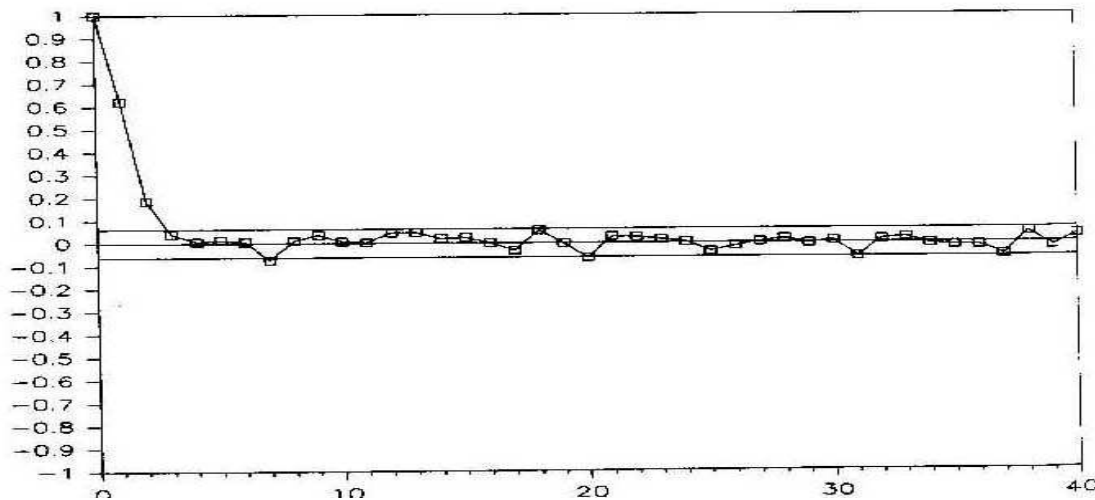
$$\begin{aligned} \hat{\phi}_{11} &= \hat{\rho}(1) = \frac{\hat{\gamma}(1)}{\hat{\gamma}(0)} = 0,6229 \\ \hat{v}_1 &= \hat{\gamma}(0)(1 - \hat{\rho}^2(1)) = 2,2545 \\ \hat{\phi}_{22} &= (\hat{\gamma}(2) - \hat{\phi}_{11}\hat{\gamma}(1)) / \hat{v}_1 = 0,1861 \\ \hat{\phi}_{21} &= \hat{\phi}_{11} - \hat{\phi}_{22}\hat{\phi}_{11} = 0,5070 \\ \hat{v}_2 &= \hat{v}_1(1 - \hat{\phi}_{22}^2) = 2,1764 \end{aligned} \tag{4.32}$$

Wyznaczenie kolejnych wartości  $\hat{\phi}_{mj}$  funkcji częściowej autokorelacji w próbie dla wzrastającej liczby  $n$ , można przeprowadzić komputerowo. Dodatkowo [za 1] zamieszczono wartości dla kolejne wartości  $\hat{\rho}(j)$  uzyskanych z próby.

Dla rekurencji (4.24) – (4.28) dla wzrastającej wartości  $n$ , otrzymujemy w rozważanym przykładzie poniższe wykresy funkcji ACF oraz PACF [1]. Granice zaznaczone na wykresach dotyczą przedziałów 95%-owych przedziałów ufności (czyli na zewnątrz są obszary krytyczne dla poziomu istotności 0,05.)



Wykres 4.2 Funkcja ACF dla procesu z Przykładu 4.2 [wykres z [1]]



Wykres 4.3 Funkcja PACF dla procesu z przykładu 4.2 [wykres z [1]]

Postacie wykresów (a szczególnie PACF) sugerują, że właściwym modelem jest proces AR(2). Zatem przerywamy rekurencję na  $m=2$ .

Wykorzystując estymatory,  $\hat{\phi}_{22}, \hat{\phi}_{21}$ , i  $\hat{v}_2$ , których wartości w próbie wyznaczono powyżej, otrzymujemy postać dopasowanego modelu:

$$X_t - 0,5070X_{t-1} - 0,1861X_{t-2} = Z_t, \{Z_t\} \sim WN(0; 2,1764). \quad (4.33)$$

Na koniec, z Twierdzenia 4.1 widzimy, że wektor błędu  $\hat{\phi} - \bar{\phi}$  ma w przybliżeniu rozkład normalny ze średnią równą  $\bar{0}$  i następującą macierzą kowariancji (dla  $p=2$ ):

$$\begin{aligned} n^{-1}\hat{v}_2\hat{\Gamma}_2^{-1} &= n^{-1}\left(1 - \sum_{j=1}^2 \hat{\rho}(j)\hat{\phi}_{2j}\right) \begin{pmatrix} 1 & \hat{\rho}(1) \\ \hat{\rho}(1) & 2 \end{pmatrix}^{-1} = \\ &= \frac{1}{1000} \left(1 - 0,6229 \cdot 0,5070 + \frac{1,8491}{3,6840} \cdot 0,1861\right) \cdot \begin{pmatrix} 1 & 0,6229 \\ 0,6229 & 1 \end{pmatrix}^{-1} = \\ &= \begin{pmatrix} 0,000965 & -0,000601 \\ -0,000601 & 0,000965 \end{pmatrix} \end{aligned} \quad (4.34)$$

Z (4.30) otrzymujemy teraz 0,95 % - owe przedziały ufności dla  $\phi_i$ ,  $i=1,2$  w postaci  $\phi_i \pm 1.96 \cdot (0,000965)^{0,5}$ , tzn. dla:

$$0,5070 \pm 0,0609 \quad \text{dla } \phi_1, \quad 0,1861 \pm 0,0609 \quad \text{dla } \phi_2 \quad (4.35)$$

Ponieważ przykład ten był wzięty wprost z [1], zatem warto za nimi zaznaczyć, że dane empiryczne pochodziły z symulacji komputerowej z procesu AR(2), dla którego  $\phi_1 = 0.5$ ,  $\phi_2 = 0.2$ , natomiast wariancja białego szumu  $\sigma^2 = 2,25$ .

## 4.2 Estymacja procesów średniej ruchomej MA( $q$ )

Tak jak można dopasować modele autoregresji rzędu  $1,2,\dots$  do danych  $x_1, \dots, x_n$  korzystając z funkcji autokowariancji z próby i stosując algorytm Durbin'a – Levinson'a, tak wykorzystując algorytm innowacyjny można dopasować modele rzędu  $m = 1,2,\dots$ , z średnią ruchomą:  $X_t = Z_t + \hat{\theta}_{m1}Z_{t-1} + \dots + \hat{\theta}_{mm}Z_{t-m}$ , gdzie  $Z_t \sim WN(0, \hat{v}_m)$ .

Poniżej zostaną podane postacie estymatorów  $\hat{\theta}_m := (\hat{\theta}_{m1}, \dots, \hat{\theta}_{mm})^T$  wektora współczynników oraz wariancji białego szumu  $\hat{v}_m$ ,  $m = 1,2,\dots$ .

**Definicja 4.1 ( innowacyjne estymatory dla parametrów procesów MA( $q$ ) )**

Jeżeli  $\hat{\gamma}(0) > 0$ , to definiujemy innowacyjne estymatory  $\hat{\theta}_m, \hat{v}_m$  z równania

$$X_t = Z_t + \hat{\theta}_{m1}Z_{t-1} + \dots + \hat{\theta}_{mm}Z_{t-m},$$

gdzie  $Z_t \sim \text{WN}(0, \hat{v}_m)$ , dla  $m = 1, 2, \dots, n-1$ , (4.36)

poprzez następujące związki rekurencyjne:

$$\hat{v}_0 = \hat{\gamma}(0) \tag{4.37}$$

$$\hat{\theta}_{m,m-k} = \hat{v}_k^{-1} \left[ \hat{\gamma}(m-k) - \sum_{j=0}^{k-1} \hat{\theta}_{m,m-j} \hat{\theta}_{k,k-j} \hat{v}_j \right], \quad k = 1, 2, \dots, m-1 \tag{4.38}$$

oraz: 
$$\hat{v}_m = \hat{\gamma}(0) - \sum_{j=0}^{m-1} \hat{\theta}_{m,m-j}^2 \hat{v}_j . \tag{4.39}$$

Poniższe twierdzenie uzasadnia stosowanie estymatorów podanych w Definicji 4.1.

**Twierdzenie 4.5 (asymptotyczne zachowanie dla  $\hat{\theta}_m$ )**

Niech  $\{X_t\}$  będzie przyczynowym i odwracalnym procesem ARMA( $p, q$ )

$$\phi(B)X_t = \theta(B)Z_t, \quad \{Z_t\} \sim \text{iid}(0, \sigma^2), \tag{4.40}$$

gdzie  $EZ_t^4 < \infty$  (4.41)

oraz niech  $\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \theta(z) / \phi(z)$ , gdzie  $|z| \leq 1$ , (4.42)

z  $\psi_0 = 1$  oraz  $\psi_j = 0$  dla  $j < 0$ .

Wtedy, dla dowolnego ciągu dodatnich liczb całkowitych  $\{m(n), n = 1, 2, \dots\}$ , takiego, że  $m < n$ ,  $m \rightarrow \infty$  oraz  $m = o(n^{1/3})$  gdy  $n \rightarrow \infty$ , zachodzi dla każdego  $k$  następująca asymptotyczna relacja:

$$n^{1/2} (\hat{\theta}_{m1} - \psi_1, \hat{\theta}_{m2} - \psi_2, \dots, \hat{\theta}_{mk} - \psi_k)^T \Rightarrow N(\bar{0}, \mathbf{A}) \tag{4.43}$$

gdzie  $\mathbf{A} = [a_{ij}]_{i,j=1}^k$  (4.44)

oraz  $a_{ij} = \sum_{r=1}^{\min(i,j)} \psi_{i-r} \psi_{j-r}$ . (4.45)

Ponadto  $\hat{v}_m \xrightarrow{P} \sigma^2$  (4.46)

Dowód znajduje się w [9].

**Uwaga.** Dla procesu MA( $q$ )  $\phi(z) \equiv 1$  i  $\psi_j = \theta_j$ .

**Uwaga.**

Pomimo, że rekurencyjne dopasowanie modeli MA( $q$ ) (przy użyciu innowacyjnego algorytmu) przebiega analogicznie do przedstawionego rekurencyjnego dopasowania modeli AR( $p$ ), to istnieje jedna ważna różnica.

Dla procesów AR( $p$ ) estymator  $\hat{\phi}_p = (\phi_{p1}, \dots, \phi_{pp})^T$  jest estymatorem zgodnym dla wektora parametrów  $\vec{\phi}_p$  (tzn.  $\hat{\phi}_p \xrightarrow{P} \vec{\phi}_p$ ) jeśli liczebność próbki  $n \rightarrow \infty$ . Natomiast dla procesów MA( $q$ ) estymator  $\hat{\theta}_q = (\hat{\theta}_{q1}, \dots, \hat{\theta}_{qq})^T$  nie jest zgodnym estymatorem prawdziwego wektora parametrów  $\vec{\theta}_q$ , gdy wielkość próby  $n \rightarrow \infty$ . Aby estymator  $\hat{\theta}_q$  był zgodnym estymatorem dla  $\vec{\theta}_q$ , należy użyć estymatorów  $(\hat{\theta}_{m1}, \dots, \hat{\theta}_{mq})^T$ , z  $\{m(n)\}$  spełniającym założenia Twierdzenia 4.5.

**Uwaga:** Wyboru  $m$  dla dowolnej ustalonej wielkości próby, można dokonać zwiększając  $m$  do momentu, gdy wektor  $(\hat{\theta}_{m1}, \dots, \hat{\theta}_{mq})^T$  ustabilizuje się.

Stwierdzono w praktyce, że istnieje duży zakres wartości  $m$ , dla których fluktuacje  $\hat{\theta}_{mj}$  są małe w porównaniu z oszacowanym, asymptotycznym odchyleniem standardowym, które dla MA( $q$ ) wynosi zgodnie (4.44) i (4.45):

$$n^{-1/2} \left( \sum_{k=0}^{j-1} \hat{\theta}_{mk}^2 \right)^{1/2} \quad (4.47)$$

Wiemy z Rozdziału 1, że dla procesów MA( $q$ ) funkcja autokorelacji  $\rho(m)$ , dla  $m > q$  jest równa zero ((1.91) w Twierdzeniu 1.11).

Ponadto z formuły Bartlett'a (Rozdział 2.2, (2.45)) wiemy, że funkcja autokorelacji z próby  $\hat{\rho}(m)$  dla  $m > q$ , ma w przybliżeniu rozkład normalny ze średnią równą 0 oraz wariancją  $n^{-1}[1 + 2\rho^2(1) + \dots + 2\rho^2(q)]$ .

Własności te umożliwiają nam użycie wykresu funkcji autokorelacji z próby  $\hat{\rho}(m)$  dla  $m = 1, 2, \dots$  by zbadać, po pierwsze czy dane mogą być modelowane procesem MA( $q$ ), oraz po drugie otrzymać wstępne oszacowanie rzędu  $q$ .

Jeśli poza zbadaniem estymatora  $\hat{\rho}(m)$ ,  $m = 1, 2, \dots$ , badamy estymator wektora współczynników  $\hat{\theta}_m$  dla  $m = 1, 2, \dots$ , wtedy jesteśmy w stanie nie tylko ocenić adekwatność wyboru modelu z ruchomą średnią i oszacować jego rząd  $q$ , ale dodatkowo otrzymać wstępne wartości estymatorów  $\hat{\theta}_{m1}, \dots, \hat{\theta}_{mq}$  współczynników.

Zatem otrzymujemy wartości  $\hat{\theta}_{m1}, \dots, \hat{\theta}_{mm}, 0, 0, \dots$  dla  $m = 1, 2, \dots$ , powiększając  $m$ , aż w końcu wartości te ustabilizują się (tzn. aż fluktuacje każdej składowej  $\hat{\theta}_m$  będą rzędu  $n^{-1/2}$ , czyli rzędu asymptotycznej wartości odchylenia standardowego dla  $\hat{\theta}_{m1}$ ).

Z Twierdzenia 4.5 (por. (4.47)) wynika, że asymptotyczna wariancja dla  $\hat{\theta}_{mj}$  wynosi:

$$\sigma_j^2(\theta_1, \dots, \theta_{j-1}) = n^{-1} \sum_{k=0}^{j-1} \theta_k^2 \quad (4.48)$$

zatem zyskujemy również granice

$$\pm 1.96\hat{\sigma}_j, \quad \text{gdzie} \quad \hat{\sigma}_j = \sigma_j(\hat{\theta}_{m1}, \dots, \hat{\theta}_{m,j-1}). \quad (4.49)$$

Wartości  $\hat{\theta}_{mj}$ , które leżą poza tymi granicami sugerują, że odpowiadający im współczynnik  $\theta_j$  jest różny od zera. Co więcej, ponieważ estymatorem współczynnika  $\theta_j$  jest  $\hat{\theta}_{mj}$ , więc największa wartość opóźnienia, dla którego  $\hat{\theta}_{mj}$  leży jeszcze poza przedziałem ufności  $\pm 1.96\hat{\sigma}_j$  jest oszacowaniem rzędu procesu MA(q).

Bardziej systematyczne podejście do wyboru rzędu wykorzystujące kryterium AICC będzie podane w Rozdziale 5.

