

Analiza szeregów czasowych

Jerzy Mycielski

21 maja 2020

- Na zajęciach będziemy zajmować się modelami, które ekonometrycznymi, które szacowane są na bazie szeregów czasowych
- Dwa główne wątki poruszane w ramach zajęć dotyczyć będą:
 - szacowania modeli prognostycznych
 - szacowania modeli z elementami struktury
- Różnica między tymi dwoma typami modeli związana jest ze sposobem ich zastosowania:
 - modele prognostyczne znajdują zastosowanie głównie w prognozowaniu przyszłych wartości zmiennych ekonomicznych
 - modele z elementami struktury służą do analizowania skutków zmian w polityce gospodarczej (policy analysis)

- Proces stochastyczny: w interesujących nas przypadkach ciąg zmiennych losowych indeksowany kolejnymi liczbami naturalnymi

$$\{X_t; t = 1, \dots, T\}$$

- Szereg czasowy jest zbiorem danych, który stanowi realizację procesu stochastycznego

$$\{x_t; t = 1, \dots, T\}$$

- Zmienną opóźnioną nazywamy zmienną która równa jest wartości danej zmiennej w poprzednim okresie
 - n.p. PKB_t może oznacza wartość PKB z danego okresu a PKB_{t-1} oznacza wartość tej samej zmiennej w poprzednim okresie
- W analizie szeregów czasowych zmienne opóźnione odgrywają bardzo dużą rolę, ponieważ:
 - służą do opisu dynamicznych własności zanalizowanych zmiennyczwiązanych z inercją/histerezą procesów ekonomicznych
 - umożliwiają formułowanie prognoz dotyczących przyszłych wartości zmiennych

- Operacje opóźniania zmiennych można zapisać za pomocą mnożenia przez specjalny operator (operator opóźnień) oznaczanego standardową literą L
 - przykładowo

$$x_{t-1} = Lx_t$$

- Operator ten ma kilka wygodnych cech, które umożliwiają często uproszczenie zapisu:

$$x_{t-2} = Lx_{t-1} = LLx_t = L^2x_t$$

- Ogólnie

$$x_{t-s} = L^s x_t$$

- Przez różnicowanie zmiennych rozumiemy operację odjęcia od obecnej wartości zmiennej jej wartości z poprzednich okresów
- Różnicowanie niesezonowe

$$\Delta x_t = x_t - x_{t-1}$$

- Operację różnicowania można zastosować wielokrotnie np:

$$\Delta(\Delta x_t) = \Delta^2 x_t$$

- Różnicowanie sezonowe

$$\Delta_s x_t = x_t - x_{t-s}$$

- Zauważmy, że

$$\Delta^2 x_t = \Delta x_t - \Delta x_{t-1} = x_t - 2x_{t-1} + x_{t-2}$$

$$= (1 - 2L + L^2) x_t = (1 - L)^2 x_t$$

$$\Delta^s x_t = (1 - L)^s x_t$$

ale

$$\Delta_s x_t = x_t - L^s x_t = (1 - L^s) x_t$$

- Miarami precyzji prognoz są między innymi:
 - pierwiastek z błędu średniokwadratowy (RMSE)

$$RMSE = \sqrt{E(\hat{y}_t - y_t)^2}$$

- średni błąd absolutny (MAE)

$$MAE = E(|\hat{y}_t - y_t|)$$

- średni absolutny błąd procentowy (MAPE)

$$MAPE = E\left(\left|\frac{\hat{y}_t - y_t}{y_t}\right|\right)$$

- Badamy jakość progностyczną modeli porównując miary jakości prognoz
- Miary jakości prognoz liczymy na podstawie próby, która *nie została* użyta w procesie szacowania modeli
- Standardowy sposób postępowania wygląda następująco: dzielimy próbę na okres *in sample* i *out of sample*:

$$\underbrace{1 \dots \dots \dots T_1}_{in\ sample} \underbrace{T_1 + 1 \dots \dots \dots T}_{out\ of\ sample}$$

- Model szacujemy na próbce *in sample*, miary jakości prognoz liczymy na podstawie próbki *out of sample*
 - Powodem, dla którego badanie jakości prognoz odbywa się na podstawie próby *out of sample* jest to, że dobrą jakość dopasowania można zawsze uzyskać używając dużej liczby parametrów.
 - Przeparametryzowany model (*overfitted*) ma zazwyczaj słabe własności *out of sample*

- O występowaniu sezonowości mówimy obserwujemy w danych okresowe zmiany pokrywające się z cyklem rocznym
- Częstotliwość zmian sezonowych powinna być równa częstotliwości próbkowania danych
 - dane miesięczne: może występować sezonowość miesięczna (cykl powtarza się co 12 miesięcy)
 - dane kwartalne: może występować sezonowość kwartalna (cykl powtarz się co 4 kwartały)

- 1 kalendarzowe: związane z nierówną liczbą dni pracujących w poszczególnych miesiącach (liczba sobót i niedziel, świąt)
- 2 administracyjno-kalendarzowe: związane z ustaleniami dotyczącymi kalendarium działania instytucji i firm (kalendarz działania instytucji edukacyjnych, roku podatkowy itp.)
- 3 pogodowe: związane z rocznym cyklem opadów, średnich temperatur etc. Tego typu cykl wpływa na produkcję rolniczą, budowlaną itd. Wpływa on na popyt na dobra konsumpcyjne (n.p. odzież)
- 4 związane z oczekiwaniami: pewne zmiany obserwowane w zmiennych mogą wynikać z oczekiwań dotyczących występowania cykli sezonowych. Przykładowo wzrost popytu na ciepłą odzież występuje jeszcze przed rozpoczęciem zimy i nawet w przypadku, gdy zima jest stosunkowo ciepła.

Sezonowość addytywna i multiplikatywna

- Załóżmy, że możliwa jest dekompozycja szeregu czasowego y_t na dwa elementy
 - element niesezonowy m_t (trend)
 - element sezonowy s_t
 - element czysto losowy r_t
- Możemy wyróżnić dwa przypadki
 - sezonowość addytywną

$$y_t = m_t + s_t + r_t$$

- sezonowość multiplikatywną

$$y_t = m_t s_t r_t$$

- Szereg czasowy z sezonowością multiplikatywną można przekształcić do szeregu z sezonowością addytywną

$$\ln(y_t) = \ln(m_t) + \ln(s_t) + \ln(r_t)$$

Metody wyrównywania sezonowego oparte na różnicowaniu sezonowym

- W przypadku sezonowości addytywnej prostym sposobem radzenia sobie z sezonowością jest różnicowanie sezonowe:

$$\Delta_s y_t = \Delta_s m_t + \Delta_s r_t$$

jeśli tylko $\Delta_s s_t = 0$ (elementy sezonowe nie zmieniają się z okresu na okres)

- Podobnie w przypadku sezonowości multiplikatywnej

$$\frac{y_t}{y_{t-s}} = \frac{m_t r_t}{m_{t-s} r_t}$$

lub

$$\ln\left(\frac{y_t}{y_{t-s}}\right) = \ln(y_t) - \ln(y_{t-s}) = \Delta_s \ln(y_t) = \Delta_s \ln(m_t) + \Delta_s \ln(r_t)$$

Do czego służy modelowanie sezonowości?

- Sezonowość danych może być bardzo interesująca z biznesowego punktu widzenia. Dobrze zidentyfikowanie rodzaju sezonowości jest ważne na przykład dla prawidłowego:
 - planowaniu wielkości produkcji
 - planowania zapasów
 - ustaleniu planów urlopowych
 - planowania wielkości mocy dostarczanych przez operatorów sieci energetycznej
 - charterowaniu lotów i hoteli przez agencje turystyczne

Dlaczego używamy danych wyrównanych sezonowo?

- W analizach ekonomicznych sezonowość traktowana jako coś oczywistego i mało interesującego
- Zmiany sezonowe są traktowane jako zakłócenia utrudniające zidentyfikowanie trendów i cykli średnio i długookresowych
- Z reguły polityka gospodarcza odnosi się właśnie do trendów i cykli średnio i długookresowych
- W szczególności działania władz fiskalnych i monetarnych, których celem jest ograniczenie wielkości wahań w ramach cyklu koniunkturalnego odnoszą się do takich średnio-okresowych zmian
- Także celem polityki rozwojowej jest wpływ na trendy średnio i długookresowe a nie na wahania sezonowe

- Dodatkowym problemem pojawiającym się w przypadku analizy danych niewyrównanych jest potencjalny problem wystąpienia pozornych zależności między zmiennymi wynikających
- Rozważmy model, w którym zmienną objaśnianą jest $y_t = m_{y,t} + s_{y,t} + r_{y,t}$ a zmienną objaśniającą $x_t = m_{x,t} + s_{x,t} + r_{x,t}$
- Analizując korelacje między y_t i x_t lub przeprowadzając regresję y_t na x_t możemy uzyskać istotne zależności nawet wtedy, gdy $m_{y,t}$ i $m_{x,t}$ od siebie nie zależą jeśli tylko $s_{y,t}$ i $s_{x,t}$ są ze sobą skorelowane
- Zależności te wynikać będą ze wspólnego cyklu sezonowego, a nie z głębszych zależności ekonomicznych

- Trendem nazywamy część składnika niesezonowego, który ma długookresowy wpływ wielkość zmiennej
- Wyróżniamy dwa typy trendów:
 - deterministyczny
 - stochastyczny
- Trend deterministyczny, to trend który można modelować wstawiając do modelu deterministyczne funkcje czasu
 - przykład: model trendu liniowego:

$$y_t = \beta_0 + \beta_1 t + \varepsilon_t$$

trend ma postać $m_t = \beta_1 t$

- Trend stochastyczny, trend o charakterze losowym:
 - przykład:

$$y_t = \beta_0 + \beta_1 \sum_{s=1}^t u_s + \varepsilon_t$$

w tym przypadku $m_t = \sum_{s=1}^t u_s$ reprezentuje trend stochastyczny

Średnia ruchoma (Moving Average)

- Średnia ruchoma jest prostym sposobem wygładzania szeregu czasowego
- Celem takiego wygładzania jest usunięcie z szeregu wpływu czynników krótkookresowych (sezonowych, losowych) i uzyskanie oszacowania tendencji długookresowej (trendu)
- Średnia ruchoma szacowana jest za pomocą następującego wzoru

$$\hat{m}_t = \frac{1}{1+2a} \sum_{k=-a}^a y_{t+k}$$

- Możliwe jest także stosowanie wzorów o nierównych wagach

- Znamy okres sezonowości s (na przykład przy danych kwartalnych równy 4)
- Załóżmy, że proces stochastyczny ma postać addytywną

$$y_t = m_t + s_t + r_t$$

- Oszacowanie trendu uzyskujemy przez zastosowanie średniej ruchomej:
 - 1 dla nieparzystego s (np. obserwacje codzienne $s = 365$ dni, cykl roczny) dla $a = \frac{s-1}{2}$ i równych wagach $w(k) = \frac{1}{1+2a}$.
 - 2 dla parzystego s (np. obserwacje kwartalne, miesięczne) dla $a = \frac{s}{2}$ i wagach $w(k) = \frac{1}{s}$ dla $k = 1, \dots, \frac{s}{2} - 1$, $w(\frac{s}{2}) = \frac{1}{2s}$.

- dla szeregu kwartalnego, $s = 4$

$$m_t = \frac{1}{4} \left(\frac{1}{2}x_{t-2} + x_{t-1} + x_t + x_{t+1} + \frac{1}{2}x_{t+2} \right)$$

- Składnik sezonowy uzyskujemy uśredniając obserwacje dla poszczególnych sezonów i odejmując średnią ogólną

$$\hat{s}_t = \sum_{k=1}^T w_s(t, k) y_t - \bar{y},$$

gdzie $w_s(t, k) = \frac{s}{T}$ dla przypadków, gdy t i k dotyczy tych samych sezonów ($(t \bmod s) = (k \bmod s)$), $w(t, k) = 0$ dla pozostałych przypadków, $\bar{y} = \frac{1}{T} \sum_{k=1}^T$

- Oszacowanie czynnika sezonowego dla drugiego kwartału:

$$\hat{s}_2 = \frac{4}{T} (y_2 + y_6 + y_{10} + \dots + y_{T-2}) - \bar{y}$$

- Czynniki losowy (resztę) znajdujemy odejmując od szeregu trend i czynniki sezonowy

$$\hat{r}_t = y_t - \hat{m}_t - \hat{s}_t$$

- Zauważmy, że przy tym sposobie liczenia, wielkość czynników sezonowych nie zmienia się w czasie.
- Powyższa metoda dekompozycji może zostać także zastosowana w przypadku sezonowości multiplikatywnej po zastosowaniu przekształcenia logarytmicznego
- Używa się także bardziej wyrafinowanych sposobów ustalania wag $w_m(t, k)$ i $w_s(t, k)$ opartych np. na procedurze LOESS (Locally Weighted Scatterplot Smoothing)

- Algorytm wyrównywania wykładniczego

$$\hat{y}_0 = y_0$$

$$\hat{y}_t = \alpha y_t + (1 - \alpha) \hat{y}_{t-1}$$

gdzie $\alpha \in (0, 1)$ jest parametrem wygładzania.

- Stosując rekursywne podstawianie uzyskujemy:

$$\hat{y}_t = \alpha \sum_{i=0}^{t-1} (1 - \alpha)^i y_{t-i} + (1 - \alpha)^t y_0$$

- Wygładzanie wykładnicze jest w istocie średnią ruchomą z wagami malejącymi wykładniczo wraz z czasem i stąd alternatywna nazwa EWMA (Exponentially weighted moving average)
- Szereg jest tym bardziej wygładzony im bliższe zera jest α
- Wygładzanie wykładnicze sprawdza się jako metoda usuwania wpływu składników losowych jednak słabo radzi sobie z uwzględnianiem trendu
- Zarówno w modelu wygładzania wykładniczego jak i podwójnego wygładzania wykładniczego można uwzględnić składnik sezonowy
- Składnik trendu, sezonowy i losowy mogą mieć charakter addywny lub multiplikatywny w różnych kombinacjach

- Podwójne wygładzanie wykładnicze stosowane jest w przypadku, gdy zmienna wykazuje wyraźny trend

$$\hat{y}_0 = y_0$$

$$\hat{m}_0 = m_0$$

$$\hat{y}_t = \alpha y_t + (1 - \alpha) (\hat{y}_{t-1} + \hat{b}_{t-1})$$

$$\hat{b}_t = \beta (\hat{y}_t - \hat{y}_{t-1}) + (1 - \beta) \hat{b}_{t-1}$$

gdzie składnik b_t można zinterpretować jako nachylenie trendu

- Zauważmy, że w tym przypadku zarówno b_t jak i składnik samo y_t uzyskiwane jest za pomocą wygładzania wykładniczego

Wygładzanie wykładnicze, ustalanie parametru α oraz wartości y_0

- Wygładzanie wykładnicze jest zasadniczo metodą ad-hoc (wymyśloną bez modelu statystycznego), choć można nadać jej interpretację statystyczną
- W związku z tym parametr
- Parametr α i wartość początkowa y_0 są często ustalone arbitralnie
- Alternatywą jest ustalenie ich tak, by przewidywania y_t^2 były najbliższe wartościom zaobserwowanym przy wykorzystaniu np. kryterium najmniejszych kwadratów

$$\min_{y_0, \alpha, \beta} \sum_{s=1}^t (\hat{y}_t - y_t)^2$$

- Na bazie wygładzania wykładniczego można w łatwy sposób generować prognozy:
- Dla wygładzania wykładniczego

$$\hat{y}_{T+h} = \hat{y}_T$$

- Dla podwójnego wygładzania wykładniczego

$$\hat{y}_{T+h} = \hat{y}_T + h\hat{m}_t$$

- Biały szum

$$y_t = \beta_0 + \varepsilon_t$$

gdzie $E(\varepsilon_t) = 0$, $Cov(\varepsilon_t, \varepsilon_{t-s}) = 0$ dla $t \neq s$ i $Var(\varepsilon_t) = \sigma^2$.