### **Uwaga.**

Twierdzenie 4.5 określa asymptotyczny przedział ufności dla wektora współczynników  $\vec{\theta}_q$  oraz dla jego składowych. Na przykład 95%-owy przedział ufności dla współczynnika  $\theta_j$  dany jest symbolicznie wzorem:

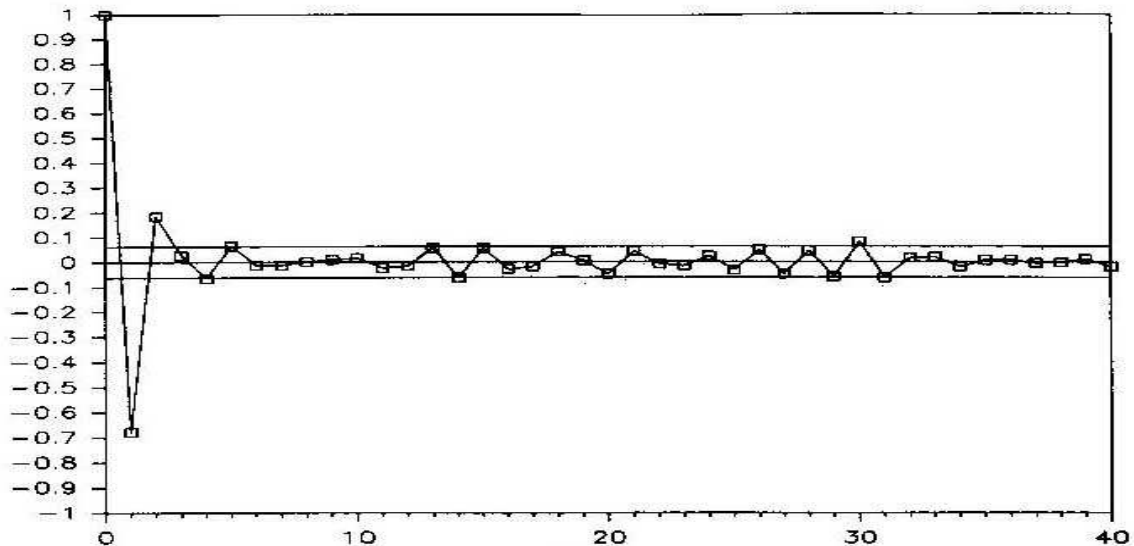
$$\left\{ \theta_j \in \mathbb{R} : |\theta_j - \hat{\theta}_{mj}| \leq 1.96n^{-1/2} \left( \sum_{k=0}^{j-1} \hat{\theta}_{mk}^2 \right)^{1/2} \right\}. \quad (4.50)$$

### Przykład 4.3

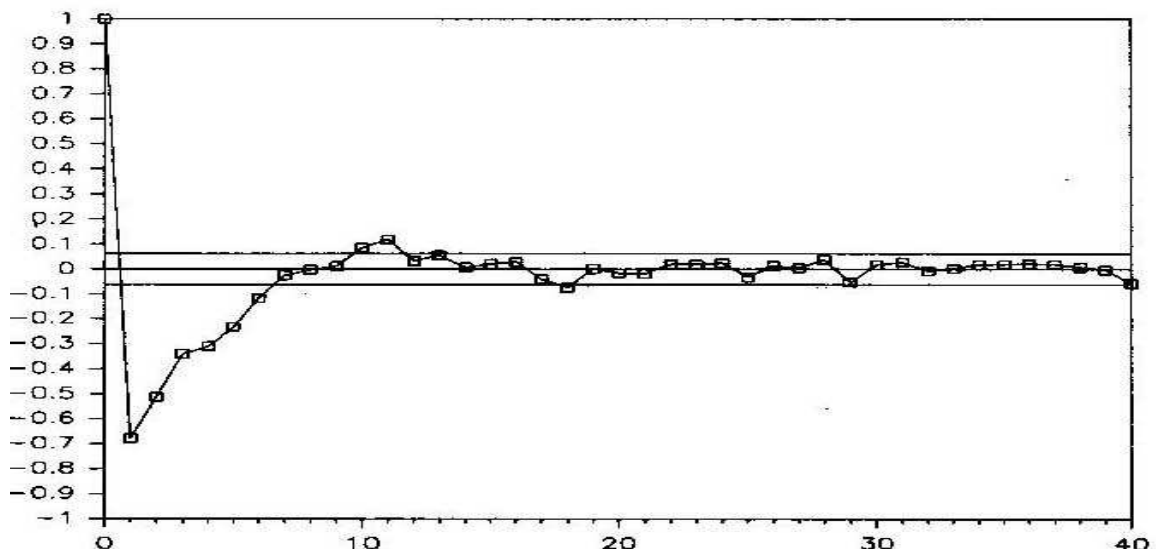
Na podstawie tysiąca obserwacji  $x_1, \dots, x_{1000}$  stacjonarnego procesu, ze średnią równą zero otrzymano (przykład za [1]) kowariancję z próby:

$$\hat{\gamma}(0) = 7,5541, \hat{\gamma}(1) = -5,1241, \hat{\gamma}(2) = 1,3805.$$

Poniżej przedstawiono wykresy funkcji autokorelacji i autokorelacji częściowej.



Wykres 4.4 Funkcja ACF dla Przykładu 4.3, za [1]



Wykres 4.5 Funkcja PACF dla Przykładu 4.3, za [1]

Sugerują one (a szczególnie wykres PACF) model średniej ruchomej rzędu 2.



Rzeczywiście, chociaż kilka wartości autokorelacji z próby dla lag większego niż 2, leży poza granicami  $\pm 1,96n^{-1/2}$  podanymi na wykresie, jednak żadna nie leży poza 95%-owymi granicami  $\pm 1,96n^{-1/2}(1 + 2\hat{\rho}^2(1) + 2\hat{\rho}^2(2))^{1/2}$ .

Stosując innowacyjny algorytm do rekurencyjnego wyznaczania oszacowań parametrów dla coraz to wyższych rzędów procesu średniej ruchomej, otrzymujemy (po skorzystaniu z (3.50) dla  $m$  oraz  $k = 0, 1, 2, \dots, m-1$ ), dla estymatorów w kolejnych rzędach  $m$ :

$$\begin{aligned}\hat{v}_0 &= 7,5541 \\ \hat{\theta}_{11} &= \hat{\rho}(1) = -0,67832 \\ \hat{v}_1 &= \hat{\gamma}(0) - \hat{\theta}_{11}^2 \hat{v}_0 = 4,0785 \\ \hat{\theta}_{22} &= \hat{v}_0^{-1} \hat{\gamma}(2) = 0,18275 \\ \hat{\theta}_{21} &= \hat{v}_1^{-1} (\hat{\gamma}(1) - \hat{\theta}_{22} \hat{\theta}_{11} \hat{v}_0) = -1,0268 \\ \hat{v}_2 &= \hat{\gamma}(0) - \hat{\theta}_{22}^2 \hat{v}_0 - \hat{\theta}_{21}^2 \hat{v}_1 = 3,0020\end{aligned}\tag{4.51}$$

Wykonując rekurencyjnie rachunki, dla większych wartości  $m$ , otrzymano wartości  $\hat{\theta}_{mj}$  i  $\hat{v}_m$ , ( $j=m-k=1, 2, \dots, m$ ), które zamieszczono w poniższej Tabelce.

$m \backslash j$	$\hat{\theta}_{mj}$										$\hat{v}_m$	
	1	2	3	4	5	6	7	8	9	10		
1	-0.68											4.08
2	-1.03	.18										3.00
3	-1.20	.37	.03									2.65
4	-1.31	.44	-.04	.07								2.40
5	-1.38	.51	-.03	-.04	.06							2.27
6	-1.41	.57	-.01	-.02	.10	-.02						2.24
7	-1.41	.60	-.01	-.02	.10	-.05	-.01					2.24
8	-1.41	.61	-.02	-.03	.10	-.07	-.02	.00				2.24
9	-1.41	.61	-.02	-.03	.10	-.08	-.02	.01	.01			2.24
10	-1.41	.61	-.02	-.02	.12	-.07	.00	.05	.04	.02		2.22
20	-1.43	.63	-.03	-.02	.11	-.08	.00	.03	.02	-.03		2.16
50	-1.43	.62	-.02	-.02	.12	-.08	.00	.03	.02	-.03		2.10
100	-1.43	.62	-.03	-.01	.11	-.08	-.01	.04	.01	-.03		2.00

Tabela.  $\hat{\theta}_{mj}$ ,  $j = 1, \dots, 10$  i  $\hat{v}_m$  dla danych z Przykładu [1].

Z tablicy tej wynika, że fluktuacje dla  $\hat{\theta}_{mj}$  od  $m=7$  do  $m=100$  są bardzo małe, rzędu  $n^{-1/2} = 1000^{-1/2} = 0,032$ .

Natomiast wartości dla  $\hat{\theta}_{7j}, j=1,2,\dots,7$  zawierają się dla  $j > 2$  (wartości z Tabelki w wierszu dla  $m=7$ : -0.01, -0.02, 0.10, -0.05, -0.01), wewnątrz przedziału  $\pm 1.96n^{-1/2}(\sum_{k=0}^{j-1} \hat{\theta}_{7k}^2)^{1/2}$ , co potwierdza, że MA(2) jest właściwym modelem.

Zatem ostatecznie model dopasowany do danych na podstawie estymatora  $\hat{\theta}_7$  jest następujący:

$$X_t = Z_t - 1,41Z_{t-1} + 0,6Z_{t-2}, \{Z_t\} \sim WN(0;2,24) \quad (4.52)$$

gdzie  $\hat{\theta}_{71} = -1,41$ ,  $\hat{\theta}_{72} = 0,60$ . Z Tabeli również widać, że estymowane współczynniki  $\hat{\theta}_{m1}$  i  $\hat{\theta}_{m2}$  wykazują bardzo małe zmiany dla  $m$  pomiędzy 7 a 100.

### 4.3 Wstępna estymacja parametrów procesów ARMA ( $p, q$ )

Niech  $\{X_t\}$  będzie przyczynowym procesem ARMA ( $p, q$ ) ze średnią równą zero:

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}, \quad \{Z_t\} \sim WN(0, \sigma^2), \quad (1.8')$$

Założenie przyczynowości (1.49) zapewnia, że:  $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, t = 0, \pm 1, \pm 2, \dots$ , gdzie  $\psi_j$

spełnia warunki (1.115-1.117):

$$\begin{cases} \psi_0 = 1 \\ \psi_j = \theta_j + \sum_{i=1}^{\min(j,p)} \phi_i \psi_{j-i} \quad \text{dla } j = 1, 2, \dots \end{cases} \quad (4.53)$$

oraz według przyjętej konwencji,  $\theta_j = 0$  dla  $j > q$  i  $\phi_j = 0$  dla  $j > p$ .

Dla oszacowania  $\psi_1, \dots, \psi_{p+q}$  możemy użyć estymatorów  $\hat{\theta}_{m1}, \dots, \hat{\theta}_{m,p+q}$  otrzymanych metodą algorytmu innowacyjnego, których asymptotyczne zachowanie opisane jest w Twierdzeniu 4.5.

Zastępując wstępnie w równaniu (4.53)  $\psi_j$  przez  $\hat{\theta}_{mj}$  (czyli wpieryw zaczynamy estymować stronę MA w (1.8')), otrzymujemy:

$$\hat{\theta}_{mj} = \theta_j + \sum_{i=1}^{\min(j,p)} \phi_i \hat{\theta}_{m,j-i}, \quad j = 1, 2, \dots, p+q \quad (4.54)$$

Rozwiązując powyższe równania otrzymujemy wstępne estymatory  $\hat{\phi}$  oraz  $\hat{\theta}$  dla parametrów  $\vec{\phi}$  i  $\vec{\theta}$ .

Z równań (4.54) dla  $j = q+1, \dots, q+p$  (zatem  $\theta_j = 0$ ) widzimy, że  $\bar{\hat{\phi}}$  powinien spełniać warunki równania:

$$\begin{bmatrix} \hat{\theta}_{m,q+1} \\ \hat{\theta}_{m,q+2} \\ \vdots \\ \hat{\theta}_{m,q+p} \end{bmatrix} = \begin{bmatrix} \hat{\theta}_{m,q} & \hat{\theta}_{m,q-1} & \cdots & \hat{\theta}_{m,q+1-p} \\ \hat{\theta}_{m,q+1} & \hat{\theta}_{m,q} & \cdots & \hat{\theta}_{m,q+2-p} \\ \vdots & \vdots & \vdots & \vdots \\ \hat{\theta}_{m,q+p-1} & \hat{\theta}_{m,q+p-2} & \cdots & \hat{\theta}_{m,q} \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix} \quad (4.55)$$

Rozwiązując (4.55) ze względu na  $\bar{\hat{\phi}}$  otrzymujemy estymator  $\hat{\phi}$  (rozwiązanie to nie musi być przyczynowe [1]).

Estymator dla  $\bar{\theta}$  znajdujemy wówczas z (4.54) :

$$\hat{\theta}_j = \hat{\theta}_{mj} - \sum_{i=1}^{\min(j,p)} \phi_i \hat{\theta}_{m,j-i}, \quad j = 1, 2, \dots, q. \quad (4.56)$$

Ostatecznie estymator wariancji białego szumu  $\sigma^2$  otrzymamy z  $\hat{\sigma}^2 = \hat{\nu}_m$  [1] gdy wykorzystać (3.50) dla estymatorów (por. (4.23) dla AR i (4.46) dla MA).

Można zauważyć, że powyższa metoda dla procesu ARMA (0, q) redukuje się do metody estymacji procesu MA(q) opisanej powyżej w Rozdziale 4.2.

#### 4.4 Estymatory metody największej wiarygodności i najmniejszych kwadratów dla procesów ARMA (p, q)

Załóżmy, że  $\{X_t\}$  jest przyczynowym procesem ARMA (p, q) spełniającym równanie (1.8):

$$X_t = \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} + \theta_0 Z_t + \dots + \theta_q Z_{t-q} \quad (4.57)$$

gdzie  $\{Z_t\} \sim \text{WN}(0, \sigma^2)$  oraz  $\theta_0 = 1$ .

Założenie przyczynowości oznacza (Twierdzenie 1.6), że:

$$1 - \phi_1 z - \dots + \phi_p z^p \neq 0 \quad \text{dla } |z| \leq 1. \quad (4.58)$$

Aby uniknąć niejednoznaczności (np. związanych z brakiem odwracalności procesu) przyjmujemy też, że współczynniki  $\theta_i$  oraz wariancja białego szumu  $\sigma^2$  zostaną tak dopasowane (bez naruszenia funkcji autokowariancji dla  $\{X_t\}$ ) by zagwarantować, że (Twierdzenie 1.8):

$$\theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q \neq 0, \quad \text{dla } |z| < 1. \quad (4.59)$$

Na początku znajdziemy estymatory MNW wektorów parametrów  $\vec{\phi} = (\phi_1, \dots, \phi_p)^T$ ,  $\vec{\theta} = (\theta_1, \dots, \theta_q)^T$  oraz wariancji białego szumu  $\sigma^2$ .

Można pokazać (Rozdział 3), że przewidywanie jednego kroku  $\hat{X}_{t+1}$  i jego średni kwadratowy błąd są dane przez następujące wyrażenia (3.99, 3.100):

$$\begin{cases} \hat{X}_{t+1} = \sum_{j=1}^i \theta_{ij} (X_{i+1-j} - \hat{X}_{i+1-j}) & , \quad 1 \leq i < m = \max(p, q) \\ \hat{X}_{t+1} = \phi_1 X_i + \dots + \phi_p X_{i+1-p} + \sum_{j=1}^q \theta_{ij} (X_{i+1-j} - \hat{X}_{i+1-j}) & , \quad i \geq m \end{cases} \quad (3.99')$$

$$\text{oraz: } v_i = E(X_{i+1} - \hat{X}_{i+1})^2 = \sigma^2 r_i \quad , \quad (3.100')$$

gdzie  $\theta_{ij}$  i  $r_i$  otrzymuje się przez zastosowywanie algorytmu innowacyjnego (3.49), (3.50). Ważna jest również informacja iż  $\theta_{ij}$  i  $r_i$  są niezależne od  $\sigma^2$  (Uwaga pod koniec Rozdziału 3.7).