- Błędzenie losowe

$$y_t = y_{t-1} + \varepsilon_t$$

- zauważmy, że

$$y_t = y_0 + \sum_{s=1}^t \varepsilon_s$$

- Błądzenie losowe z dryfem

$$y_t = y_{t-1} + \beta_0 + \varepsilon_t$$

- zauważmy, że

$$y_t = y_0 + \beta_0 t + \sum_{s=1}^t \varepsilon_s$$

- O procesie stochastycznym y_t mówimy, że jest słabo stacjonarny jeśli spełnione są następujące założenia:

$$E(y_t) = \mu$$

$$\text{Var}(y_t) = \sigma^2$$

$$\text{Cov}(y_t, y_{t-h}) = \sigma_h$$

- Intuicyjnie proces stochastyczny jest stacjonarny jeśli struktura korelacji między y_t z różnych okresów nie zmienia się w czasie
- Proces stochastyczny y_t jest trendostacjonarny, jeśli wartość oczekiwana nie jest stała (podlega trendowi deterministycznemu) $E(y_t) = \mu_t$ ale

$$\text{Var}(y_t - \mu_t) = \sigma^2$$

$$\text{Cov}(y_t - \mu_t, y_{t-h} - \mu_t) = \sigma_h$$

- O silnej stacjonarności mówimy wtedy, gdy rozkład wektora zmiennych losowych $[y_t, \dots, y_{t+s}]$ nie zależy t dla każdego s

- Szczególnym rodzajem zmiennej stacjonarnej jest zmienna zintegrowana rzędu 0 $x_t \sim I(0)$. Zmienną $I(0)$ definiuje się jako zmienną stacjonarną.
- My ograniczamy się do przypadku, gdy $I(0)$ można przedstawić jako proces liniowy, czyli proces, który można przedstawić jako $x_t = \sum_{i=0}^{\infty} \theta_i \varepsilon_{t-i}$, gdzie ε_i i ε_j są nieskorelowane $i \neq j$ i stacjonarne.
- Zmienna zintegrowana rzędu d oznaczana jako $x_t \sim I(d)$, to taka zmienna, która po zastosowaniu d -tych różnic staje się zmienną zintegrowaną rzędu 0: $\Delta^d x_t \sim I(0)$.

- Przykładem zmiennej niestacjonarnej zmienna zachowująca się zgodnie z modelem błędzenia przypadkowego

$$y_t = y_{t-1} + \varepsilon_t$$

- Rekursywnie podstawiając

$$y_t = y_0 + \sum_{s=1}^t \varepsilon_s$$

- Jeśli założymy, że $y_0 = 0$, to

$$E(y_t) = 0$$

$$\text{Var}(y_t) = \sum_{s=1}^t \text{Var}(\varepsilon_s) = t\sigma^2$$

$$\text{Cov}(y_t, y_{t-h}) = \sum_{s=1}^{t-h} \text{Var}(\varepsilon_s) = (t-h)\sigma^2$$

- Zauważmy, że dla błędzenia przypadkowego korelacja między realizacjami maleje liniowo

$$\rho_h = \frac{\text{Var}(y_t)}{\text{Cov}(y_t, y_{t-h})} = \frac{t-h}{t}$$

Błądzenie przypadkowe, proces $I(1)$

- Wariancja i kowariancje błędzenia przypadkowego zależą od czasu - zmienna ta jest niestacjonarna! Zauważmy, że kowariancje między realizacjami zmiennej maleją bardzo wolno - w przybliżeniu liniowo.
- Odejmując od obu stron równania dla błędzenia losowego y_{t-1} otrzymujemy:

$$y_t - y_{t-1} = \Delta y_t = \varepsilon_t$$

- Po zastosowaniu pierwszych różnic otrzymaliśmy zmienną zachowującą się zgodnie z modelem białego szumu - zmienną $I(0)$. Wnioskujemy z tego, że błędzenie przypadkowe jest $I(1)$

- Rozważmy model

$$y_t = \mu + \alpha y_{t-1} + \varepsilon_t$$

dla $\varepsilon_t \sim IID(0, \sigma^2)$ i $|\alpha| < 1$

- Stosując rekursywne podstawianie uzyskujemy

$$y_t = \sum_{i=0}^{\infty} \alpha^i \mu + \sum_{i=0}^{\infty} \alpha^i \varepsilon_{t-i} = \frac{\mu}{1-\alpha} + \sum_{i=0}^{\infty} \alpha^i \varepsilon_{t-i}$$

- Weryfikacja założeń stacjonarności

$$E(y_t) = \frac{\mu}{1 - \alpha}$$

$$\text{Var}(y_t) = \frac{\sigma^2}{1 - \alpha^2}$$

$$\text{Cov}(y_t, y_{t-h}) = \frac{\alpha^h \sigma^2}{1 - \alpha^2}$$

- Zauważmy, że w przypadku procesu AR(1) korelacja między realizacjami maleje wykładniczo

$$\rho_h = \frac{\text{Cov}(y_t, y_{t-h})}{\text{Var}(y_t)} = \alpha^h$$

- W ekonomii ważne są jedynie zmienne $I(1)$ i $I(2)$
- Wiele zmiennych w ekonomii wygląda tak na zmienne $I(1)$.
- W znanym artykule Nelson i Plosser (1982) pokazali, że znacząca część szeregów makroekonomicznych wygląda na szeregi $I(1)$

- Funkcja autokorelacji *ACF* (**A**utocorrelation **F**unction)
- Funkcja *ACF* dla pewnego k , oznaczane jako ρ_k , jest równa

$$\rho_k = \frac{\text{Cov}(y_t, y_{t-k})}{\text{Var}(y_t)}$$

- Funkcja autokorelacji pozwala stwierdzić jak silna jest korelacja między poszczególnymi realizacjami zmiennej y_t .

- Funkcja autokorelacji cząstkowej *PACF* (Partial Autocorrelation Function).
- Funkcja *PACF* dla pewnego k , oznaczana jako α_{kk} , jest równa wyestymowanemu współczynnikowi α_k w modelu $AR(k)$

$$y_t = \mu + \alpha_1 y_{t-1} + \dots + \alpha_k y_{t-k} + \varepsilon_t$$

- Funkcja autokorelacji cząstkowej pozwala zbadać istotność odległych opóźnień.

Test DF (Dickey-Fullera)

- Badamy, czy zmienna jest niestacjonarna
- Najprostszy przypadek

$$y_t = \beta y_{t-1} + \varepsilon_t$$

- $H_0 : \beta = 1$, wtedy y_t jest niestacjonarna (jest błędzeniem przypadkowym - zmienną $I(1)$)
- $H_1 : |\beta| < 1$, wtedy y_t jest stacjonarna
- Przekształcając

$$\begin{aligned}\Delta y_t &= (\beta - 1)y_{t-1} + \varepsilon_t \\ &= \rho y_{t-1} + \varepsilon_t\end{aligned}$$

- $H_0 : \rho = 0$, wtedy y_t jest niestacjonarna
- $H_1 : \rho \in (-2, 0)$, wtedy y_t jest stacjonarna

- Hipotezy, że $\rho = 0$ nie można badać za pomocą standardowych tablic - używamy do tego specjalnych tablic dla testu *DF*
 - Sposób przeprowadzania testu:
 - 1 Przeprowadzamy regresję Δy_t na y_{t-1}
 - 2 Porównujemy statystykę t dla y_{t-1} z wartościami krytycznymi testu *DF*. Jeśli statystyka jest mniejsza od wartości krytycznej odrzucamy H_0 .
- W teście Dickey Fullera:
 - H_0 zmienna jest *niestacjonarna*
 - H_1 zmienna jest *stacjonarna*

Test ADF (Rozszerzony test Dickey-Fullera)

- Często w regresji $\Delta y_t = \rho y_{t-1} + \varepsilon_t$ występuje autokorelacja
- Autokorelacja w regresjach z opóźnioną zmienną zależną jest bardzo niekorzystna - prowadzi do wystąpienia problemu równoczesności.
- Stosujemy wtedy test *DF* z rozszerzeniem (test *ADF* - **A**ugmented **D**ickey **F**uller)

$$\Delta y_t = \rho y_{t-1} + \sum_{i=1}^k \gamma_i \Delta y_{t-i} + \varepsilon_t$$

- Ilość opóźnień k dobieramy w taki sposób, by w modelu nie występowała autokorelacja.
- Podobnie jak w teście *DF* do testowania używamy statyki t (oznaczanej czasami jako z) policzonej dla parametru ρ
- Hipoteza zerowa i alternatywna identyczne jak w teście *DF*

- Sposób uwzględnienia stałej lub trendu w regresji ADF:

① $\Delta y_t = \rho y_{t-1} + \sum_{i=1}^k \gamma_i \Delta y_{t-i} + \varepsilon_t$

② $\Delta y_t = a + \rho y_{t-1} + \sum_{i=1}^k \gamma_i \Delta y_{t-i} + \varepsilon_t$

③ $\Delta y_t = a + bt + \rho y_{t-1} + \sum_{i=1}^k \gamma_i \Delta y_{t-i} + \varepsilon_t$

- Dla każdej z tych możliwości stosuje się inną wersję tablic *ADF*

- Możliwe jest także testowanie stacjonarności za pomocą testów, w których H_0 mówi, że zmienna jest $I(0)$
- Tego rodzaju testem test *KPSS*
- W modelu

$$y_t = \delta + \zeta_t + \varepsilon_t$$

gdzie ε_t jest $I(0)$ oraz

$$\zeta_t = \zeta_{t-1} + u_t, u_t \sim IID(0, \sigma_u^2)$$

a więc ζ_t jest błędzeniem przypadkowym.