Załóżmy, że  $\{X_i\}$  jest procesem Gaussa ze średnią równą zero i funkcją autokowariancji  $\kappa(i, j) = E(X_i, X_j)$ . Wstawiając (3.100') do (7.23) (Uzupełnienie 2), otrzymujemy Gaussowską wiarygodność wektora obserwacji  $\vec{X}_n = (X_1, \dots, X_n)^T$ , równą:

$$L(\vec{\phi}, \vec{\theta}, \sigma^2) = (2\pi\sigma^2)^{-n/2} (r_0 \cdots r_{n-1})^{-1/2} \exp\left[-\frac{1}{2}\sigma^{-2} \sum_{j=1}^n (X_j - \hat{X}_j)^2 / r_{j-1}\right] \quad (4.60)$$

skąd :

$$\ln L(\vec{\phi}, \vec{\theta}, \sigma^2) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2} \sum_{j=1}^{n-1} \ln r_j - \frac{1}{2} \sigma^{-2} \sum_{j=1}^n (X_j - \hat{X}_j)^2 / r_{j-1} \quad (4.61)$$

Różniczkując  $\ln L(\vec{\phi}, \vec{\theta}, \sigma^2)$  względem  $\sigma^2$ , pamiętając przy tym, że  $\hat{X}_j$  i  $r_i$  są niezależne od  $\sigma^2$ , otrzymujemy:

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2} \sigma^{-2} + \frac{1}{2} \sigma^{-4} \sum_{j=1}^n (X_j - \hat{X}_j)^2 / r_{j-1} = 0 \quad (4.62)$$

skąd:

$$\hat{\sigma}^2 \equiv \sigma^2 = \frac{1}{n} \sum_{j=1}^n (X_j - \hat{X}_j)^2 / r_{j-1} \equiv \frac{1}{n} S(\hat{\phi}, \hat{\theta}) \quad (4.63)$$

gdzie 
$$S(\hat{\phi}, \hat{\theta}) = \sum_{j=1}^n (X_j - \hat{X}_j)^2 / r_{j-1} \quad (4.64)$$

gdzie  $\hat{\phi}, \hat{\theta}$  są wartościami  $\vec{\phi}, \vec{\theta}$ , które maksymalizują  $\ln L$ , więc również  $L$ .

**Uwaga.**

Rozwiązujemy nie tylko równanie (4.62), ale cały układ równań wiarygodności:

$$\begin{cases} \frac{\partial \ln L}{\partial \sigma^2} = 0 \\ \frac{\partial \ln L}{\partial \vec{\phi}} = 0 \\ \frac{\partial \ln L}{\partial \vec{\theta}} = 0 \end{cases} \quad (4.65)$$

skąd otrzymujemy  $\hat{\sigma}^2, \hat{\phi}, \hat{\theta}$ .

Następnie sprawdzamy, czy istotnie w  $\hat{\sigma}^2, \hat{\phi}, \hat{\theta}$ , funkcja  $\ln L$  ma maksimum.

Zauważmy z (4.61), że maksymalizacja  $\ln L$  ze względu na  $\vec{\phi}, \vec{\theta}$  oznacza minimalizację ze względu na  $\vec{\phi}, \vec{\theta}$  funkcji:

$$\frac{1}{2} \sum_{j=1}^{n-1} \ln r_j + \frac{1}{2} \sigma^{-2} \sum_{j=1}^n (X_j - \hat{X}_j)^2 / r_{j-1} \quad (4.66)$$

oraz funkcji:

$$l(\vec{\phi}, \vec{\theta}) = \ln(n^{-1} S(\vec{\phi}, \vec{\theta})) + n^{-1} \sum_{j=1}^n \ln r_{j-1} \quad (4.67)$$

$l(\vec{\phi}, \vec{\theta})$  będziemy nazywali „zredukowaną wiarygodnością”.

**Uwaga.**

Wyznaczenie  $l(\vec{\phi}, \vec{\theta})$  powinno nastąpić z jednoczesnym wyznaczeniem metodą rekurencyjną parametrów  $\theta_{i-1,j}, r_{i-1}$  oraz przewidywań  $\hat{X}_i$ , dla dowolnej pary wektorów parametrów  $\vec{\phi}$  oraz  $\vec{\theta}$ , które minimalizują  $l(\vec{\phi}, \vec{\theta})$  (czyli dla estymatorów tych parametrów wyznaczonych metodą największej wiarygodności). Oznacza to w praktyce konieczność skorzystania z rachunków komputerowych.

W celu przyspieszenia stosowanych komputerowych procedur numerycznych, najlepiej zacząć z wartościami parametrów  $\vec{\phi}_0, \vec{\theta}_0$ , które leżą blisko minimum dla  $l(\vec{\phi}, \vec{\theta})$ . Skąd konieczne staje się umiejętne oszacowanie parametrów  $\vec{\phi}, \vec{\theta}$  omówione w Rozdziałach 4.1, 4.2 i 4.3.

**Uwaga.**

SAS przeprowadza całą analizę samodzielnie, bez konieczności (a nawet możliwości) wstępnego, tzn. ręcznego oszacowania parametrów  $\vec{\phi}$  i  $\vec{\theta}$ .

**Metoda najmniejszych kwadratów.**

Alternatywną procedurą estymacji jest minimalizacja ważonej sumy kwadratów:

$$S(\vec{\phi}, \vec{\theta}) = \sum_{j=1}^n (X_j - \hat{X}_j)^2 / r_{j-1} \quad (4.67)$$

ze względu na  $\vec{\phi}$  i  $\vec{\theta}$ .

Estymatory otrzymane w ten sposób będziemy nazywali estymatorami metody najmniejszych kwadratów (MNK) parametrów  $\vec{\phi}$  i  $\vec{\theta}$ . Oznaczone będą przez  $\vec{\phi} \approx, \vec{\theta} \approx$ .

**Uwaga.**

Okazuje się, że minimalizacja  $S(\vec{\phi}, \vec{\theta})$  oznacza nie tylko ograniczenie się do takich  $\vec{\phi}$ , dla których proces  $\{X_t\}$  jest przyczynowy, ale i do takich wartości  $\vec{\theta}$ , dla których proces ten jest również odwracalny [1]. Bez warunku odwracalności nie istniałby w ogólności skończone wartości  $(\vec{\phi}, \vec{\theta})$ , dla których  $S(\vec{\phi}, \vec{\theta})$  osiągałoby minimum [1].

Okazuje się również, że ponieważ dla szeregów odwracalnych  $r_n \rightarrow 1$  (gdy  $n \rightarrow \infty$ ) (Uwaga na końcu Rozdziału 3.7), zatem w tym przypadku wyrażenie (4.66):

$$n^{-1} \sum_{j=1}^n \ln r_{j-1} \quad (4.68)$$

jest asymptotycznie zaniedbywalne w porównaniu z  $\ln S(\vec{\phi}, \vec{\theta})$ , i minimalizacja funkcji zredukowanej wiarygodności  $l(\vec{\phi}, \vec{\theta})$  jest równoważna z minimalizacją  $S(\vec{\phi}, \vec{\theta})$ , a estymatory MNK oraz MNW mają wtedy podobne właściwości asymptotyczne.

Estymator  $\tilde{\sigma}_{LS}^2$  MNK parametru okazuje się mieć wtedy postać [1]:

$$\tilde{\sigma}_{LS}^2 = (n - p - q)^{-1} S(\tilde{\phi}, \tilde{\theta}) \quad (4.69)$$

gdzie dzielnik  $(n - p - q)$  pojawia się podobnie jak w standardowej analizie regresji jako wynik tego, że statystyka:

$$\sigma^{-2} S(\tilde{\phi}, \tilde{\theta}) \quad (4.70)$$

ma w przybliżeniu rozkład chi – kwadrat z liczbą stopni swobody  $(n - p - q)$ .

## 4.5 Asymptotyczne własności estymatorów MNW i MNK

Jeśli  $\{X_t\}$  jest przyczynowym, odwracalnym procesem (1.8):

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}, \quad Z_t \sim \text{iid}(0, \sigma^2) \quad (4.71)$$

i jeśli  $\phi(\cdot)$ ,  $\theta(\cdot)$  nie mają wspólnych zer, wtedy estymator największej wiarygodności:

$$\hat{\beta}^T = (\hat{\phi}_1, \dots, \hat{\phi}_p; \hat{\theta}_1, \dots, \hat{\theta}_q) = (\hat{\phi}^T, \hat{\theta}^T) \quad (4.72)$$

nazywa się przyczynową, odwracalną wartością dla  $\bar{\beta}^T = (\bar{\phi}^T, \bar{\theta}^T)$ , która minimalizuje zredukowaną wiarygodność  $l(\bar{\phi}, \bar{\theta})$ , określoną w (4.67).

Natomiast estymatory  $\tilde{\phi}, \tilde{\theta}$  metody najmniejszych kwadratów są przyczynowymi, odwracalnymi wartościami dla  $\bar{\phi}$  i  $\bar{\theta}$ , które minimalizują wielkość (por. (4.67)):

$$\ln(n^{-1} S(\bar{\phi}, \bar{\theta})) = l(\bar{\phi}, \bar{\theta}) - n^{-1} \sum_{j=1}^n \ln r_{j-1} \quad (4.73)$$

Jak wspomnieliśmy, w przypadku odwracalności procesu wyrażenie

$$n^{-1} \sum_{j=1}^n \ln r_{j-1} \quad (4.74)$$

jest zaniedbywalne asymptotycznie (dla  $n \rightarrow \infty$ ), a estymatory  $\tilde{\phi}, \tilde{\theta}$  mają asymptotycznie takie same własności jak  $\hat{\phi}$  i  $\hat{\theta}$ .

Poniższe Twierdzenie 4.6 [1] o asymptotycznym rozkładzie estymatora  $\hat{\beta}$  (4.72), pozwala na wyznaczenie asymptotycznych obszarów ufności dla  $\bar{\phi}$  i  $\bar{\theta}$  zarówno z estymatorów MNW i MNK.

### **Twierdzenie 4.6**

Jeśli  $\{Z\} \sim \text{iid}(0, \sigma^2)$  i  $\phi(\cdot)$  oraz  $\theta(\cdot)$  są przyczynowe, odwracalne i nie mają wspólnych zer wtedy:

$$n^{1/2}(\hat{\beta} - \bar{\beta}) \Rightarrow N(0, V(\bar{\beta})) \quad (4.75)$$

gdzie asymptotyczna macierz kowariancji  $V(\bar{\beta})$  ma dla  $p \geq 1$  i  $q \geq 1$  postać:

$$V(\bar{\beta}) = \sigma^2 \begin{bmatrix} E\bar{U}_t\bar{U}_t^T & E\bar{U}_t\bar{V}_t^T \\ E\bar{V}_t\bar{U}_t^T & E\bar{V}_t\bar{V}_t^T \end{bmatrix}^{-1} \quad (4.76)$$

gdzie:

$$\bar{U}_t = (U_t, \dots, U_{t+1-p})^T \quad (4.77a)$$

$$\bar{V}_t = (V_t, \dots, V_{t+1-q})^T \quad (4.77b)$$

natomiast  $\{U_t\}$  i  $\{V_t\}$  są procesami autoregresji :

$$\phi(B)U_t = Z_t \quad (4.78a)$$

$$\theta(B)V_t = Z_t \quad (4.78b)$$

przy czym dla  $p = 0$  :

$$V(\bar{\beta}) = \sigma^2 [E\bar{V}_t\bar{V}_t^T]^{-1} \quad (4.79a)$$

a dla  $q = 0$  :

$$V(\bar{\beta}) = \sigma^2 [E\bar{U}_t\bar{U}_t^T]^{-1} \quad (4.79b)$$

### **Przykład 4.4**

Dla procesu AR( $p$ ) z równania (4.76) mamy:

$$V(\vec{\phi}) = \sigma^2 [E\bar{U}_t\bar{U}_t^T]^{-1} \quad (4.80)$$

gdzie  $\phi(B)U_t = U_t - \phi_1 U_{t-1} - \dots - \phi_p U_{t-p} = Z_t$  (4.78a')

Równanie (4.80) możemy zapisać jako:

$$V(\vec{\phi}) = \sigma^2 \Gamma_p^{-1} \quad (4.81)$$

gdzie ze względu na to, że procesy  $\{U_t\}$  i  $\{X_t\}$  mają takie same  $\vec{\phi}$ , mamy:

$$\Gamma_p = E(\bar{U}_t\bar{U}_t^T) = [EX_i X_j]_{i,j=1}^p \quad (4.82)$$

W końcu z (4.12) mamy:

$$\hat{\vec{\phi}} \sim AN(\vec{\phi}, n^{-1} \sigma^2 \Gamma_p^{-1}) \quad (4.83)$$



W przypadkach gdy  $p = 1$  otrzymujemy (por. 1.40b):

$$\text{AR}(1): \hat{\phi} \sim AN(\phi, n^{-1}(1 - \phi^2)) \quad (4.84)$$

Natomiast dla  $p = 2$ , wykorzystując (1.40b) w  $\Gamma_2 = \begin{pmatrix} \gamma(0) & \gamma(-1) \\ \gamma(1) & \gamma(0) \end{pmatrix}$  i odwróceniu macierzy  $\Gamma_2$ , otrzymujemy:

$$\text{AR}(2): \begin{bmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \end{bmatrix} \sim AN \left( \begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix}, n^{-1} \begin{bmatrix} 1 - \phi_2^2 & -\phi_1(1 + \phi_2) \\ -\phi_1(1 + \phi_2) & 1 - \phi_2^2 \end{bmatrix} \right) \quad (4.85)$$

### **Przykład 4.5**

Dla procesu MA( $q$ ) z równania (4.76) otrzymujemy:

$$V(\vec{\theta}) = \sigma^2 [E\vec{V}_t\vec{V}_t^T]^{-1} \quad (4.86)$$

gdzie  $\theta(B)V_t = Z_t$  (4.78b')

Równanie (4.86) możemy zapisać jako:

$$V(\vec{\phi}) = \sigma^2 [\Gamma_q^*]^{-1} \quad (4.87)$$

gdzie:  $\Gamma_q^* = [EV_i V_j]_{i,j=1}^q$  (4.88)

jest macierzą autokowariancji procesu autoregresyjnego:

$$V_t + \theta_1 V_{t-1} + \dots + \theta_q V_{t-q} = Z_t. \quad (4.89)$$

Korzystając z poprzedniego Przykładu 4.4 (jedynie znak przed parametrami  $\theta_j$  w (4.89) i  $\phi_j$  w (4.78a') jest przeciwny), otrzymujemy w przypadkach gdy:

$q = 1$       MA(1):  $\hat{\theta} \sim AN(\theta, n^{-1}(1 - \theta^2))$  (4.90)

$q = 2$       MA(2):  $\begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix} \sim AN \left( \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}, n^{-1} \begin{bmatrix} 1 - \theta_2^2 & \theta_1(1 - \theta_2) \\ \theta_1(1 - \theta_2) & 1 - \theta_2^2 \end{bmatrix} \right)$  (4.91)

### **Przykład 4.6**

Dla procesu ARMA(1,1) z równania (4.76) otrzymujemy:

$$V(\phi, \theta) = \sigma^2 \begin{bmatrix} EU_t^2 & EU_t V_t \\ EV_t U_t & EV_t^2 \end{bmatrix}^{-1} \quad (4.92)$$

gdzie:  $U_t - \phi U_t = Z_t$  (4.93)

oraz  $V_t - \theta V_t = Z_t$  (4.94)

skąd po prostych obliczeniach otrzymujemy:

$$V(\phi, \theta) = \sigma^2 \begin{bmatrix} (1-\phi^2)^{-1} & (1+\phi\theta)^{-1} \\ (1+\phi\theta)^{-1} & (1-\theta^2)^{-1} \end{bmatrix}^{-1} \quad (4.95)$$

gdzie :

$$\begin{bmatrix} \hat{\phi} \\ \hat{\theta} \end{bmatrix} \sim AN \left( \begin{bmatrix} \phi \\ \theta \end{bmatrix}, n^{-1} \frac{1+\phi\theta}{(\phi+\theta)^2} \begin{bmatrix} (1-\phi^2)(1+\phi\theta) & -(1-\theta^2)(1-\phi^2) \\ -(1-\theta^2)(1-\phi^2) & (1-\theta^2)(1+\phi\theta) \end{bmatrix} \right) \quad (4.96)$$

## **4.6 Przedziały ufności dla parametrów przyczynowych i odwracalnych procesów ARMA [1].**

Obszary ufności dla wektorów współczynników  $\vec{\phi}$  i  $\vec{\theta}$  dla przyczynowych i odwracalnych procesów ARMA, mogą być w przypadku dużych próbek wyznaczone z asymptotycznych rozkładów estymatorów MNW, w sposób analogiczny do podanego poprzednio w Rozdziale 4.1 asymptotycznego rozkładu estymatora Yule'a – Walker'a dla  $\vec{\phi}$ .