- Testujemy $H_0 : \sigma_u = 0$ co jest równoważnie hipotezie, że $\zeta_t = 0$.

- Jeśli H_0 jest prawdziwa, błędzenie przypadkowe nie odgrywa roli i y_t jest stacjonarne.
- Jeśli odrzucimy H_0 , to musimy przyjąć hipotezę alternatywną, że zmienna jest niestacjonarna.
- Często się zdarza (szczególnie jeśli mamy mało obserwacji), że zarówno H_0 w teście DF (niestacjonarność) jak H_0 w teście $KPSS$ (stacjonarność), nie mogą być odrzucone.
- Sposób pozbywania się wpływu korelacji między ε_t jest w przypadku testu $KPSS$ dosyć skomplikowany.

Wielomiany operatora opóźnień

- Zdefiniowaliśmy

$$L^s x_t = x_{t-s}$$

- Ponieważ

$$\left[1 + aL + (aL)^2 + \dots + (aL)^n\right] (1 - aL) = 1 - (aL)^{n+1}$$

więc

$$1 + aL + (aL)^2 + \dots + (aL)^n = (1 - aL)^{-1} \left[1 - (aL)^{n+1}\right]$$

- Jeśli $|a| < 1$ to

$$(1 - aL)^{-1} = \sum_{i=0}^{\infty} (aL)^i$$

- Wniosek: $1 - aL$ można odwrócić jeśli $|a| < 1$

- Model

$$y_t = \mu + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p} + \varepsilon_t$$

można zapisać:

$$A(L)y_t = \mu + \varepsilon_t$$

- gdzie $A(L)$ jest wielomianem operatora opóźnień:

$$A(L) = 1 - a_1 L - a_2 L^2 - \dots - a_p L^p$$

Wielomiany operatora opóźnień c.d.

- Wielomian o tych samych współczynnikach dla liczby zespolonej x

$$A(x) = 1 - a_1x - a_2x^2 - \dots - a_px^p$$

takie wielomian można sfaktoryzować następująco (fundamentalne twierdzenie algebry):

$$A(x) = (1 - \lambda_1x) \times \dots \times (1 - \lambda_px)$$

gdzie $\mu_i = \lambda_i^{-1}$ są pierwiastkami tego wielomianu (pierwiastki te mogą być zespolone)

- Identyczna faktoryzacja będzie możliwa dla wielomianu operatora opóźnień

$$A(L) = (1 - \lambda_1L) \times \dots \times (1 - \lambda_pL)$$

- $1 - \lambda_i L$ może być odwrócony jeśli $|\lambda_i| < 1$.
- $A(L)$ może być odwrócony jeśli $|\lambda_i| < 1$ (albo $|\mu_i| > 1$) dla $i = 1, \dots, p$, jego pierwiastki leżą *poza* kołem jednostkowym
- Odwracalny $A(L)^{-1}$ może zostać zapisany następująco

$$\begin{aligned} A(L)^{-1} &= (1 - \lambda_1 L)^{-1} \times \dots \times (1 - \lambda_s L)^{-1} \\ &= \left[\sum_{i=1}^{\infty} (\lambda_1 L)^i \right] \times \dots \times \left[\sum_{i=1}^{\infty} (\lambda_p L)^i \right] \\ &= \sum_{i=0}^{\infty} \psi_i L^i \end{aligned}$$

gdzie $\psi_0, \psi_1, \psi_2, \dots$ są współczynnikami malejącymi wykładniczo do zera.

- Test HEGY: decompozycja wielomianu opóźnień dla sezonowości kwartalnej

$$\begin{aligned}\Delta_4 = 1 - L^4 &= (1 - L^2)(1 + L^2) \\ &= (1 - L)(1 + L)(1 - iL)(1 + iL) \\ &= (1 - L)(1 + L + L^2 + L^3) \\ &= (1 + L)(1 - L + L^2 - L^2)\end{aligned}$$

- Pierwiastki wielomianu opóźnień
 - $\mu_1 = 1, 1 - L$, pierwiastek niesezonowy
 - $\mu_2 = -1, 1 + L$, jeśli $\mu_1 = 1$ i , to $(1 - L)(1 + L) = (1 - L^2)$ pierwiastki sezonowe półroczne
 - $\mu_1 = 1, \mu_4 = -1, \mu_3 = i$ i $\mu_4 = -i$, to $(1 - L)(1 + L)(1 - iL)(1 + iL) = (1 - L^4)$ pierwiastki sezonowe kwartalne

- Sprawdzamy, czy w szeregu występuje pierwiastek $\mu = 1$. W najprostszym przypadku $H_0 : \mu_1 = 1$:

$$(1 - \mu_1 L)(1 + L + L^2 + L^3)y_t = \varepsilon_t$$

- Testowanie pierwiastka niesezonowego

$$\Delta_4 y_t = (\mu_1 - 1)L(1 + L + L^2 + L^3)y_t + \varepsilon_t = \rho y_{1,t-1} + \varepsilon_t$$

gdzie $\rho = \mu_1 - 1$ i $y_{1,t} = (1 + L + L^2 + L^3)y_t$, $H_0 : \rho = 0$

- Sprawdzanie istnienia pozostałych pierwiastków przebiega analogicznie
- Regresje definiujemy w ten sposób, by można było także przetestować łącznie istnienie poszczególnych pierwiastków

- Regresja HEGY

$$\begin{aligned}\Delta_4 y_t &= \pi_1 y_{1,t-1} + \pi_2 y_{2,t-1} + \pi_3 y_{3,t-1} + \pi_4 y_{3,t-2} \\ &+ \sum_{s=1}^4 \alpha_s D_{st} + \gamma T + \sum_{i=1}^k \phi_i \Delta_4 y_{t-i} + \varepsilon_t\end{aligned}$$

gdzie

$$\begin{aligned}y_{1,t} &= (1 + L + L^2 + L^3) y_t \\ y_{2,t} &= -(1 - L + L^2 - L^2) y_t \\ y_{3,t} &= -(1 - L^2) y_t\end{aligned}$$

- Hipotezy testowane:

- $H_0 : \pi_1 = 0$ pierwiastek jednostkowy niesezonowy
- $H_0 : \pi_1 = 0, \pi_2 = 0$ pierwiastki sezonowe półroczne
- $H_0 : \pi_1 = 0, \pi_2 = 0, \pi_3 = 0, \pi_4 = 0$ pierwiastki sezonowe kwartalne

- Proces $ARMA(p, q)$ (**A**uto**R**egressive **M**oving **A**varege) można także zapisać następująco

$$y_t = \mu + \underbrace{\alpha_1 y_{t-1} + \dots + \alpha_p y_{t-p}}_{AR} + \underbrace{\varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}}_{MA}$$

i zakładamy, że innowacje ε_t są nieskorelowane $E(\varepsilon_t, \varepsilon_s) = 0$.

- Zapis modelu $ARMA(p, q)$ przy użyciu wielomianów operatora opóźnień

$$A(L)y_t = \mu + B(L)\varepsilon_t$$

$$A(L) = 1 - \alpha_1 L - \alpha_2 L^2 - \dots - \alpha_p L^p$$

$$B(L) = 1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q$$

- Model

$$y_t = \mu + \alpha_1 y_{t-1} + \dots + \alpha_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

nie można oszacować MNK ponieważ ε_{t-i} są nieobserwowalne

- Definiując $u_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$ uzyskujemy model:

$$y_t = \mu + \alpha_1 y_{t-1} + \dots + \alpha_p y_{t-p} + u_t$$

ale w tym modelu u_t będzie wykazywała autokorelację a w konsekwencji y_t będzie skorelowane z u_t

- Rzeczywiście n.p. zmienna objaśniająca y_{t-1} w modelu dla y_t zależy od ε_{t-1} ale od ε_{t-1} zależy także u_t a zatem y_{t-1} jest skorelowane z u_t i mamy problem endogeniczności

- Jeśli $q = 0$, to model $AR(p)$, można wyestymować *MNK*.
- Jeśli $q \neq 0$, estymatory *MNK* asymptotycznie obciążone (równoczesność).
- Należy zastosować estymatory *MNW* (Metodę Największej Wiarygodności).
- Do testowania hipotez można zastosować testy *LM*, *LR* lub *W*.
- Ustalenie wielkości parametrów p i q .
 - na bazie kształtu funkcji ADF i PAF
 - kryteria informacyjne
 - od ogólnego do szczegółowego

Własności ACF i PACF dla modeli ARMA(p,q)

- Kształty funkcji *ACF* i *PACF* zależą od wielkości parametrów p i q .
- Wykresów *ACF* i *PACF* używa się do celów diagnostycznych.
- Można udowodnić, że:
- W przypadku, gdy $p > 0$
 - funkcja *ACF* maleje wykładniczo
 - funkcja *PACF* istotna jedynie dla p pierwszych wartości α_{kk}
- W przypadku gdy $q > 0$
 - funkcja *ACF* istotna dla pierwszych q wartości ρ_k
 - funkcja *PACF* maleje wykładniczo

Ustalanie p i q na podstawie kształtu ACF i $PACF$

- Wielkość p i q można mniej więcej oszacować na podstawie kształtu funkcji ACF i $PACF$
- Przy ostatecznym ustalaniu wielkości p i q lepiej jednak posłużyć się formalnymi procedurami testowania
- Jeśli funkcja ACF i $PACF$ nie maleje powinniśmy d razy zróżnicować y_t , tak by ACF i $PACF$ zaczęły maleć - ilość takich różnicowań można także ustalić na podstawie testów
- Jeśli w ACF i $PACF$ pojawiają się regularne skoki w krotnościach równych próbkowaniu danych (np. dla danych kwartalnych w wielokrotnościach 4), sugerować to będzie sezonowość. Można ją usunąć stosując różnicowanie sezonowe lub umieszczając elementy sezonowe (model $SARIMA(p,d,q)(P,D,Q)$)

- Model $ARMA(p, q)$

$$y_t = \mu + \alpha_1 y_{t-1} + \alpha_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

- Nie znamy wielkości p, q .
- Możemy uzyskać zgodne oszacowania tylko wtedy jeśli w modelu nie występuje autokorelacja.
- Celem jest znalezienie takiego modelu, który miałby możliwie mało parametrów (małe p, q), i w którym nie występuje autokorelacja.
- Szukamy modelu spośród modeli branych pod uwagę, który ma najniższą wartość kryteriów informacyjnych

- Metoda od ogólnego do szczegółowego:
- wybieramy największe sensowne p, q (zazwyczaj wystarczy $p = 3, q = 3$)
 - 1 testujemy sekwencyjnie łączną nieistotność coraz większej liczby ostatnich opóźnień (n.p. $\alpha_3 = 0$ i $\theta_3 = 0$, następnie $\alpha_3 = \alpha_2 = 0$ i $\theta_3 = 0$, następnie $\alpha_3 = \alpha_2 = 0$ i $\theta_3 = \theta_2 = 0$ etc.)
 - 2 Wybieramy pierwszy model, dla którego musieliśmy odrzucić H_0 o nieistotności opóźnień
- Kryteria informacyjne: spośród modeli wybieramy ten, dla którego wartość kryterium informacyjnego była najniższa.
- Na koniec testujemy, czy reszty nie są skorelowane - możemy do tego użyć testu *partmanteau* nazywanego także testem Ljunga-Boxa, H_0 : zmienne jest białym szumem (ciągami nieskorelowanych zmiennych losowych)

- Czy istnieją miary dopasowania, które uwzględniałyby ilość traconych w procesie estymacji stopni swobody?
- Statystyki takie istnieją i noszą ogólną nazwę kryteriów informacyjnych.
- Jedną z takich statystyk jest skorygowane R^2 jednak tę statystykę można zdefiniować jedynie dla modelu liniowego
- Dla modeli szacowanych na bazie modelu statystycznego można znaleźć tak zwane kryteria informacyjne

- Najbardziej popularnymi z nich jest kryterium informacyjne Akaike *AIC* (**A**kaike **I**nformation **C**riterion)
- Bayesowskie kryterium informacyjne Schwartz'a *BIC* (**B**ayes **I**nformation **C**riterion) - określane skrótami *SC*, *SBC*, *SIC*
- W modelu szacowanym *MNK* kryteria ta dane są wzorami:

$$BIC = -2\ell(\hat{\theta}) + K \log(T)$$

$$AIC = -2\ell(\hat{\theta}) + 2K$$

$$AIC_c = AIC + \frac{2K(K+1)}{T-K-1}$$

gdzie K to liczba parametrów a T to liczba obserwacji

- Za najlepszy uznaje się ten model, dla którego kryterium informacyjne uzyskuje *najmniejszą* wartość.
- Pokazano, że dla pewnych ogólnych założeń, dla $T \rightarrow \infty$ wybrany na podstawie *AIC*, *AIC_c*, *BIC* model zawierać będzie poprawny zbiór zmiennych objaśniających.

- Proste narzędzie służące do prognozowania przyszłych y_t
- Szczególną zaletą tej metodologii prognozowania jest jej prostota
- Rozpatrzmy wzór na y_{t+1}

$$y_{T+1} = \mu + \alpha_1 y_T + \alpha_p y_{T-p+1} + \varepsilon_{T+1} + \theta_1 \varepsilon_T + \dots + \theta_q \varepsilon_{T-q+1}$$

- Wartość ε_{T+1} jest niemożliwa do oszacowania (ε_t są nieskorelowane w czasie) więc ją pomijamy
- $\varepsilon_t, \dots, \varepsilon_{t-q+1}$ zastępujemy przez ich oszacowania e_t, \dots, e_{t-q+1}

- Oczywistym wzorem na prognozę y_t dla okresu $T + 1$ jest więc:

$$\begin{aligned}\hat{y}_{T+1} &= \hat{\mu} + \hat{\alpha}_1 y_T + \hat{\alpha}_2 y_{T-1} + \dots + \hat{\alpha}_p y_{T-p+1} \\ &\quad + \hat{\theta}_1 e_T + \hat{\theta}_2 e_{T-2} + \dots + \hat{\theta}_q e_{T-q+1}\end{aligned}$$

- By sformułować prognozę y_t na okres $t + 2$ możemy posłużyć prognozą dla \hat{y}_{t+1}

$$\begin{aligned}\hat{y}_{T+2} &= \hat{\mu} + \hat{\alpha}_1 \hat{y}_{T+1} + \hat{\alpha}_2 y_T + \dots + \hat{\alpha}_p y_{T-p+2} \\ &\quad + \hat{\theta}_2 e_T + \dots + \hat{\theta}_q e_{T-q+2}\end{aligned}$$

- Prognozę dla \hat{y}_{T+s} uzyskujemy podstawiając za $y_{T+1}, \dots, y_{T-p+s}$ rekurencyjnie policzone prognozy $\hat{y}_{T+1}, \dots, \hat{y}_{T-p+s}$

- Dla dużych s wpływ y_T, y_{T-1}, \dots i $\varepsilon_T, \varepsilon_{T-1}, \dots$ na y_{T+s} maleje do zera
- Z tego powodu prognozy wygenerowane dla procesu *ARMA* zbiegają do $y^* = \mu^* = \frac{\mu}{1 - \alpha_1 - \dots - \alpha_k}$

- W modelach ARIMA problem niestacjonarności rozwiązuje się poprzez d-krotne różnicowanie zmiennej modelowanej
- Uzyskujemy w rezultacie model ARIMA(p,d,q)

$$\Delta^d y_t = \mu + \alpha_1 y_{t-1} + \dots + \alpha_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

bądź

$$A(L) \Delta^d x_t = B(L) \varepsilon_t$$

- W kontekście modelowania za pomocą modeli *ARMA* problem niestacjonarności wykrywany był za pomocą funkcji *ACF* i *PACF*.
- Zmienne niestacjonarne charakteryzują się
 - wysoką korelacją między realizacjami - wolno malejąca funkcja *ACF*
 - bliską jedności wartością współczynniki przy pierwszym opóźnieniu - *PACF* dla pierwszego opóźnienia bliskie jedności

- Jeśli funkcja ACF nie maleje powinniśmy d razy zróżnicować y_t , tak by ACF i $PACF$ zaczęły maleć - ilość takich różnicowań można także ustalić na podstawie testów na występowanie pierwiastków jednostkowych
- Jeśli w ACF i $PACF$ pojawiają się regularne skoki w krotnościach równych próbkowaniu danych (np. dla danych kwartalnych w wielokrotnościach 4), sugerować to będzie sezonowość.