Niech  $\vec{\beta}^T = (\vec{\phi}^T, \vec{\theta}^T)$  i niech  $\vec{\beta}$  jest estymatorem MNW  $\vec{\beta}$ . Wtedy definiując  $V(\vec{\beta})$  tak jak w (4.76), przybliżony (dokładny asymptotycznie) obszar ufności dla  $\vec{\beta}$ , ma na poziomie ufności  $1-\alpha$  postać, którą można symbolicznie zapisać jako:

$$\left\{ \vec{\beta} \in \mathbb{R}^{p+q} : (\vec{\beta} - \hat{\vec{\beta}})^T V(\hat{\vec{\beta}}) (\vec{\beta} - \hat{\vec{\beta}}) \leq n^{-1} \chi_{1-\alpha}^2(p+q) \right\} \quad (4.97)$$

Można również podać przybliżoną postać przedziałów ufności dla pojedynczego parametru (na poziomie ufności  $1-\alpha$ ). Niech  $v_{jj}(\hat{\vec{\beta}})$  będzie  $j$ -tym diagonalnym elementem macierzy  $V(\hat{\vec{\beta}})$ .

Wtedy  $(1 - \alpha) \cdot 100\%$  - owy przedział ufności dla  $\beta_j$  ma symboliczną postać:

$$\left\{ \beta \in \mathbb{R} : |\beta - \hat{\beta}_j| \leq n^{-1/2} u_\alpha v_{jj}^{1/2}(\hat{\beta}) \right\} \quad (4.98)$$

gdzie  $u_\alpha = \Phi(1 - \alpha/2)$  jest wartością kwantyla rzędu  $1 - \frac{\alpha}{2}$  z rozkładu  $N(0,1)$ .

## 4.7 Asymptotyczna normalność estymatorów parametrów

W poniższym rozdziale omówimy dla przyczynowych, odwracalnych procesów ARMA( $p, q$ ), własność asymptotycznej normalności estymatora wektora współczynników, który ma asymptotycznie ten sam rozkład co estymatory MNK i MNW. Asymptotyczne rozkłady estymatorów MNK i MNW są podane w [1].

W Rozdziale 4.4 stwierdziliśmy, że estymatory MNK minimalizują sumę kwadratów (4.64):

$$S(\vec{\phi}, \vec{\theta}) = \sum_{t=1}^n (X_t - \hat{X}_t)^2 / r_{t-1} \quad (4.64')$$

Poniżej rozważymy pewne użyteczne przybliżenie dla  $S(\vec{\phi}, \vec{\theta})$ . Wpierw przybliżmy „standardowe innowacje”  $(X_t - \hat{X}_t)/(r_{t-1})^{1/2}$  poprzez poniższy układ rekurencyjnych równań:

$$\left\{ \begin{array}{l} Z_1(\vec{\phi}, \vec{\theta}) = X_1, \\ Z_2(\vec{\phi}, \vec{\theta}) = X_2 - \phi_1 X_1 - \theta_1 Z_1(\vec{\phi}, \vec{\theta}) \\ \vdots \\ Z_n(\vec{\phi}, \vec{\theta}) = X_n - \phi_1 X_{n-1} - \dots - \phi_p X_{n-p} - \theta_1 Z_{n-1}(\vec{\phi}, \vec{\theta}) - \dots - \theta_q Z_{n-q}(\vec{\phi}, \vec{\theta}) \end{array} \right. \quad (4.99)$$

Z powodu założonej odwracalności możemy zapisać zmienne białego szumu  $Z_t$  w postaci (1.81):

$$Z_t = X_t + \sum_{j=1}^{\infty} \pi_j X_{t-j}, \quad (4.50)$$

a jak można pokazać [1] układ (4.99), odpowiada poniższemu:

$$Z_t(\vec{\phi}, \vec{\theta}) = X_t + \sum_{j=1}^{t-1} \pi_j X_{t-j} \quad (4.51)$$

Można również pokazać [1], że zachodzą nierówności:

$$\begin{aligned} \|Z_t(\vec{\phi}, \vec{\theta}) - Z_t\| &\leq c_1 a^t \\ \|Z_t(\vec{\phi}, \vec{\theta}) - (X_t - \hat{X}_t)/(r_{t-1})^{1/2}\| &\leq c_2 a^t \\ \|Z_t(\vec{\phi}, \vec{\theta})\| &\leq k \end{aligned} \quad (4.52)$$

gdzie  $a, c_1, c_2$  i  $k$  są stałymi oraz  $0 < a < 1$ .

Okazuje się, że pożytecznie jest wykonać dalsze przybliżenie dla  $(X_t - \hat{X}_t)/(r_{t-1})^{1/2}$  poprzez linearyzację  $Z_t(\vec{\phi}, \vec{\theta})$  wokół (korzystnie wybranych) początkowych oszacowań  $(\vec{\phi}_0, \vec{\theta}_0)$  dla  $(\vec{\phi}, \vec{\theta})$ . Jeśli zatem  $\vec{\beta}^T = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)$  i  $\vec{\beta}_0^T = (\vec{\phi}_0^T, \vec{\theta}_0^T)$ , wtedy przybliżamy  $Z_t(\vec{\beta})$  poprzez :

$$Z_t(\vec{\beta}_0) - \vec{D}_t^T (\vec{\beta} - \vec{\beta}_0), \quad (4.53)$$

gdzie  $\vec{D}_t^T = (D_{t,1}(\vec{\beta}_0), \dots, D_{t,p+q}(\vec{\beta}_0))$  , (4.54)

natomiast:

$$D_{t,i}(\vec{\beta}) = -\frac{\partial Z_t(\vec{\beta})}{\partial \beta_i}, \quad i = 1, 2, \dots, p+q. \quad (4.55)$$

Minimalizując sumę kwadratów:

$$\sum_{t=1}^n (Z_t(\vec{\beta}_0) - \vec{D}_t^T (\vec{\beta} - \vec{\beta}_0))^2, \quad (4.56)$$

która ze względu na związki (4.52) jest dobrym przybliżeniem dla  $S(\vec{\phi}, \vec{\theta})$ , otrzymujemy pewien estymator  $\vec{\beta}^*$  dla  $\vec{\beta}$ , który ma takie same asymptotyczne własności jak estymator  $\tilde{\vec{\beta}}$  MNK.

Wprowadźmy oznaczenia

$$\vec{Z}(\vec{\beta}_0) = (Z_1(\vec{\beta}_0), \dots, Z_n(\vec{\beta}_0))^T \quad (4.57)$$

oraz  $n \times (p+q)$  wymiarową macierz:

$$D = (\vec{D}_1, \dots, \vec{D}_n)^T, \quad (4.58)$$

Można pokazać, że estymatorem regresji liniowej dla

$$\Delta\vec{\beta} = \vec{\beta} - \vec{\beta}_0 \quad (4.59)$$

jest  $\hat{\Delta}\vec{\beta} = (D^T D)^{-1} D^T \vec{Z}(\vec{\beta}_0)$  (4.60)

tak, że  $\vec{\beta}^* = \vec{\beta}_0 + \hat{\Delta}\vec{\beta}$  (4.61)

Asymptotyczna normalność estymatora ustanowiona jest w poniższym twierdzeniu.

### **Twierdzenie 4.7**

Niech  $\{X_t\}$  jest przyczynowym, odwracalnym procesem ARMA( $p, q$ ):

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}, \quad \{Z_t\} \sim \text{iid}(0, \sigma^2), \quad (1.8')$$

a  $\phi(z)$  i  $\theta(z)$  nie mają wspólnych zer.

Założmy, że  $\vec{\beta}_0 = (\beta_{01}, \dots, \beta_{0,p+q})^T$  jest wstępnym oszacowaniem dla  $\vec{\beta} = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)^T$  takim, że  $\vec{\beta}_0 - \vec{\beta} = o_p(n^{-1/4})$ , i że  $\vec{\beta}^*$  jest estymatorem utworzonym z  $\vec{\beta}_0$  jak to opisano powyżej w (4.61).

Wtedy:

$$(i) \quad n^{-1} D^T D \xrightarrow{P} \sigma^2 V^{-1}(\vec{\beta}) \quad (4.62)$$

gdzie  $V(\vec{\beta})$  jest  $(p+q) \times (p+q)$  nieosobliwą macierzą

$$(ii) \quad n^{1/2}(\vec{\beta}^* - \vec{\beta}) \Rightarrow N(\vec{0}, V(\vec{\beta})) \quad (4.63)$$

oraz jak wiemy, dla estymatorów MNK, mamy:

$$(iii) \quad n^{1/2}(\tilde{\vec{\beta}} - \vec{\beta}) \Rightarrow N(\vec{0}, V(\vec{\beta})) \quad (4.64)$$

Macierz  $V(\vec{\beta})$  ma postać:

$$V(\vec{\beta}) = \sigma^2 \begin{bmatrix} E\vec{U}_1\vec{U}_1^T & E\vec{U}_1\vec{V}_1^T \\ E\vec{V}_1\vec{U}_1^T & E\vec{V}_1\vec{V}_1^T \end{bmatrix}^{-1} \quad (4.65)$$

gdzie  $\{U_t\}$  i  $\{V_t\}$  są procesami autoregresji :

$$U_t = \theta^{-1}(B)X_t = \phi^{-1}(B)Z_t \quad (4.66)$$

$$V_t = \theta^{-1}(B)Z_t \quad (4.67)$$

Powyższe określenia i twierdzenie są istotne przy konstrukcji kryterium AICC w Rozdziale 5.2

## Rozdział 5. Kryteria wyboru rzędu parametrów $p$ i $q$

### 5.1 Kryterium FPE

We wstępie przedstawiliśmy problem przeffitowania i potrzeby jego uniknięcia „nakładając cenę” na zwiększenie liczby parametrów dopasowanego modelu. Jednym z wyjść uniknięcia przeffitowania w prostych modelach autoregresyjnych, jest minimalizacja końcowego błędu przewidywania FPE (final prediction error) zaproponowana przez Akaike’a [7].

FPE jest oszacowaniem średnio – kwadratowego błędu predykcji w jednym kroku dla realizacji procesu, która jest niezależna od realizacji właśnie obserwowanej. Jeśli dopasujemy proces  $AR(p)$  o stale rosnącym rzędzie  $p$ , do obserwowanych danych, wtedy oszacowanie MNW dla wariancji białego szumu będzie zazwyczaj spadać wraz ze wzrostem  $p$ . Jednakże błędy oszacowań w poszerzonym ciągle zbiorze dopasowanych parametrów, spowodują w końcu wzrost FPE. Zgodnie z kryterium FPE wybieramy wtedy jako rząd dopasowanego modelu tą wartość  $p$ , dla której FPE jest najmniejsze. Aby zastosować to kryterium, należy tylko wyrazić FPE poprzez dane  $X_1, \dots, X_n$ .

Niech  $\{X_1, \dots, X_n\}$  będzie realizacją procesu  $AR(p)$  z współczynnikami  $\phi_1, \dots, \phi_p$  ( $p < n$ ) oraz niech  $\{Y_1, \dots, Y_n\}$  będzie niezależną od poprzedniego procesu realizacją tego samego procesu  $AR(p)$ . Jeśli  $\hat{\phi}_1, \dots, \hat{\phi}_p$  są estymatorami MNW współczynników dla realizacji  $\{X_1, \dots, X_n\}$  i jeśli użyjemy ich do wyznaczenia predykcji w jednym kroku  $\hat{\phi}_1 Y_n + \dots + \hat{\phi}_p Y_{n+1-p}$  dla  $Y_{n+1}$ , wtedy średnio – kwadratowy błąd predykcji wynosi:

$$E\left(Y_{n+1} - \hat{\phi}_1 Y_n - \dots - \hat{\phi}_p Y_{n+1-p}\right)^2 =$$

$$= E\left(\{Y_{n+1} - \phi_1 Y_n - \dots - \phi_p Y_{n+1-p}\} - \{(\hat{\phi}_1 - \phi_1) Y_n + \dots + (\hat{\phi}_p - \phi_p) Y_{n+1-p}\}\right)^2 = \quad (5.1a)$$

$$= \sigma^2 + E\left[\left(\hat{\phi}_p - \bar{\phi}_p\right)^T \left[Y_{n+1-i} Y_{n+1-j}\right]_{i,j=1}^p \left(\hat{\phi}_p - \bar{\phi}_p\right)\right] \quad (5.1b)$$

gdzie  $\bar{\phi}_p^T = (\phi_1, \dots, \phi_p)$ ,  $\hat{\phi}_p^T = (\hat{\phi}_1, \dots, \hat{\phi}_p)$ , natomiast  $\sigma^2$  jest wariancją białego szumu modelu  $AR(p)$ . W (5.1a) skorzystano z faktu, że dwa wyrażenia  $\{ \}$  i  $\{ \}$  pod  $E$ , są nieskorelowane.

Korzystając dla drugiego składnika w (5.1b) z niezależności  $\{X_1, \dots, X_n\}$  i  $\{Y_1, \dots, Y_n\}$ , możemy zależność (5.1) zapisać następująco:

$$E\left(Y_{n+1} - \hat{\phi}_1 Y_n - \dots - \hat{\phi}_p Y_{n+1-p}\right)^2 = \sigma^2 + E\left[\left(\hat{\phi}_p - \bar{\phi}_p\right)^T \Gamma_p \left(\hat{\phi}_p - \bar{\phi}_p\right)\right] \quad (5.2)$$

gdzie :  $\Gamma_p = E\left[Y_i Y_j\right]_{i,j=1}^p$ .

Możemy teraz przybliżyć ostatni człon poprzez założenie, że  $n^{1/2}\left(\hat{\phi}_p - \bar{\phi}_p\right)$  ma, zgodnie z Twierdzeniem 4.1, asymptotycznie rozkład  $N(0, \sigma^2 \Gamma_p^{-1})$ . (5.3)

Otrzymujemy wtedy:

$$E\left(Y_{n+1} - \hat{\phi}_1 Y_n - \dots - \hat{\phi}_p Y_{n+1-p}\right)^2 \cong \sigma^2 \left(1 + \frac{p}{n}\right). \quad (5.4)$$

Jeśli  $\hat{\sigma}^2$  jest estymatorem MNW dla  $\sigma^2$ , wtedy  $n\hat{\sigma}^2 / \sigma^2$  ma dla dużych  $n$  w przybliżeniu rozkład rozkładem chi – kwadrat z  $(n - p)$  stopniami swobody [1]. Dlatego zastępujemy w (5.4),  $\sigma^2$  przez estymator  $n\hat{\sigma}^2 / (n - p)$ , aby otrzymać ostatecznie oszacowany średnio – kwadratowy błąd predykcji dla  $Y_{n+1}$ .

$$\text{FPE} \cong \hat{\sigma}^2 \frac{n + p}{n - p}. \quad (5.5)$$

Typowe zachowanie dla FPE jest takie, że początkowo, wraz ze wzrostem  $p$ , wartość FPE maleje, aby znowu zacząć rosnąć, przy dalszym wzroście  $p$ .

## 5.2 Kryterium AICC

Przy wyborze modelu jeszcze bardziej przydatnym kryterium niż FPE jest informacyjne kryterium AIC Akaike’a [7]. Zostało ono skonstruowane tak, aby być w przybliżeniu nieobciążonym estymatorem (zdefiniowanego poniżej) wskaźnika Kullback’a – Leibler’a dopasowanego modelu względem prawdziwego modelu.

Będziemy jednak używać zmodyfikowanej nieobciążonej wersji AIC, zwanej AICC, zasugerowanej przez Hurvich’a i Tsai’a [8].

Jeśli  $\vec{X}$  jest  $n$  wymiarowym wektorem losowym, którego gęstość prawdopodobieństwa należy do rodziny  $\{f(\cdot; \psi), \psi \in \Psi\}$ , wtedy różnica Kullback’a – Leibler’a między  $f(\cdot; \psi)$  i  $f(\cdot; \theta)$  jest zdefiniowana jako:

$$d(\psi | \theta) = \Delta(\psi | \theta) - \Delta(\theta | \theta) \quad (5.6)$$

gdzie:

$$\Delta(\psi | \theta) = E_{\theta}(-2 \ln f(\vec{X}; \psi)) = \int_{R^n} -2 \ln(f(\vec{x}; \psi)) f(\vec{x}; \theta) d\vec{x} \quad (5.7)$$

jest wskaźnikiem Kullback'a – Leibler'a dla  $f(\cdot; \psi)$  w stosunku do  $f(\cdot; \theta)$  (w ogólności  $\Delta(\psi | \theta) \neq \Delta(\theta | \psi)$ ).

Po zastosowaniu nierówności Jensen'a otrzymujemy [1]:

$$d(\psi | \theta) = \int_{R^n} -2 \ln \left( \frac{f(\vec{x}; \psi)}{f(\vec{x}; \theta)} \right) f(\vec{x}; \theta) d\vec{x} \geq -2 \ln \int_{R^n} \frac{f(\vec{x}; \psi)}{f(\vec{x}; \theta)} f(\vec{x}; \theta) d\vec{x} = -2 \ln \int_{R^n} f(\vec{x}; \psi) d\vec{x} = 0 \quad (5.8)$$

z równością zachodzącą wtedy i tylko wtedy, gdy  $f(\vec{x}; \theta) = f(\vec{x}; \psi)$  prawie zawsze z miarą  $[f(\cdot; \theta)]$ .

Mając obserwacje  $X_1, \dots, X_n$  z procesu ARMA z nieznanymi parametrami  $\theta = (\vec{\beta}, \sigma^2)$ , prawdziwy model może zostać zidentyfikowany, jeśli byłoby możliwe wyznaczenie różnicy Kullback'a – Leibler'a pomiędzy prawdziwym modelem a wszystkimi kandydującymi modelami. Ponieważ nie jest to możliwe, oszacujemy różnice Kullback'a – Leibler'a i wybierzemy model, dla którego oszacowana różnica (bądź wskaźnik) jest najmniejsza. W tym celu zakładamy, że prawdziwy model oraz wszystkie alternatywne do niego są gaussowskie. Wtedy dla każdego zadanego  $\theta = (\vec{\beta}, \sigma^2)$ , funkcja  $f(\cdot; \theta)$  jest gęstością prawdopodobieństwa dla  $(Y_1, \dots, Y_n)^T$ , gdzie  $\{Y_t\}$  jest gaussowskim procesem ARMA( $p, q$ ) z wektorem współczynników  $\vec{\beta}$ , oraz wariancją białego szumu  $\sigma^2$ . Zależność parametrów  $\theta$  od  $p$  i  $q$  ma miejsce poprzez wymiar wektora współczynników autoregresji i ruchomej średniej w  $\vec{\beta}$ .

Przypuśćmy, że nasze obserwacje  $X_1, \dots, X_n$  pochodzą z gaussowskiego procesu ARMA z parametrowym wektorem  $\theta = (\vec{\beta}, \sigma^2)$ , oraz założmy na chwile, że prawdziwy rząd modelu wynosi  $(p, q)$ . Niech  $\hat{\theta} = (\hat{\vec{\beta}}, \hat{\sigma}^2)$  będzie estymatorem MNW dla  $\theta$  wyznaczonym z  $X_1, \dots, X_n$ , i niech  $Y_1, \dots, Y_n$  będzie niezależną realizacją prawdziwego procesu (z parametrem  $\theta$ ).

Wtedy [1]:

$$-2 \ln L_Y(\hat{\vec{\beta}}, \hat{\sigma}^2) = -2 \ln L_X(\hat{\vec{\beta}}, \hat{\sigma}^2) + \hat{\sigma}^{-2} S_Y(\hat{\vec{\beta}}) - n \quad (5.9)$$

co daje:



$$E_{\theta}(\Delta(\hat{\theta} | \theta)) = E_{\vec{\beta}, \sigma^2}(-2 \ln L_Y(\hat{\vec{\beta}}, \hat{\sigma}^2)) = E_{\vec{\beta}, \sigma^2}(-2 \ln L_X(\hat{\vec{\beta}}, \hat{\sigma}^2)) + E_{\vec{\beta}, \sigma^2} \left( \frac{S_Y(\hat{\vec{\beta}})}{\hat{\sigma}^2} \right) - n \quad (5.10)$$

gdzie  $S_Y(\hat{\vec{\beta}})$  jest ważoną sumą kwadratów odchyleń postaci (4.64), ale dla realizacji procesu  $Y_1, \dots, Y_n$  przy parametrach oszacowanych dla procesu  $X_1, \dots, X_n$ .

Dokonując linearyzacji i posługując się oznaczeniami i wynikami z Rozdziału 4.7, możemy dla dużych  $n$  napisać:

$$\begin{aligned} S_Y(\hat{\vec{\beta}}) &\approx S_Y(\vec{\beta}) + (\hat{\vec{\beta}} - \vec{\beta}) \frac{\partial S_Y(\vec{\beta})}{\partial \vec{\beta}} + \frac{1}{2} (\hat{\vec{\beta}} - \vec{\beta})^T \left[ \frac{\partial^2 S_Y(\vec{\beta})}{\partial \beta_i \partial \beta_j} \right]_{i,j=1}^n (\hat{\vec{\beta}} - \vec{\beta}) \approx \\ &\approx S_Y(\vec{\beta}) + (\hat{\vec{\beta}} - \vec{\beta}) 2 \sum_{t=1}^n \frac{\partial Z_t(\vec{\beta})}{\partial \vec{\beta}} Z_t(\vec{\beta}) + (\hat{\vec{\beta}} - \vec{\beta})^T D^T D (\hat{\vec{\beta}} - \vec{\beta}). \end{aligned} \quad (5.11)$$

Ze względu na (4.62),  $n^{-1} D^T D \xrightarrow{P} \sigma^2 V^{-1}(\vec{\beta})$ , i ponieważ z Twierdzenia 4.6 wynika, że  $\hat{\vec{\beta}} \sim AN(\vec{\beta}, n^{-1} V(\beta))$ , oraz ponieważ  $\frac{\partial Z_t}{\partial \vec{\beta}}(\vec{\beta}) Z_t(\vec{\beta})$  ma średnią zero, zatem zastępując  $D^T D$  przez  $n \sigma^2 V^{-1}(\beta)$  i zakładając, że  $(\hat{\vec{\beta}} - \vec{\beta})$  ma macierz kowariancji  $V(\vec{\beta})$ , otrzymujemy dla warunkowej wartości oczekiwanej z (5.11):

$$E_{\vec{\beta}, \sigma^2} [S_Y(\hat{\vec{\beta}})] \approx E_{\vec{\beta}, \sigma^2} [S_Y(\vec{\beta})] + \sigma^2 E_{\vec{\beta}, \sigma^2} [(\hat{\vec{\beta}} - \vec{\beta})^T V^{-1}(\vec{\beta}) (\hat{\vec{\beta}} - \vec{\beta})] \approx \sigma^2 n + \sigma^2 (p + q) \quad (5.12)$$

gdzie skorzystano z faktu, że  $\frac{\partial Z_t}{\partial \vec{\beta}}(\vec{\beta}) Z_t(\vec{\beta})$  jest niezależne od  $\hat{\vec{\beta}} - \vec{\beta}$ , oraz z faktu, że

$E(\vec{U}^T \Sigma^{-1} \vec{U}) = Tr(\Sigma \Sigma^{-1}) = k$  dla każdego  $k$  – wymiarowego wektora losowego  $\vec{U}$  (ze średnią zero i z nieosobliwą macierzą kowariancji  $\Sigma$ ).

Statystyka  $n \hat{\sigma}^2 = S_X(\hat{\vec{\beta}})$  ma asymptotycznie (dla dużych  $n$ ), rozkład  $\sigma^2 \chi^2(n - p - q)$ , oraz jest asymptotycznie niezależna od  $\hat{\vec{\beta}}$ . Wraz z niezależnością  $\{X_1, \dots, X_n\}$  oraz  $\{Y_1, \dots, Y_n\}$  implikuje to nam warunek, że  $\hat{\sigma}^2$  jest asymptotycznie niezależne od  $S_Y(\hat{\vec{\beta}})$ .

W konsekwencji otrzymujemy:

$$E_{\bar{\beta}, \sigma^2} \left( \frac{S_Y(\hat{\beta})}{\hat{\sigma}^2} \right) - n \cong \sigma^2 (n + p + q) (E_{\bar{\beta}, \sigma^2} \hat{\sigma}^{-2}) - n \cong \sigma^2 (n + p + q) \left( \sigma^2 \frac{n - p - q - 2}{n} \right)^{-1} - n =$$

$$= \frac{2(p + q + 1)n}{n - p - q - 2}. \quad (5.13)$$

W ten sposób wielkość:

$$-2 \ln L_X(\hat{\beta}, \hat{\sigma}^2) + 2(p + q + 1)n / (n - p - q - 2)$$

jest w przybliżeniu nieobciążonym estymatorem wartości oczekiwanej wskaźnika Kullback'a – Leibler'a  $E_\theta(\Delta(\hat{\theta} | \theta))$ , (5.10).

Ponieważ powyższe obliczenia (oraz estymatory MNW  $\hat{\beta}$  oraz  $\hat{\sigma}^2$ ) są oparte na założeniu, iż prawdziwy model jest rzędu  $(p, q)$ , dlatego wybieramy wartości  $p$  i  $q$  dla naszego dopasowanego modelu jako te, które minimalizują  $AICC(\hat{\beta})$ , gdzie:

$$AICC(\bar{\beta}) := -2 \ln L_X(\bar{\beta}, S_X(\bar{\beta})/n) + 2(p + q + 1)n / (n - p - q - 2) \quad (5.14)$$

Statystyka AIC, zdefiniowana jako:

$$AIC(\bar{\beta}) := -2 \ln L_X(\bar{\beta}, S_X(\bar{\beta})/n) + 2(p + q + 1) \quad (5.15)$$

może być użyta w ten sam sposób.

Oba kryteria  $AICC(\bar{\beta}, \sigma^2)$  oraz  $AIC(\bar{\beta}, \sigma^2)$  mogą zostać zdefiniowane dla dowolnego  $\sigma^2$  poprzez zastąpienie w powyższych definicjach  $S_X(\bar{\beta})/n$  przez  $\sigma^2$ .

Jednakże korzystniej jest używać  $AICC(\bar{\beta})$  oraz  $AIC(\bar{\beta})$  tak jak zostały zdefiniowane powyżej. Wynika to stąd, że obie wartości kryteriów AICC i AIC są minimalizowane dla dowolnych  $\bar{\beta}$  przez podstawienie  $\sigma^2 = S_X(\bar{\beta})/n$ .

## **Zakończenie.**

Praca przedstawia procedurę modelowania szeregów czasowych za pomocą liniowych modeli  $ARMA(p, q)$ , oraz  $ARIMA(p, d, q)$ . Na początku przedstawiona została budowa oraz podstawowe właściwości procesów  $ARMA$ , wraz z identyfikacją „prostych” modeli  $AR(p)$ ,  $MA(q)$  za pomocą funkcji  $ACF$  i  $PACF$ . Omówione zostały również własności tych funkcji i ich estymacja z próby, gdyż są one pomocne w interpretacji autokorelogramów oraz przy wyborze wartości parametrów  $p$  i  $q$ . Wybór parametrów naszego modelu został rozszerzony o kryteria  $AICC$  oraz  $FPE$ , które dzięki swoim własnościom naprowadzają nas na wybór odpowiedniego modelu (Rozdział 5). Praca ta została rozszerzona również o podstawowe wiadomości analizy spektralnej szeregów czasowych, co umożliwi nam dalszy rozwój pracy w tym kierunku.

Bardzo ważnym zagadnieniem jest pojęcie stacjonarności szeregu gdyż w przypadku, gdy szereg jest niestacjonarny (występuje trend lub sezonowość) wnioskowanie nie jest możliwe, dlatego też podano procedury (log, różnicowanie), które sprowadzają szereg niestacjonarny ( $ARIMA$ ) do stacjonarnego ( $ARMA$ ).

Następnie pokazano, w jaki sposób wykorzystując własności przestrzeni Hilberta w prosty sposób przeprowadzić predykcję szeregów (Rozdział 3), a więc główny cel analizy. Dzięki tej prostocie możliwe jest „dobre” wyznaczenie przyszłych wartości nie tylko szeregów stacjonarnych, ale również i niestacjonarnych. W rozdziale tym wprowadzono dwie metody rekurencyjne (algorytm Durbin’a – Levinson’a, oraz algorytm innowacyjny), które znacznie upraszczają obliczenia, co pozwala na szybkie wyliczanie przyszłych wartości procesu.

W rozdziale 4 omówiono najważniejszą część budowania modeli, czyli estymację parametrów  $(\phi, \theta)$ . W estymacji szeregów czasowych bardzo istotną rolę odgrywa metoda największej wiarygodności, oraz standardowa metoda najmniejszych kwadratów, (która okazuje się w sensie statystycznym, jednym z przypadków MNW dla modelu gaussowskiego). Metody te okazują się uniwersalne nie tylko w estymacji szeregów czasowych, ale również w dowolnej estymacji parametrów funkcji.

Podsumowując w pracy została zamieszczona cała procedura analizy szeregów za pomocą liniowych modeli  $ARIMA$ , co niesie ze sobą korzyści ze względu na prostą budowę modelu, oraz ciekawą liniową predykcję. Sprawia to, że modelowanie to jest szeroko wykorzystywane w różnych dziedzinach np. ekonomii, medycynie i fizyce.

## Uzupełnienie 1

### Definicja 1

*Szeregiem czasowym* nazywamy zbiór wartości cechy w uporządkowanych chronologicznie różnych momentach (przedziałach) czasu. Można go też zdefiniować jako: rodzinę zmiennych losowych  $\{X_t, t \in T\}$  zdefiniowaną na przestrzeni probabilistycznej  $(\Omega, \mathfrak{F}, P)$ .

### Definicja 2

*Funkcja autokowariancji.* Jeśli  $\{X_t\}$  jest procesem dla którego zachodzi:

$Var(X_t) < \infty$ , wtedy funkcja autokowariancji  $\gamma_X(.,.)$  zdefiniowana jest następująco:

$$\gamma_X(i, j) = Cov(X_i, X_j) = E[(X_i - EX_i)(X_j - EX_j)] \quad (U1.1)$$

### Definicja 3

*Stacjonarność szeregu (słaba).* Szereg  $\{X_t\}$  jest stacjonarny (słabo) jeśli dla wszystkich  $t$  zachodzi:

$$1. \quad EX_t = \mu = const \quad (U1.2)$$

$$2. \quad E|X_t|^2 < \infty \quad (U1.3)$$

$$3. \quad \gamma_X(i, j) = \gamma_X(i+h, j+h) \quad (U1.4)$$

### Uwaga

Dokładny opis stacjonarności (słabej i mocnej) zawarty jest w pracy [3].