- W modelu *SARIMA* uwzględniamy dodatkowo elementy sezonowe
- Sezonowość może być uwzględniona zarówno przez różnicowanie sezonowe jak i umieszczenie w modelu sezonowych elementów autoregresyjnych bądź związanych ze średnią ruchomą
- Zapis modelu *SARMIA*(p, d, q)(P, D, Q, s)

$$A(L)A_s(L)\Delta^d\Delta_s^D x_t = B(L)B_s(L)\varepsilon_t$$

$$A(L) = 1 - \alpha_1 L - \alpha_2 L^2 - \dots - \alpha_p L^p$$

$$B(L) = 1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_p L^p$$

$$A_s(L) = 1 - \alpha_{s,1} L^s - \alpha_{s,2} L^{2s} - \dots - \alpha_{s,p} L^{Ps}$$

$$B_s(L) = 1 + \theta_{s,1} L^s + \theta_{s,2} L^{2s} + \dots + \theta_{s,Q} L^{Qs}$$

- Do oszacowania s , P i Q możemy posłużyć się funkcjami ACF i $PACF$. Efekty sezonowe pojawiają się w nich jako periodyczne (z okresem s) spadki lub wzrosty
- Liczbę różnicowań Q ustalamy na bazie wyników testów na sezonowe pierwiastki jednostkowe
- Do ustalenia prawidłowej wielkości p, d, P, Q możemy użyć kryteriów informacyjnych bądź metody od ogólnego do szczegółowego

- Ważne przy estymacji modeli autoregresyjnych jest brak autokorelacji, ponieważ skutkuje ona brakiem zgodności estymatorów
- Testem na brak autokorelacji w przypadku modeli $ARIMA(p, d, q)$ jest test Ljunga-Box'a (patmanteau):

$$Q = n(n+2) \sum_{k=1}^m \frac{\hat{\rho}_k^2}{n-k} \xrightarrow{D} \chi_{m-p-q}^2,$$

gdzie

$$\hat{\rho}_k = \frac{(T-k) \sum_{t=k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y})}{T^{-1} \sum_{t=1}^T (y_t - \bar{y})^2}$$

jest współczynnikiem korelacji rzędu k a m jest liczbą $\hat{\rho}_k$ uwzględnionych w trakcie liczenia statystyki.

- Model, w którym występuje jedynie część *DL* (**D**istributed **L**ags)
- Prosty w analizie jeśli \mathbf{x}_t i opóźnione \mathbf{x}_t są nielosowe
- Model może spełniać założenia *KMRL*.

$$y_t = \mu + \beta_0 \mathbf{x}_t + \dots + \beta_p \mathbf{x}_{t-p} + \varepsilon_t = \mu + D(L) \mathbf{x}_t + \varepsilon_t$$

- W przypadku modeli dynamicznych współczynniki przy zmiennych objaśniających opisują reakcję zmiennej zależnej z okresu t na zmiany zmiennych niezależnych z okresu t oraz okresów wcześniejszych
- W związku z tym przy interpretacji tych współczynników modelu należy dokładnie określić czy chodzi nam o:
 - zależności między zmiennymi w położeniu równowagi
 - natychmiastową (krótkookresową) reakcje zmiennej zależnej na zmianę w zmiennej niezależnej

- Wielkość natychmiastowej reakcję y_t na zmianę jednostkową zmianę x_t nazywamy mnożnikiem bezpośrednim (*impact multiplier*)
- Jeśli x_t pozostanie na tym wyższym poziomie także w τ kolejnych okresach, to wielkość reakcja y_t na taką zmianę x_t nazywamy mnożnikiem skumulowanym (*cumulated multiplier*)
- Wielkość wpływu trwałej zmiany x_t na y_t nazywamy mnożnikiem długookresowym (*long-run multiplier*)

- W przypadku modeli *DL*:
 - Wpływ zmiany \mathbf{x}_t o $\Delta\mathbf{x}_t$ powoduje zmianę y_t o Δy_t równą:

$$\begin{aligned}E(y_t + \Delta y_t) &= \mu + \beta_0(\mathbf{x}_t + \Delta\mathbf{x}_t) + \dots + \beta_p\mathbf{x}_{t-p} \\ &= E(y_t) + \beta_0\Delta\mathbf{x}_t\end{aligned}$$

Wpływ ten mierzymy *mnożnikiem bezpośrednim*:

$$\frac{E(\Delta y_t)}{\Delta\mathbf{x}_t} = \beta_0$$

- Wpływ zmiany \mathbf{x} o $\Delta \mathbf{x}$, która nastąpiła τ okresów temu powoduje zmianę y_t o Δy_t równą

$$\begin{aligned} E(y_t + \Delta y_t) &= \mu + \beta_0(\mathbf{x}_t + \Delta \mathbf{x}) + \dots + \beta_\tau(\mathbf{x}_{t-\tau} + \Delta \mathbf{x}) \\ &\quad + \beta_{\tau+1}\mathbf{x}_{t-\rho+1} + \dots + \beta_\rho\mathbf{x}_{t-\rho} \\ &= E(y_t) + \left(\sum_{i=0}^{\tau} \beta_i \right) \Delta \mathbf{x} \end{aligned}$$

Zmianę tę mierzymy *mnożnikiem skumulowanym*:

$$\beta_\tau = \frac{E(\Delta y_{t+\tau})}{\Delta \mathbf{x}} = \sum_{i=0}^{\tau} \beta_i$$

- Długookresowy wpływ zmiany \mathbf{x} o $\Delta \mathbf{x}$ mierzymy *mnożnikiem długookresowym*. Jest on równy mnożnikowi skumulowanemu dla $\tau \rightarrow \infty$

$$\frac{E(\Delta y)}{\Delta \mathbf{x}} = \beta = \sum_{i=0}^{\infty} \beta_i$$

- Miarą szybkości reakcji zmiennej zależnej na zmiany zmiennych

Modele autoregresyjne o rozłożonych opóźnieniach (ADL)

- Często okazuje się, że wprowadzając do modelu jako zmienną objaśniającą, opóźnioną zmienną zależną uzyskuje się model o lepszym dopasowaniu i mniejszej ilości parametrów
- Można to wytłumaczyć między innymi tym, że wiele zjawisk ekonomicznych charakteryzuje się sporą inercją
- Model, w którym występują opóźnione zmienne zależne nazywamy modelem autoregresyjnym
- Ogólną klasą takich modeli są modele *ADL* (**A**utoregressive **D**istributed **L**ags) postaci

$$y_t = \underbrace{\alpha_1 y_{t-1} + \dots + \alpha_p y_{t-p}}_{AR} + \underbrace{\mu + \beta_0 x_t + \beta_1 x_{t-1} + \dots + \beta_s x_{t-s}}_{DL} + \varepsilon_t$$

- Model *ADL* można też zapisać za pomocą wielomianów operatora opóźnień

$$A(L)y_t = \mu + \mathbf{D}(L)\mathbf{x}_t + \varepsilon_t$$

gdzie

$$A(L) = 1 - \alpha_1 L - \alpha_2 L^2 - \dots - \alpha_p L^p$$

$$\mathbf{D}(L) = \beta_0 + \beta_1 L + \dots + \beta_s L^s$$

- Tak samo jak w przypadku modeli *DL* interesują nas dwa typy mnożników: bezpośredni i długookresowy.
- Mnożnik bezpośredni mierzy wpływ zmiany \mathbf{x}_t o $\Delta\mathbf{x}_t$ na oczekiwany poziom y_t . Oznaczmy zmianę y_t jako Δy_t .