### Definicja 4

Proces  $\{X_t\}$  jest zespolonym i stacjonarnym (słabo) procesem jeśli zachodzi:

$$1. \quad E|X_t|^2 < \infty \quad (U1.5)$$

$$2. \quad EX_t \text{ nie zależy od } t \quad (U1.6)$$

$$3. \quad E(X_{t+h} \overline{X_t}) \text{ nie zależy od } t \quad (U1.7)$$

### **Uwaga.**

Dla wygody, w przypadku procesu stacjonarnego, będziemy używać funkcję autokowariancji z jedną zmienną zdefiniowaną przez:

$$\gamma_X(h) \equiv \gamma_X(h,0) = \gamma_X(t+h,t) = \text{Cov}(X_{t+h}, X_t) \quad (\text{U1.8})$$

### **Definicja 5**

Funkcja autokowariancji  $\gamma(\cdot)$  dla zespolonego stacjonarnego procesu jest określona następująco:

$$\gamma(h) = E\left(X_{t+h} \overline{X_t}\right) - E(X_{t+h})E(\overline{X_t}) \quad (\text{U1.9})$$

dla zespolonej funkcji autokowariancji zachodzi:

$$1. \quad \gamma(0) \geq 0 \quad (\text{U1.10})$$

$$2. \quad |\gamma(h)| \leq \gamma(0) \text{ dla wszystkich całkowitych } h \quad (\text{U1.11})$$

$$3. \quad \gamma(\cdot) \text{ jest funkcją hermitowską, tzn. } \gamma(h) = \overline{\gamma(-h)} \quad (\text{U1.12})$$

### **Definicja 6**

*Funkcja autokorelacji.* Funkcja Autokorelacji (ACF) zdefiniowana jest jako funkcja, której wartość w opóźnieniu  $h$  wyraża się wzorem:

$$\rho_X(h) \equiv \gamma_X(h) / \gamma_X(0) \quad (\text{U1.13})$$

Funkcje autokowariancji mają też subtelniejszą własność, tzw. nieujemnej określoności, zdefiniowaną poniżej [1, 3].

### **Definicja 7** (Nieujemna Określoność) [1, 3].

Funkcja o wartościach rzeczywistych określona na zbiorze liczb całkowitych,  $\kappa: Z \rightarrow \mathfrak{R}$ , jest określona jako nieujemna wtedy i tylko wtedy, gdy

$$\sum_{i,j=1}^n a_i \kappa(t_i - t_j) a_j \geq 0, \quad (\text{U1.14})$$

dla wszystkich liczb naturalnych  $n$  i dla wszystkich wektorów  $\mathbf{a} = (a_1, \dots, a_n)^T \in \mathfrak{R}^n$ , i

$\mathbf{t} = (t_1, \dots, t_n)^T \in Z^n$ , lub, według równoważnego stwierdzenia, wtedy i tylko wtedy, gdy

$$\sum_{i,j=1}^n a_i \kappa(i-j) a_j \geq 0, \text{ dla każdego naturalnego } n \text{ i } \mathbf{a} = (a_1, \dots, a_n)^T \in \mathfrak{R}^n.$$

**Twierdzenie 1** (Charakterystyka funkcji autokowariancji). [3]

1) Jeśli  $\gamma(\cdot)$  jest funkcją autokowariancji pewnego stacjonarnego procesu  $\{X_t\}$ , to jest ona parzysta i nieujemnie określona, 2) a jeśli układ zmiennych  $(X_{t_1}, \dots, X_{t_n})$  posiada  $n$ -wymiarowy rozkład Gaussowski, to istnieje pewien stacjonarny proces  $\{X_t\}$  ( $t = t_1, \dots, t_n$ ) oraz pewna parzysta i nieujemnie określona funkcja  $\kappa: Z \rightarrow \mathfrak{R}$ , która jest funkcją autokowariancji  $\gamma(\cdot)$  tego procesu.

**Dowód** tego twierdzenia znajduje się w [3]. W dowodzie tego twierdzenia pokazano, że każdy Gaussowski szereg czasowy jest stacjonarny, a jego funkcja autokowariancji jest parzystą i nieujemnie określoną funkcją  $\gamma: Z \rightarrow \mathfrak{R}$ .

Można pokazać, że zachodzi twierdzenie [1]:

**Twierdzenie 2**

Jeśli  $X_1, \dots, X_n$  są  $n$  obserwacjami stacjonarnego procesu a estymatory funkcji autokorelacyjnej w próbie mają postać (2.15):

$$\hat{\gamma}(h) = \begin{cases} \frac{1}{n} \sum_{t=1}^{n-h} (X_t - \bar{X}_n)(X_{t+h} - \bar{X}_n), & \text{dla } |h| < n \\ 0 & \text{dla } |h| \geq n \end{cases} \quad (\text{U1.15})$$

wtedy funkcja  $\hat{\gamma}(\cdot)$  jest nieujemnie określona i zgodnie z Twierdzeniem A, jest funkcją autokowariancji pewnego Gaussowskiego (zatem stacjonarnego) procesu  $\{Y_t\}$ . Wtedy z Twierdzenia 1.11 wynika, że  $\{Y_t\}$  jest procesem MA( $n-1$ ). Z Twierdzenia 3.1 wynika, że jeśli  $\hat{\gamma}(0) > 0$ , wtedy  $\hat{\Gamma}_n$  jest nieosobliwa dla każdego  $n$ .

**Definicja 8**

Przestrzeń  $L^2(\Omega, \mathfrak{F}, P)$ , gdzie trójka  $(\Omega, \mathfrak{F}, P)$  jest przestrzenią probabilistyczną, tworzy zbiór  $C$  wszystkich zmiennych losowych  $X$  zdefiniowanych na przestrzeni zdarzeń  $\Omega$ , które spełniają warunek  $EX^2 = \int_{\Omega} X(\omega)^2 P(d\omega) < \infty$ . (U1.16)

Z mnożeniem przez skalar i dodawaniem zmiennych losowych, zbiór  $C$  tworzy przestrzeń wektorową. Między innymi dla wszystkich  $a \in \mathfrak{R}$  i  $X, Y \in C$  spełnione są własności

$$E(aX)^2 = a^2 EX^2 < \infty, E(X+Y)^2 \leq 2EX^2 + 2EY^2 < \infty \quad (\text{U1.17})$$

i wszystkie pozostałe własności przestrzeni wektorowej. Między innymi  $C$  ma element zerowy, którym jest zmienną losową tożsamościowo równą zero na  $\Omega$ .

### **Definicja 9 (Iloczynu wewnętrznego)**

Zmienne losowe  $X$  o wartościach zespolonych, określone na przestrzeni probabilistycznej  $(\Omega, \mathfrak{F}, P)$  i spełniające warunek  $E|X_t|^2 < \infty$ , tworzą przestrzeń Hilberta z iloczynem wewnętrznym określonym następująco:

$$\langle X, Y \rangle = E(X, \bar{Y}) \quad (\text{U1.18})$$

Dla zmiennych rzeczywistych zachodzi :

$$\langle X, Y \rangle = E(X, Y) \quad (\text{U1.19})$$

### **Uwaga**

Tak określony iloczyn wewnętrzny spełnia wszystkie własności iloczynu wewnętrznego z wyjątkiem tej, że z  $\langle X, X \rangle = 0$  nie wynika, że  $X(\omega) = 0$  dla wszystkich  $\omega \in \Omega$ , a jedynie, że  $P(X=0) = 1$ . Trudność tą omija się mówiąc, że zmienne  $X$  i  $Y$  są równoważne jeśli  $P(X=Y) = 1$ , lub, że są równe z prawdopodobieństwem równym 1. Ta relacja równoważności dzieli przestrzeń  $C$  na klasy równoważności zmiennych losowych, tak, że dowolne dwie zmienne w tej samej klasie są równe z prawdopodobieństwem równym 1.

**Definicja10 (Przestrzeń Hilberta)** Przestrzeń Hilberta  $H$  jest przestrzenią zupełną z iloczynem wewnętrznym, tzn. jest to przestrzeń z iloczynem wewnętrznym, w której każdy ciąg Cauchy'ego  $\{x_n\}$  zbiega się w normie do pewnego elementu  $x \in H$ .

### **Definicja 11 (Podprzestrzeń domknięta)**

Liniowa podprzestrzeń  $M$  przestrzeni Hilberta  $H$ , jest domkniętą podprzestrzenią  $H$ , jeśli  $M$  zawiera wszystkie swoje punkty graniczne, tzn. że warunek:

$$x_n \in M \text{ oraz } \|x_n - x\| \rightarrow 0 \quad (\text{U1.20})$$

implikuje, że  $x \in M$ .

### **Definicja 12 (Ortogonalne dopełnienie)**

Ortogonalne dopełnienie podzbioru  $M$  należącego do przestrzeni Hilberta  $H$ , jest zdefiniowane jako zbiór  $M^\perp$  wszystkich elementów z  $H$ , które są ortogonalne do wszystkich elementów podzbioru  $M$ .

Otrzymujemy więc, że  $x \in M^\perp$  wtedy i tylko wtedy, gdy  $\langle x, y \rangle = 0$  (zapisywane jako  $x \perp y$ ) dla wszystkich  $y \in M$ .

### **Uwaga**

Jeśli  $M$  jest dowolnym podzbiorem przestrzeni Hilberta  $H$ , wtedy  $M^\perp$  jest domkniętą podprzestrzenią Hilberta.

Dowód [1]. Kluczowym elementem dowodu jest skorzystanie z własności ciągłości iloczynu wewnętrznego w przestrzeni  $H$ .

### **Twierdzenie 3**

Jeśli  $M$  jest domkniętą podprzestrzenią przestrzeni Hilberta  $H$ , wtedy:

1. istnieje jedyny element  $\hat{x} \in M$  (nazywany ortogonalnym rzutowaniem  $x$  na  $M$ ) taki, że:

$$\|x - \hat{x}\| = \inf_{y \in M} \|x - y\|, \quad (\text{U1.21})$$

2.  $\hat{x} \in M$  oraz  $\|x - \hat{x}\| = \inf_{y \in M} \|x - y\|$  wtedy i tylko wtedy, gdy  $\hat{x} \in M$  oraz  $(x - \hat{x}) \in M^\perp$

Dowód [1].

### **Wnioski**

Jeśli  $M$  jest zamkniętą podprzestrzenią przestrzeni Hilberta  $H$ , natomiast  $I$  jest odwzorowaniem identyfikacyjnym w  $H$ , wtedy istnieje jedyne odwzorowanie  $P_M$  przestrzeni  $H$  na przestrzeń  $M$  takie, że :

$$I - P_M \text{ rzutuje } H \text{ na przestrzeni } M^\perp.$$

$P_M$  nazywamy odwzorowaniem rzutowym przestrzeni  $H$  na przestrzeń  $M$ .



## Dowód

Z Twierdzenia 1 wynika, że dla każdego  $x \in H$ , istnieje jedno  $\hat{x} \in M$  takie, że :

$$x - \hat{x} \in M^\perp.$$

Właściwym odwzorowaniem jest zatem  $P_M x = \hat{x}$ , dla  $x \in H$ . (U1.22)

## Propozycja 1

Niech  $H$  będzie przestrzenią Hilberta i niech  $P_M$  będzie odwzorowaniem rzutowym na zamkniętą podprzestrzeń  $M$ . Wtedy:

1.  $P_M(\alpha x + \beta y) = \alpha P_M x + \beta P_M y$ ,  $x, y \in H$ ,  $\alpha, \beta \in C$ ,
2.  $\|x\|^2 = \|P_M x\|^2 + \|(I - P_M)x\|^2$ ,
3. dla każdego  $x \in H$ ,  $x = P_M x + (I - P_M)x$ , (U1.23)
4.  $P_M x_n \rightarrow P_M x$  jeśli  $\|x_n - x\| \rightarrow 0$ ,
5.  $x \in M$  wtedy i tylko wtedy, gdy:  $P_M x = x$ ,
6.  $x \in M^\perp$  wtedy i tylko wtedy, gdy:  $P_M x = 0$
7.  $M_1 \subseteq M_2$  wtedy i tylko wtedy, gdy:  $P_{M_1} P_{M_2} x = P_{M_1} x$  dla wszystkich  $x \in H$

Dowód [1].

## Definicja 13 Równanie Przewidywania

Mając przestrzeń Hilberta  $H$ , zamkniętą podprzestrzeń  $M$  i element  $x \in H$ , z Twierdzenia 1 wynika, że element w  $M$  najbliższy elementowi  $\hat{x} \in M$ , jest jedynym elementem  $\hat{x} \in M$ , takim, że:

$$\langle x - \hat{x}, y \rangle = 0 \text{ dla wszystkich } y \in M. \quad (\text{U1.24})$$

$\hat{x} = P_M x$  interpretujemy jako najlepsze przewidywanie  $x$  w podprzestrzeni  $M$ .

## Uwaga.

W szczególnym przypadku gdy  $H = \mathbb{R}^3$  a  $M$  jest podprzestrzenią  $H$  i gdy  $\hat{y} = P_M \bar{y}$ , wtedy równanie (U1.24) jest stwierdzeniem, że  $\hat{y} - \bar{y}$  jest ortogonalne do  $M$ . Twierdzenie o rzutowaniu mówi nam, że  $\hat{y} = P_M \bar{y}$  jest jednoznacznie określone przez warunek (U1.15) dla dowolnej przestrzeni Hilberta  $H$  i domkniętej podprzestrzeni  $M$ .

**Przykład 1** (Minimalny średnio – kwadratowy błąd liniowej predykcji dla procesu stacjonarnego)

Niech  $\{X_t, t = 0, \pm 1, \dots\}$  będzie stacjonarnym procesem na przestrzeni zdarzeń  $(\Omega, \mathfrak{F}, P)$  ze średnią zero i funkcją autokowariancji  $\gamma(\cdot)$ . Rozważmy problem znalezienia liniowej kombinacji

$\hat{X}_{n+1} = \sum_{j=1}^n \phi_{nj} X_{n+1-j}$  będącej najlepszym przybliżeniem dla  $X_{n+1}$  w takim sensie, że

$$E \left| X_{n+1} - \sum_{j=1}^n \phi_{nj} X_{n+1-j} \right|^2 \text{ jest minimalne.}$$

Problem ten łatwo rozwiązać z pomocą twierdzenia o rzutowaniu. Niech  $H = L^2(\Omega, \mathfrak{F}, P)$  oraz

$$M = \left\{ \sum_{j=1}^n \alpha_j X_{n+1-j}; \alpha_1, \dots, \alpha_n \in R \right\}.$$

Minimalizacja  $E \left| X_{n+1} - \hat{X}_{n+1} \right|^2$  jest równoważna minimalizacji kwadratu normy  $\|X_{n+1} - \hat{X}_{n+1}\|^2$ , dzięki czemu z Twierdzenia 1 od razu otrzymujemy, że :

$$\hat{X}_{n+1} = P_M X_{n+1}. \quad (\text{U1.25})$$

Z równania (U1.24) otrzymujemy równanie predykcji :

$$\left\langle X_{n+1} - \sum_{j=1}^n \phi_{nj} X_{n+1-j}, Y \right\rangle = 0, \text{ dla każdego } Y \in M \quad (\text{U1.26a})$$

które, po skorzystaniu z liniowości iloczynu skalarnego, jest równoważne n równaniom postaci:

$$\left\langle X_{n+1} - \sum_{j=1}^n \phi_{nj} X_{n+1-j}, X_k \right\rangle = 0, \quad k = n, n-1, \dots, 1. \quad (\text{U1.26b})$$

Korzystając z Definicji 5 (U1.10) iloczynu wewnętrznego  $\langle X, Y \rangle = E(X, Y)$  w przestrzeni  $L^2(\Omega, \mathfrak{F}, P)$ , widzimy, że równania predykcji można zapisać w formie:

$$\bar{\Gamma}_n \bar{\phi}_n = \bar{\gamma}_n \quad (\text{U1.27})$$

gdzie  $\bar{\phi}_n = (\phi_{1n}, \dots, \phi_{nn})^T$ ,  $\bar{\gamma}_n = (\gamma(1), \dots, \gamma(n))^T$  oraz  $\bar{\Gamma}_n = [\gamma(i-j)]_{i,j=1}^n$ .

Teoria rzutowania gwarantuje, że istnieje co najmniej jedno rozwiązanie  $\bar{\phi}_n$  równania (U1.27). Jeśli  $\bar{\Gamma}_n$  jest osobliwa wtedy (U1.27) ma nieskończenie wiele rozwiązań, jednakże twierdzenie o rzutowaniu gwarantuje to, że każde rozwiązanie daje to samo przewidywanie  $\hat{X}_{n+1}$ .

### Uwaga.

Jeśli  $M = \overline{sp}\{x_1, \dots, x_n\}$ , wtedy dla każdego  $x \in H$ ,  $P_M x$  jest jedynym elementem takim, że:

$$P_M x = \alpha_1 x_1 + \dots + \alpha_n x_n, \quad (\text{U1.28})$$

dla którego zachodzi:

$$\langle x - P_M x, y \rangle = 0, \text{ dla } y \in M \quad (\text{U1.29a})$$

$$\text{lub równoważnie: } \langle P_M x, x_j \rangle = \langle x, x_j \rangle, \text{ dla } j = 1, 2, \dots, n \quad (\text{U1.29b})$$

Równanie (U1.29b) może być zapisane również jako zbiór liniowych równań dla  $\alpha_1, \dots, \alpha_n$ . Jednoznaczność  $P_M x$  oznacza, że wszystkie rozwiązania równania (U1.29b) muszą dawać ten sam element  $\alpha_1 x_1 + \dots + \alpha_n x_n$ .

### Definicja 14 (Zbiór ortonormalny)

Zbiór  $\{e_t, t \in T\}$  nazywamy ortonormalnym, jeśli dla każdego  $s, t \in T$  zachodzi:

$$\langle e_s, e_t \rangle = \begin{cases} 1 & \text{dla } s = t \\ 0 & \text{dla } s \neq t \end{cases} \quad (\text{U1.30})$$

### Przykład 2

Każdy ciąg  $\{Z_t, t \in Z\}$  niezależnych standaryzowanych zmiennych losowych o rozkładzie normalnym jest ortonormalnym zbiorem w  $L^2(\Omega, \mathfrak{F}, P)$ .

### Twierdzenie 4

Jeśli  $\{e_1, \dots, e_k\}$  jest ortonormalnym podzbiorem w przestrzeni Hilberta  $H$ , oraz jeśli  $M = \overline{sp}\{e_1, \dots, e_k\}$ , wtedy:

$$P_M x = \sum_{i=1}^k \langle x, e_i \rangle e_i \text{ dla wszystkich } x \in H \quad (\text{U1.31})$$

### **Twierdzenie 5**

Funkcja  $\gamma(\cdot)$  o wartościach zespolonych, określona na liczbach całkowitych jest funkcją autokowariancji stacjonarnego procesu  $\{X_t, t = 0, \pm 1, \pm 2, \dots\}$ , wtedy i tylko wtedy, gdy:

1. można ją zapisać jako:

$$\gamma(h) = \int_{(-\pi, \pi]} e^{ihv} dF(v) \text{ dla wszystkich } h = 0, \pm 1, \dots \quad (\text{U1.32})$$

gdzie  $F$  jest funkcją prawostronnie ciągłą, niemalejącą i ograniczoną na  $[-\pi, \pi]$ , oraz  $F(-\pi) = 0$ . Funkcja  $F$  nazywana jest widmową funkcją rozkładu dla  $\gamma(\cdot)$

Ponadto funkcja autokowariancji jest nieujemnie określona, wtedy i tylko wtedy (twierdzenie Hergoltz'a), gdy zachodzi powyższy warunek (U1.32).

Jeśli przy tym:

$$F(\lambda) = \int_{-\pi}^{\lambda} f(v) dv, \text{ dla } -\pi \leq \lambda \leq \pi \quad (\text{U1.33})$$

wtedy  $f(\cdot)$  nazywana jest gęstością widmową funkcji autokowariancji  $\gamma(\cdot)$ .

$$2. \quad \sum_{i,j=1}^n a_i \gamma(i-j) \bar{a}_j \geq 0 \quad (\text{U1.34})$$

dla wszystkich  $n$  i wszystkich  $\vec{a} = (a_1, a_2, \dots, a_n)^T \in C^n$ .

Widmowa funkcja rozkładu  $F(\cdot)$  i (jeśli istnieje jedna) odpowiadająca jej gęstość widmowa  $f(\cdot)$  są nazywane funkcją widmowego rozkładu  $F$  oraz funkcją gęstości widmowej  $f$  dla zarówno funkcji autokowariancji  $\gamma(\cdot)$  jak i dla procesu  $\{X_t\}$ .

### **Twierdzenie 6**

Funkcja  $\gamma(\cdot)$  określona na liczbach całkowitych, dla której szereg  $\sum_{n=-\infty}^{\infty} |\gamma(n)|$  jest zbieżny, jest funkcją autokowariancji stacjonarnego procesu wtedy i tylko wtedy, gdy:

$$f(\lambda) := \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} e^{-in\lambda} \gamma(n) \geq 0, \text{ dla wszystkich } \lambda \in [-\pi, \pi] \quad (\text{U1.35})$$

jest gęstością widmową funkcji  $\gamma(\cdot)$ .

Dowód twierdzenia [1].

### **Uwaga.**

Powyższe twierdzenie jest mocnym narzędziem dla sprawdzania nieujemnej określoności funkcji  $f$ .

### **Uwaga.**

Twierdzenie 4 pokazuje, że ponieważ dla każdego procesu  $\{X_t\}$  będącego procesem ARMA( $p, q$ ) zachodzi  $\sum_{n=-\infty}^{\infty} |\gamma(n)| < \infty$  [1], zatem każdy proces ARMA( $p, q$ ) ma gęstość widmową  $f(\cdot)$ , oraz zachodzi:

$$n\sigma^2(\overline{X}_n) \xrightarrow{n \rightarrow \infty} \sum_{h=-\infty}^{\infty} \gamma(h) = 2\pi f(0) \quad (\text{U1.36})$$

gdzie 
$$f(0) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-ih0} \gamma(h) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) \quad (\text{U1.37})$$

Natomiast jeśli proces ARMA ( $p, q$ ) spełnia dodatkowo zależność:

$$X_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j} \quad (\text{U1.38})$$

z warunkiem, że 
$$\sum_{j=-\infty}^{\infty} |\psi_j| < \infty, \quad (\text{U1.39})$$

wtedy  $\sum_{n=-\infty}^{\infty} |\gamma(n)| < \infty$  oraz:

$$n\sigma^2(\overline{X}_n) \rightarrow \sum_{h=-\infty}^{\infty} \gamma(h) = 2\pi f(0) = \sigma^2 \left( \sum_{j=-\infty}^{\infty} \psi_j \right)^2. \quad (\text{U1.40})$$

### **Twierdzenie 7**

Jeśli  $\{X_t\}$  jest dwustronną nieskończoną średnią ruchomą

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j} \quad \{Z_t\} \sim \text{WN}(0, \sigma^2) \quad (\text{U1.41})$$

gdzie 
$$\sum_{j=-\infty}^{\infty} |\psi_j| < \infty, \quad (\text{U1.39}')$$

wtedy : 
$$\sum_{j=-\infty}^{\infty} |\gamma(h)| < \infty \quad (\text{U1.42})$$

gdzie  $\gamma(\cdot)$  jest funkcją autokowariancji procesu  $\{X_t\}$ .

### **Twierdzenie 8**

Jeśli  $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + Z_t$ ,  $t = 1, 2, \dots$ , gdzie  $Z_t \sim IID(0, \sigma^2)$  (U1.43)

i  $Z_t$  nie są skorelowane z  $\{X_j, j < t\}$  dla każdego  $t$ , wtedy można pokazać wykorzystując równanie predykcji [1], że najlepszym przewidywaniem średnio – kwadratowym (tzn. metody najmniejszych kwadratów) dla  $X_{n+1}$  w  $\overline{sp}\{X_j, -\infty < j \leq n\}$  jest :

$$\hat{X}_{n+1} = \phi_1 X_n + \phi_2 X_{n-1} + \dots + \phi_p X_{n+1-p} \quad (\text{U1.44})$$

### **Twierdzenie 9**

Niech  $\vec{Z} = (Z_1, Z_2, \dots, Z_n)$  jest zmienną (wektorem) losową na  $(\Omega, \mathfrak{F}, P)$  i niech  $X \in L^2(\Omega, \mathfrak{F}, P)$ .

Najlepszym przewidywaniem  $\hat{X}$  dla  $X$  w  $M(\vec{X})$ , które jest rozwiązaniem metody najmniejszych kwadratów jest warunkowa wartość oczekiwana  $E(X | \vec{Z} = \vec{z})$ .

### **Dowód**

Niech  $\hat{X} = \hat{X}(\vec{Z}, \vec{\theta})$ , gdzie  $\vec{\theta} = (\theta_1, \dots, \theta_k)$  są parametrami linii regresji wyznaczonymi metodą najmniejszych kwadratów. Zatem z minimalizacji średniej sumy kwadratów odchyleń  $X$  od  $\hat{X}$ , tzn. z minimalizacji  $E(|X - \hat{X}|^2) = \|X - \hat{X}\|^2$  otrzymujemy:

$$\frac{\partial}{\partial \vec{\theta}} E(|X - \hat{X}|^2) = \frac{\partial}{\partial \vec{\theta}} \iint (x - \hat{x})^2 f(x, \vec{z}) dx d\vec{z} = -2 \iint (x - \hat{x}) \frac{\partial \hat{x}}{\partial \vec{\theta}} f(x, \vec{z}) dx d\vec{z} = 0 \quad (\text{U1.45})$$

skąd:

$$\iint x \frac{\partial \hat{x}}{\partial \vec{\theta}} f(x, \vec{z}) dx d\vec{z} = \iint \hat{x} \frac{\partial \hat{x}}{\partial \vec{\theta}} f(x, \vec{z}) dx d\vec{z} \quad (\text{U1.46a})$$

$$\int \frac{\partial \hat{x}}{\partial \vec{\theta}} \left( \int x f(x, \vec{z}) dx \right) d\vec{z} = \int \left[ \hat{x} \frac{\partial \hat{x}}{\partial \vec{\theta}} \int f(x, \vec{z}) dx \right] d\vec{z} \quad (\text{U1.46b})$$

ponieważ:  $\int f(x, \vec{z}) dx = f(\vec{z})$

Zatem z równości funkcji podcałkowych, otrzymujemy:

$$\frac{\partial \hat{x}}{\partial \bar{\theta}} \int x f(x, \bar{z}) dx = \hat{x} \frac{\partial \hat{x}}{\partial \bar{\theta}} f(\bar{z}) \quad (\text{U1.47})$$

skąd:

$$\hat{x} = \int x \frac{f(x, \bar{z})}{f(\bar{z})} dx = \int x f(x | \bar{z}) dx \equiv E(X | \bar{Z} = \bar{z}) \quad (\text{U1.48})$$

tzn.:  $\hat{x} = E(X | \bar{Z} = \bar{z})$ . c.n.d.

## Uzupełnienie 2

**Rekurencyjne wyznaczanie funkcji wiarygodności dowolnego procesu Gaussa ze średnią równą zero.**

Założmy, że  $\{X_t\}$  jest procesem Gauss'a ze średnią równą zero i funkcją autokowariancji

$$\kappa(i, j) = EX_i X_j.$$

Niech  $\vec{X}_n = (X_1, \dots, X_n)^T$  (U2.1)

i  $\hat{X}_n = (\hat{X}_1, \dots, \hat{X}_n)^T$  (U2.2)

gdzie  $\hat{X}_1 = 0$  (U2.3a)

i  $\hat{X}_j = E(X_j | X_1, \dots, X_{j-1}) = P_{sp\{X_1, \dots, X_{j-1}\}} X_j$ ,  $j \geq 2$ , (U2.3b)

gdzie (U2.3b) jest cechą charakterystyczną wielowymiarowego rozkładu normalnego.

Niech  $\Gamma_n$  oznacza macierz kowariancji:

$$\Gamma_n = E(\vec{X}_n \vec{X}_n^T) \quad (\text{U2.4})$$

i założmy, że  $\Gamma_n$  jest nieosobliwa.

Wiarygodność (dla)  $\vec{X}_n$  jest zdefiniowana jako:

$$L(\Gamma_n) = (2\pi)^{-n/2} (\det \Gamma_n)^{-1/2} \exp\left(-\frac{1}{2} \vec{X}_n^T \Gamma_n^{-1} \vec{X}_n\right) \quad (\text{U2.5})$$

Można uniknąć bezpośredniego wyznaczania  $\det \Gamma_n$  oraz  $\Gamma_n^{-1}$ , korzystając z możliwości wyrażenia  $\Gamma_n$  poprzez przewidywanie jednego kroku  $\hat{X}_j$  oraz średniego błędu kwadratowego  $v_{j-1}$ ,  $j = 1, 2, \dots, n$ , które można policzyć rekurencyjnie, korzystając z algorytmu innowacyjnego. (Twierdzenie 3.4)

Niech  $\theta_{ij}$ ,  $j = 1, 2, \dots, i$ ,  $i = 1, 2, \dots$ , oznacza współczynniki wyznaczone po zastosowaniu algorytmu innowacyjnego do funkcji kowariancji  $\kappa$  zmiennych  $\{X_t\}$ , i przyjmijmy  $\theta_{i0} = 1$ ,  $\theta_{ij} = 0$ , dla  $j < 0$ ,  $i = 1, 2, \dots$ . Warunek ten wynika z postaci rozwinięcia (3.49) dla  $\hat{X}_{n+1}$ :

$$\hat{X}_{n+1} = \begin{cases} 0 & \text{dla } n = 0 \\ \sum_{j=1}^n \theta_{nj} (X_{n+1-j} - X_{n+1-j}) & \text{dla } n \geq 1 \end{cases} \quad (\text{U2.6})$$

Wprowadźmy macierz:

$$C = [\theta_{i,i-j}]_{i,j=0}^{n-1}, \quad (\text{U2.7})$$

która ma niezerowe elementy na i poniżej diagonalnej, oraz  $n \times n$  wymiarową, diagonalną macierz:

$$D = \text{diag}(v_0, v_1, \dots, v_{n-1}); \text{ gdzie } v_j = E|X_{j+1} - \hat{X}_{j+1}|^2 \quad (\text{U2.8})$$

Reprezentacja innowacyjna (3.50) dla  $\hat{X}_j$ ,  $j = 1, 2, \dots, n$  może być teraz zapisana następująco:

$$\hat{X}_n = (C - I)(\bar{X}_n - \hat{X}_n) \quad (\text{U2.9})$$

gdzie  $I$  jest  $n \times n$  wymiarową macierzą jednostkową. Macierz  $(C - I)$  ma elementy niezerowe poniżej przekątnej.

Z (U2.9) otrzymujemy:

$$\hat{X}_n = C(\bar{X}_n - \hat{X}_n) - (\bar{X}_n - \hat{X}_n)$$

$$\text{skąd: } \bar{X}_n = C(\bar{X}_n - \hat{X}_n) \quad (\text{U2.10})$$

$$\text{Ponieważ } (X_j - \hat{X}_j) \perp (X_i - \hat{X}_i), \quad \text{gdzie } i \neq j \quad (\text{U2.11})$$

$$\text{skąd } E\{(X_j - \hat{X}_j)(X_i - \hat{X}_i)\} = 0 \quad \text{dla } i \neq j \quad (\text{U2.12})$$

zatem macierz  $D$  jest macierzą kowariancji dla  $(\bar{X}_n - \hat{X}_n)$ .