$$E(y_t + \Delta y_t) = E(y_t) + \beta_0 \Delta \mathbf{x}_t$$

$$\frac{E(\Delta y_t)}{\Delta \mathbf{x}_t} = \beta_0$$

a więc mnożnik krótkookresowy jest równy β_0

- Zmieniamy $\mathbf{x}_t, \mathbf{x}_{t-1}, \dots$ o $\Delta \mathbf{x}$. W tym przypadku oczekiwana zmiana Δy_t

$$\begin{aligned} E(y_t + \Delta y_t) &= \alpha_1 E(y_{t-1} + \Delta y_{t-1}) + \dots + \alpha_p E(y_{t-p} + \Delta y_{t-p}) \\ &\quad + \mu + (\mathbf{x}_t + \Delta \mathbf{x}) \beta_0 \\ &\quad + (\mathbf{x}_{t-1} + \Delta \mathbf{x}) \beta_1 + \dots + (\mathbf{x}_{t-s} + \Delta \mathbf{x}) \beta_s \end{aligned}$$

jest równa

$$\begin{aligned} E(\Delta y_t) &= \alpha_1 E(\Delta y_{t-1}) + \dots + \alpha_p E(\Delta y_{t-p}) \\ &\quad + (\beta_0 + \beta_1 + \dots + \beta_s) \Delta \mathbf{x} \end{aligned}$$

- Wpływ zmiany x na y w długim okresie stabilizuje się:

$$E(\Delta y) = E(\Delta y_t) = E(\Delta y_{t-1}) = \dots = E(\Delta y_{t-p})$$

- Długookresowy mnożnik β jest więc równy

$$\frac{E(\Delta y)}{\Delta x} = \beta = \frac{\beta_0 + \beta_1 + \dots + \beta_s}{1 - \alpha_1 - \dots - \alpha_p} = A(1)^{-1} D(1)$$

- Stan równowagi długookresowej (rozwiązanie długookresowe) jest to stan, w którym wartość oczekiwana zmiennej zależnej pozostaje stała w czasie o ile nie zmieniają się zmienne niezależne.
- Rozwiązanie długookresowe jest bardzo ważne , ponieważ większość teorii ekonomicznych dotyczy relacji między zmiennymi w stanie równowagi.
- Stan równowagi długookresowej znajdujemy używając jego definicji definicji, a więc zakładając, że wartość oczekiwana zmiennej zależnej i wartości zmiennych niezależnych są stałe w czasie

$$y^* = E(y_t) = E(y_{t-1}) = \dots = E(y_{t-p})$$

$$x^* = x_t = x_{t-1} = \dots = x_{t-s}$$

Modele ADL: stan równowagi długookresowej

- Wstawiając to do definicji modelu *ADL* otrzymujemy:

$$(1 - \alpha_1 - \dots - \alpha_p) y^* = \mu + \beta_0 \mathbf{x}^* + \beta_1 \mathbf{x}^* + \dots + \beta_s \mathbf{x}^*$$

- Dla modelu *ADL* implikuje to, że równowaga długookresowa zachodzi, gdy spełnione są następujące zależności między zmiennymi:

$$y^* = \mu^* + \mathbf{x}^* \boldsymbol{\beta} = A(1)^{-1} \mu + A(1)^{-1} \mathbf{D}(1) \mathbf{x}^*$$

gdzie $\mu^* = \frac{\mu}{1 - \alpha_1 - \dots - \alpha_p} = A(1)^{-1} \mu$ a

$$\boldsymbol{\beta} = \frac{\beta_0 + \beta_1 + \dots + \beta_s}{1 - \alpha_1 - \dots - \alpha_p} = A(1)^{-1} \mathbf{D}(1).$$

- Zauważmy, że $\boldsymbol{\beta}$ jest dokładnie równa mnożnikowi długookresowemu, który znaleźliśmy wcześniej
- Dla modelu czysto autoregresyjnego, rozwiązanie długookresowe będzie miało postać

$$y^* = \mu^*$$

- W ogólniejszym przypadku mamy do czynienia z modelami *ADL* z błędem losowym w postaci *ARMA*(p, q)

$$\Delta^d y_t = \mu + \beta_0 \mathbf{x}_t + \dots + \beta_s \mathbf{x}_{t-s} + u_t$$

$$u_t = \alpha_1 u_{t-1} + \dots + \alpha_p u_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

- Model *ARIMAX* jest więc modelem $D(L)$ z błędem losowym, mającym postać procesu *ARIMA*(p, q)

- Zapis przy użyciu wielomianów operatora opóźnień

$$A(L) u_t = B(L) \varepsilon_t$$

$$u_t = A^{-1}(L) B(L) \varepsilon_t$$

$$y_t = \mu + \mathbf{D}(L) \mathbf{x}_t + A^{-1}(L) B(L) \varepsilon_t$$

- Równoważna reprezentacja, model *ADL* z błędem v_t w postaci procesu $MA(q)$

$$y_t = \mu^* + \alpha_1 y_{t-1} + \dots + \alpha_p y_{t-p} + \beta_0 \mathbf{x}_t + \dots + \beta_s \mathbf{x}_{t-s} \\ + \underbrace{\varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}}_{v_t}$$

$$A(L)y_t = \mu^* + \mathbf{D}^*(L)\mathbf{x}_t + B(L)\varepsilon_t$$

gdzie $\mu^* = A(1)\mu$, $\mathbf{D}^*(L) = A(L)\mathbf{D}(L)$

- Analiza dynamicznych własności błędów losowych w tych modelach wygląda identycznie jak analiza własności y_t w modelu *ARMA*.

- Co to jest przyczyna i skutek?
- Ważna cecha przyczynowości:
 - zdarzenia wcześniejsze powodują zdarzenia późniejsze
 - zajście przyczyny umożliwia przewidzenie zajścia skutku
- Przyczynowość w sensie Grangera:

Definicja

Zmienna jest przyczyną w sensie Grangera jeśli bieżące wartości zmiennej y można dokładniej prognozować przy użyciu przeszłych wartości x niż bez ich wykorzystania.

- W modelu *ADL*

$$y_t = a(t) + \sum_{i=1}^k \alpha_i y_{t-i} + \sum_{i=1}^k \beta_i x_{t-i} + \varepsilon_t$$

gdzie $a(t)$ jest częścią deterministyczną modelu (n.p. $a(t) = \gamma_0 + \gamma_1 t$), jeśli $\beta_1 = \beta_2 = \dots = \beta_k = 0$, to x nie jest przyczyną w sensie Grangera y

Test przyczynowości Grangera

Test hipotezy

H_0 : x nie jest przyczyną w sensie Grangera y

przeprowadzamy testując hipotezę łączną, że wszystkie współczynniki przy opóźnionych x są równe zero:

$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$

- Zmienne objaśniające w modelu *ADL* muszą być z góry ustalone (egzogeniczne lub endogeniczne opóźnione)
- Jeśli modelu występują opóźnione zmienne zależne i autokorelacja, to wystąpi równoczesność
 - estymator \mathbf{b}_{MNK} będzie w tym przypadku nie będzie zgodny
- Z tego powodu w modelach tych trzeba zawsze dążyć do usunięcia autokorelacji
- Do testowania autokorelacji można używać testu Breuscha-Godfrey
- W modelach z częścią autoregresyjną *nie powinno* używać się testu *DW*, może dawać nieprawidłowe wyniki
- Autokorelacje można zazwyczaj usunąć zwiększając liczbę opóźnień

- O zmiennych $x_{1t} \sim I(1)$ i $x_{2t} \sim I(1)$ mówimy, że są skointegrowane jeśli istnieje takie β , że $x_{1t} + \beta x_{2t}$ jest $I(0)$.
- W przypadku wektora \mathbf{x}_t , jeśli każdy element wektora \mathbf{x}_t jest $I(1)$, to mówimy, że elementy tego wektora są skointegrowane wektorem kointegrującym β , jeśli $\beta' \mathbf{x}_t$ jest $I(0)$
- Istnieje taka kombinacja liniowa elementów wektora \mathbf{x}_t , która jest $I(0)$, mimo że elementy tego wektora są $I(1)$

- Problem regresji pozorną może się pojawić jeśli zmienne w modelu nie są $I(0)$
- Może to dotyczyć zarówno zmiennych objaśnianych jak i objaśniających
- W regresji ze zmiennymi $I(1)$ uzyskujemy w dużych próbach istotne statystyki t w MNK nawet jeśli zmienna endogeniczna i zmienne egzogeniczne są od siebie niezależne. Estymator MNK w takiej regresji jest dalej zgodny.