Mamy więc:

$$\Gamma_n = E(\bar{X}_n \bar{X}_n^T) = E\{C(\bar{X}_n - \hat{X}_n)(\bar{X}_n - \hat{X}_n)^T C^T\} = CE\{(\bar{X}_n - \hat{X}_n)(\bar{X}_n - \hat{X}_n)^T\}C^T = CDC^T \quad (\text{U2.13})$$



Zatem:

$$\Gamma_n = CDC^T \quad (\text{U2.14})$$

gdzie  $D = E\{(\bar{X}_n - \hat{X}_n)(\bar{X}_n - \hat{X}_n)^T\} = \text{diag}(v_0, v_1, \dots, v_{n-1})$  (U2.15)

i  $D^{-1} = \text{diag}(v_0^{-1}, v_1^{-1}, \dots, v_{n-1}^{-1})$ . (U2.16)

Korzystając z (U2.10) otrzymujemy:

$$\bar{X}_n^T \Gamma_n^{-1} \bar{X}_n = (\bar{X}_n - \hat{X}_n)^T C^T \Gamma_n^{-1} C (\bar{X}_n - \hat{X}_n) \quad (\text{U2.17})$$

i ponieważ  $C$  jako macierz trójkątna z jedynekami na przekątnej spełnia warunek:

$$C^T = C^{-1} \quad (\text{U2.18})$$

a  $\Gamma_n^{-1} = (CDC^T)^{-1} = (C^T)^{-1} D^{-1} C^{-1} = CD^{-1} C^T$  (U2.19)

skąd  $D^{-1} = C^T \Gamma_n^{-1} C$  (U2.20)

więc  $\bar{X}_n^T \Gamma_n^{-1} \bar{X}_n = (\bar{X}_n - \hat{X}_n)^T D^{-1} (\bar{X}_n - \hat{X}_n) = \sum_{j=1}^n (X_j - \hat{X}_j)^2 \cdot v_{j-1}^{-1}$ . (U2.21)

Wyznacznik  $\Gamma_n$  jest równy:

$$\det \Gamma_n = \det(CDC^T) = (\det C)^2 \det D = 1 \cdot v_0 \cdot v_1 \cdot \dots \cdot v_{n-1}$$

tzn.  $\det \Gamma_n = v_0 \cdot v_1 \cdot \dots \cdot v_{n-1}$  (U2.22)

Ostatecznie otrzymujemy, że wiarygodność (U2.5) wektora  $\bar{X}_n$  jest równa:

$$L(\Gamma_n) = (2\pi)^{-n/2} (v_0 \cdot v_1 \cdot \dots \cdot v_{n-1})^{-1/2} \exp\left(-\sum_{j=1}^n (X_j - \hat{X}_j)^2 / v_{j-1}\right) \quad (\text{U2.23})$$

Stosując teraz algorytm innowacyjny (3.50) do funkcji kowariancji  $\kappa$ , otrzymujemy zgodnie z Twierdzeniem 3.4 rekurencyjnie  $\hat{X}_1, \hat{X}_2, \dots$ ;  $v_0, v_1, \dots$ , a zatem otrzymujemy funkcję wiarygodności  $L(\Gamma_n)$ .

Jeśli  $\Gamma_n$  można wyrazić poprzez skończoną liczbę nieznanymi parametrów  $\beta_1, \dots, \beta_r$ , tak jak to ma miejsce gdy  $\{X_t\}$  jest procesem ARMA( $p, q$ ) z  $r = p + q + 1$  parametrami (1 parametr dotyczy wariancji  $\sigma^2$  białego szumu), wtedy zazwyczaj musimy estymować parametry z otrzymanych danych dla  $\bar{X}_n$ .

Standardowa procedura estymacyjna [12], polega na maksymalizacji wiarygodności  $L(\beta_1, \dots, \beta_r)$  w konkretnej próbce (realizacji  $\bar{x}_n$ ) ze względu na  $\beta_1, \dots, \beta_r$ .

W przypadku gdy  $X_1, X_2, \dots$  są niezależne i mają takie same rozkłady, wiadomo, że przy dość ogólnych warunkach, estymatory maksymalnej wiarygodności są zgodne (dla  $n \rightarrow \infty$ ), oraz mają asymptotycznie rozkład normalny z wariancjami nie większymi niż inne asymptotycznie normalne estymatory, czyli posiadają własność asymptotycznej efektywności w klasie rozkładów normalnych.

Zatem zwykłą drogą postępowania dla procesów Gauss'a jest maksymalizacja funkcji wiarygodności  $L(\Gamma_n)$  zadanej w postaci (U2.23), względem parametrów  $\beta_1, \dots, \beta_r$ .

### **Uwaga.**

W rozważanych przez nas procesach zmienne  $\{X_t\}$  nie są jednak niezależne. Ze względu na ten fakt, zależność ciągu zmiennych  $\{X_t\}$  musi być wzięta pod uwagę przy badaniu asymptotycznych własności estymatorów (por. 4.75).

### **Uwaga.**

Należy pamiętać, że funkcja wiarygodności jest traktowana jako funkcja parametrów  $\beta_1, \dots, \beta_r$ , a nie funkcja próby  $X_1, X_2, \dots, X_n$ , chociaż jej interpretacja jako rozkładu (gęstości) prawdopodobieństwa pojawienia się próbki  $x_1, \dots, x_n$ , nie stoi tu, dla ustalonych  $x_1, \dots, x_n$ , w sprzeczności.

## **Uzupełnienie 3. Uwagi dotyczące klasycznych testów dla reszt modelu**

W literaturze poświęcono zagadnieniu testowania normalności rozkładu zmiennej losowej, w tym i reszt, oraz testowaniu niezależności reszt modelu wiele uwagi [2, 10, 11]. Podstawowe standardowe testy zgodności, to np. test chi-kwadrat Pearsona i test Kołmogorowa-Smirnowa. W testach zgodności porównuje się funkcję rozkładu prawdopodobieństwa badanej cechy w próbie z jej hipotetycznym rozkładem w populacji. Ze względu na obszerną literaturę nie będę testów tych omawiał.

Poniżej podam jedynie omówienie dwóch testów dotyczących testowania hipotez o braku autokorelacji reszt modelu [za 2] oraz test braku stacjonarności [13].

## Test Durbina-Watsona

W tym teście weryfikuje się hipotezę o zerowaniu się współczynnika autokorelacji.

Niech  $\{U_i, i = 1, 2, \dots, n\}$  będą zwykłymi resztami modelu obserwowanymi w próbie. Rozpatrujemy parę hipotez, gdzie hipoteza zerowa oznacza nie występowanie autokorelacji (pierwszego rzędu) składnika losowego badanego modelu w populacji.

$$H_0: \rho = 0 \quad (\text{U3.1})$$

$$H_1: \rho < 0 \text{ lub } \rho > 0 \quad (\text{U3.2})$$

gdzie  $\rho$  jest wartością współczynnika autokorelacji rzędu pierwszego w populacji. Jego estymatorem w próbie jest współczynnik autokorelacji w próbie  $\hat{\rho}$ :

$$\hat{\rho} = \frac{\sum_{i=2}^n (U_i - \bar{U}) \cdot (U_{i-1} - \bar{U})}{\sqrt{\sum_{i=2}^n (U_i - \bar{U})^2} \cdot \sqrt{\sum_{i=2}^n (U_{i-1} - \bar{U})^2}} \quad (\text{U3.3})$$

Gdy  $\bar{U} = 0$  wtedy estymatorem  $\rho$  jest

$$\hat{\rho} = \frac{\sum_{i=2}^n U_i U_{i-1}}{\sqrt{\sum_{i=2}^n U_i^2} \cdot \sqrt{\sum_{i=2}^n U_{i-1}^2}} \quad (\text{U3.4})$$

Statystyka testowa jest dana wzorem:

$$d = \frac{\sum_{i=2}^n (U_i - U_{i-1})^2}{\sum_{i=1}^n U_i^2} \quad (\text{U3.5})$$

Ma ona rozkład Durbina-Watsona. Jej związek z estymatorem  $\hat{\rho}$  jest następujący:

$$d = 2 \left( 1 - \hat{\rho} \right) \quad (\text{U3.6})$$

Dla ustalonego poziomu istotności  $\alpha$ , dla liczby czynników  $k$  (czyli liczby szacowanych parametrów  $k+1$ ) i liczebności próby  $n$ , odczytujemy z tablic rozkładu Durbina-Watsona dwie wartości krytyczne, (dolną)  $d_l$  i (górną)  $d_u$ .

Natomiast metoda weryfikacji zależy w zależności, w którym miejscu przedziału (0,4) znajduje się obliczana wartość statystyki  $d$ .

a) W przypadku gdy w próbie  $\hat{\rho} > 0$  ( $d \in (0,2)$ ), wtedy hipoteza alternatywna formułowana jest w postaci:

$$H_1 : \rho > 0 \quad (\text{U3.7})$$

czyli hipoteza o dodatniej autokorelacji reszt.

Wartości statystyki  $d$  porównujemy z wartościami krytycznymi  $d_l$  i  $d_u$ .

Jeśli  $d < d_l$  to  $H_0$  odrzucamy na korzyść  $H_1$  i wnioskujemy, że autokorelacja jest dodatnia. Jeśli  $d > d_u$  to nie ma podstaw do odrzucenia  $H_0$  i wnioskujemy, że nie ma autokorelacji (dodatniej) reszt.

b) W przypadku gdy w próbie  $\hat{\rho} < 0$  ( $d \in (2,4)$ ), wtedy hipoteza alternatywna formułowana jest w postaci:

$$H_1 : \rho < 0, \quad (\text{U3.8})$$

czyli hipoteza o ujemnej autokorelacji reszt.

W przypadku tym obliczamy wartość statystyki

$$d' = 4 - d. \quad (\text{U3.9})$$

Wartości statystyki  $d'$  również porównujemy z wartościami krytycznymi  $d_l$  i  $d_u$ .

Jeśli  $d' < d_l$  to  $H_0$  odrzucamy na korzyść  $H_1$  i wnioskujemy, że autokorelacja jest ujemna. Jeśli  $d' > d_u$  to nie ma podstaw do odrzucenia  $H_0$  i wnioskujemy, że nie ma autokorelacji (ujemnej) reszt.

**Uwaga:** Jeśli dla przypadku (a)  $d_u \geq d \geq d_l$  lub dla przypadku (b)  $d_u \geq d' \geq d_l$ , wtedy powyższy test nie daje podjąć decyzji statystycznej dotyczącej występowania autokorelacji reszt.

Jeśli mamy dwie cechy mierzalne  $X$  i  $Y$ , które mają dwuwymiarowy rozkład normalny, wtedy można sobie poradzić stosując statystykę  $t$ :

$$t = \frac{\hat{\rho}}{\sqrt{1 - \hat{\rho}^2}} \cdot \sqrt{n - 2} \quad (\text{U3.10})$$

Przy prawdziwości hipotezy zerowej (U3.1), statystyka  $t$  ma rozkład t-Studenta z  $n-2$  stopniami swobody. Jeśli  $t_\alpha$  jest kwantylem rzędu  $(1 - \alpha / 2)$  rozkładu t-Studenta, to, gdy  $|t| \geq t_\alpha$  wtedy na poziomie istotności  $\alpha$  odrzucamy  $H_0$  na korzyść  $H_1$  i wnioskujemy, że istnieje autokorelacja. Jeśli  $|t| < t_\alpha$  to nie ma podstaw do odrzucenia  $H_0$ .

Statystykę  $t$  tą można by zresztą, przy spełnieniu odpowiednich dla niej założeń, stosować i w pozostałych przypadkach.

### **Test Ljung'a – Box'a**

Hipoteza zerowa: Reszty  $\{U_t\}$  są białym szumem.

W SAS'ie wykorzystywana jest statystyka  $\chi^2$  określona według formuły Ljung'a – Box'a:

$$\chi_m^2 = n(n+2) \sum_{k=1}^m \frac{R_k^2}{(n-k)} \quad (\text{U3.11})$$

gdzie

$$R_k = \frac{\sum_{t=1}^{n-k} U_t U_{t+k}}{\sum_{t=1}^n U_t^2} \quad (\text{U3.12})$$

natomiast  $U_t$  oznacza resztę.

Hipoteza zerowa tej statystyki zakłada, że reszty są białym szumem.

Ostatnim podanym testem jest test stacjonarności badanego procesu.

### **Test Dickey'a – Fuller'a**

**Definicja:** Mówimy, że zmienna  $X$  jest zintegrowana stopnia  $d$ , co oznaczamy:  $X \sim I(d)$ , jeśli sama zmienna jest niestacjonarna, lecz jej przyrosty są stacjonarne, przy czym  $d$  jest najmniejszą liczbą całkowitą, dla której  $\nabla^d X$  jest stacjonarne.

Niech  $Z_t$  będą niezależnymi zmiennymi stacjonarnymi o jednakowym rozkładzie i wartości oczekiwanej równej zero.

Wtedy: 
$$X_t = X_{t-1} + Z_t \tag{U3.13}$$

byłaby zmienną zintegrowaną stopnia 1.

Jeśli jednak chcemy sprawdzić, czy pewna zmienna jest generowana przez proces postaci (U3.13) nie możemy zastosować sposobu polegającego na oszacowaniu regresji (U3.13) metodą najmniejszych kwadratów i zastosowaniu zwykłego testu *t*-Studenta. Zmienne występujące w równaniu przy założeniu prawdziwości poniższej hipotezy zerowej są bowiem niestacjonarne.

Hipoteza zerowa i alternatywna:

$H_0$ : zmienna  $X$  jest niestacjonarna wskutek występowania pierwiastka jednostkowego,

$H_1$ : zmienna  $X$  jest stacjonarna.

Odpowiadają one przypadkom, gdy w równaniu  $X_t = \phi_1 X_{t-1} + Z_t$  (U3.14)

parametr  $\phi_1$  jest, odpowiednio, równy 1, lub mniejszy co do modułu od 1 (Rozdział 1).

Metodą najmniejszych kwadratów należy oszacować równanie

$$\nabla X_t = \phi_1 X_{t-1} + Z_t \tag{U3.15}$$

Nowa hipoteza zerowa odpowiada przypadkowi, gdy parametr  $\phi_1$  jest równy zero,

nowa hipoteza alternatywna – gdy  $\phi_1$  jest mniejszy od zera.

Przy założeniu prawdziwości hipotezy zerowej równanie (U3.15), w przeciwieństwie do (U3.14), ma po lewej stronie zmienną stacjonarną.

Statystyka testu Dickey’a-Fuller’a jest obliczana jako  $DF = \frac{\hat{\phi}_1}{S_{\hat{\phi}_1}}$ , czyli analogicznie jak statystyka

*t*-Studenta. Ma jednak inny rozkład: *asymetryczny i o ujemnej wartości oczekiwanej*. Nie wolno więc stosować tablic rozkładu *t*-Studenta!

Sposób przeprowadzenia testu:

Obliczoną wartość statystyki Dickey’a – Fuller’a porównujemy z odczytaną z tablic wartością krytyczną dla odpowiedniej liczby obserwacji i dla przyjętego poziomu istotności. *Jeśli obliczona wartość jest mniejsza niż wartość krytyczna, odrzucamy hipotezę zerową na rzecz alternatywnej, oznaczającej stacjonarność badanej zmiennej. Jeśli jednak obliczona wartość statystyki DF jest większa niż wartość krytyczna, nie ma podstaw do odrzucenia hipotezy zerowej.*

## Bibliografia:

1. P.J. Brockwell, R.A. Davis. "*Time series: Theory and Methods*", Springer – Verlag 1991r.
2. P. Kruczek, „*Diagnostyka reszt w modelach regresji liniowej.*” – praca licencjacka.
3. Z. Wierzbicka „*Wykorzystanie programu statystycznego SAS jako narzędzia do analizy szeregów czasowych*” – praca licencjacka.
4. Box G.E.P. and Jenkins G.M. „*Time series analysis : forecasting and control*” Holden – Dey, 1970 San Francisco
5. <http://inflacja.icm.edu.pl/jmyc/Notatki/SW10T.pdf>
6. Ruey S. Tsay "*Analysis of Financial Time Series*" Wiley series in probability and statistics 2002r.
7. Akaike H.,(1969) "*Fitting autoregressive models for prediction*", Annals of the institute of statistical mathematics Tokyo, 21, 243-247
8. Hurvich, C.M. and Tsai, C.L. (1989), "*Regression and time series model selection in small samples*" Biometrika 76, 297-307
9. Brockwell P.J., Davis R.A. (1988b), "*Simple consistent estimation of the coefficient of a linear filter*" Stoch. Processes
10. M. Sobczyk „*Statystyka*” PWN 1998
11. Józwiak J., Podgórski J. „*Statystyka od podstaw*” PWE, Warszawa 1997.
12. E. L. Lehmann, "Theory of point estimation", John Wiley, New York, (1983).
13. Pomoc pakietu SAS’a wersja 9.1.