Eksperyment Grangera i Newbolda (1974)

- Przeprowadźmy regresję y_t na x_t , gdzie y_t i x_t są *niezależnymi* zmiennymi $I(1)$.

$$y_t = \beta_0 + \beta_1 x_t + e_t$$

- Wartości p z rozkładu t -studenta i uzyskane z symulacji dla hipotezy $H_0 : \beta_1 = 0$.

liczba obserwacji	t_α	p-stwo, że $ t > t_\alpha$	
		t-student	symulacja
10	2.31	0.05	0.28
100	1.98	0.05	0.76
1000	1.96	0.05	0.93

- Dla dużej liczby obserwacji jest niemal pewne, że $|t| > t_\alpha$ i że w konsekwencji odrzucimy prawdziwą hipotezę zerową o braku związku między zmiennymi!
- Problem regresji pozornej jest potencjalnie bardzo poważny, ponieważ może doprowadzić do zbudowania modelu, w którym zmienne są całkowicie niepowiązane ze sobą mimo pozornie istotnych statystyk t
- UWAGA: mimo, że statystyki t mają całkowicie nietypowe rozkłady w przypadku, kiedy zmienna zależna jest $I(1)$ i zmienne niezależne są $I(1)$, to jednak estymatory współczynników regresji są zgodne.

- Proste rozwiązanie problemu regresji pozornej: w przypadku zmiennych $I(1)$ możemy oszacować model na pierwszych różnicach - te będą $I(0)$ więc statystyki t będą miały standardowe rozkłady.
- Problem: za pomocą regresji na pierwszych różnicach nie da się oszacować równowagi długookresowej:

$$\begin{aligned}\Delta y_t &= \Delta \mathbf{x}_t \boldsymbol{\beta} + \varepsilon_t \\ E(y_t - y_{t-1}) &= E(\mathbf{x}_t - \mathbf{x}_{t-1}) \boldsymbol{\beta}\end{aligned}$$

i przy założeniu, że $y^* = E(y_t) = E(y_{t-1}) = \dots$ i $\mathbf{x}^* = E(\mathbf{x}_t) = E(\mathbf{x}_{t-1}) = \dots$ otrzymujemy mało mówiącą tożsamość $\mathbf{0} = \mathbf{0}\boldsymbol{\beta}$.

Twierdzenie Grangera i model korekty błędem (ECM)

- Praktyczne znaczenie pojęcia kointegracji związane jest z twierdzeniem Grangera.

Twierdzenie

*Jeśli (y_t, \mathbf{x}_t) są skointegrowane, oraz y_t i \mathbf{x}_t są $I(1)$, to y_t można przedstawić w postaci Mechanizmu Korekty Błędem (ECM – **Error Correction Mechanism**)*

$$\Delta y_t = \alpha (y_{t-1} - \mathbf{x}_{t-1} \beta) + \sum_{i=1}^{k-1} \theta_i \Delta y_{t-i} + \sum_{i=0}^{k-1} \Delta \mathbf{x}_{t-i} \gamma_i + \varepsilon_t$$

gdzie $y_{t-1} - \mathbf{x}_{t-1} \beta \sim I(0)$

- Znaczenie twierdzenia Grangera: umożliwia ono nam interpretacje wektora kointegrującego jako równowagi długookresowej:

$$\begin{aligned}y^* &= E(y_t) = \dots = E(y_{t-k}) \\ \implies E(\Delta y_t) &= \dots = E(\Delta y_{k-1}) = 0 \\ \mathbf{x}^* &= E(\mathbf{x}_t) = \dots = E(\mathbf{x}_{t-k}) \\ \implies E(\Delta \mathbf{x}_t) &= \dots = E(\Delta \mathbf{x}_{k-1}) = 0\end{aligned}$$

a więc z *ECM*:

$$0 = \alpha(y^* - \mathbf{x}^*\beta)$$

- Równowaga długookresowa dana wzorem $y^* = \mathbf{x}^*\beta$ to relacja między zmiennymi, do której zmienne dążą przy braku zakłóceń związanych z błędami losowymi

- $y_t - \mathbf{x}_t\beta$ jest odchyleniem od tej równowagi długookresowej.
- Odchylenie to jest korygowane za pośrednictwem $\alpha(y_{t-1} - \mathbf{x}_{t-1}\beta)$.
- Współczynnik α związany jest więc szybkością dostosowania y_t do poziomu równowagi.
- Współczynniki θ_i i γ_i związane są z krótkookresową dynamiką zmiennej zależnej.

- Do modelu powinniśmy wstawić taką liczbę opóźnionych różnic y_t i x_t , by wyeliminować autokorelację reszt. Przy ustalaniu k można posłużyć się metodą od ogólnego do szczegółowego bądź kryteriami informacyjnymi.
- Testowanie kointegracji ma sens jedynie wtedy, gdy y_t i *wszystkie* zmienne zawarte w x_t są $I(1)$.
- Pierwszym etapem analizy kointegracji musi więc być przetestowanie, czy wszystkie analizowane zmienne są $I(1)$

1 STOPIEŃ: Estymujemy równanie na poziomach

$$y_t = \mathbf{x}_t \beta + u_t$$

- otrzymujemy potencjalny wektor kointegrujący $1, -\hat{\beta}'$.
Ponieważ w regresji zmiennej $I(1)$ na $I(1)$ oszacowanie β jest zgodne więc reszty \hat{u}_t stanowią "dobre" oszacowanie błędów losowych u_t .
- Przeprowadzamy test *ADF* dla \hat{u}_t : jeśli $\hat{u}_t \sim I(0)$, to występuje kointegracja
- Rozkład statystyki testu *ADF* w przypadku testowania kointegracji zależy od liczby zmiennych objaśniających zawartych w wektorze \mathbf{x}_t .

2 STOPIEŃ: Szacujemy *ECM* wykorzystując otrzymane $\hat{\beta}$

$$\Delta y_t = \alpha \left(y_{t-1} - \mathbf{x}_{t-1} \hat{\beta} \right) + \sum_{i=1}^{k-1} \theta_i \Delta y_{t-i} + \sum_{i=0}^{k-1} \Delta \mathbf{x}_{t-i} \gamma_i + \varepsilon_t$$

- Tak uzyskane estymatory parametrów α i β są zgodne ale nieefektywne
 - Istnieją także inne metody szacowania współczynników w *ECM*

- Prognozy uzyskiwane z dużych strukturalnych modeli wielorównaniowych nie są lepsze niż prognozy uzyskiwane z *ARIMA*
- Teoria ekonomii opisuje warunki równowagi ale nie ma wiele do powiedzenia na temat dynamiki dostosowań
 - z tego powodu klasyczne dynamiczne modele strukturalne są całkowicie arbitralne
- Podział na zmienne endogeniczne i egzogeniczne jest całkowicie arbitralny - w rzeczywistości, z odpowiednim opóźnieniem, wszystko zależy od wszystkiego
- Sims zaproponował następujące rozwiązanie zasugerowanych problemów:
 - badania ilościowe powinny koncentrować się na właściwościach dynamicznych modelu zamiast na jego strukturze
 - szczególnie interesujące jest rozkład w czasie reakcji na szoki takie jak interwencje rządowe, zmiany regulacyjne, czy szoki egzogeniczne

VAR (Vector Autoregressive Model): forma strukturalna

- W strukturalnym modelu VAR, zmienna zależna x_{kt} zależy od równoczesnych wartości pozostałych elementów \mathbf{x}_t oraz od opóźnionych wartości wszystkich zmiennych w modelu
- Zapis modelu VAR:

$$\mathbf{A}\mathbf{x}_t = \mathbf{B}_1\mathbf{x}_{t-1} + \dots + \mathbf{B}_k\mathbf{x}_{t-k} + \boldsymbol{\Psi}\mathbf{D}_t + \mathbf{u}_t$$

$$\mathbf{u}_t \sim IID(0, \boldsymbol{\Sigma})$$

gdzie:

- \mathbf{x}_t jest $p \times 1$ wektorem zmiennych endogenicznych
- \mathbf{D}_t jest nielosowym $m \times 1$ wektorem elementów/trendów deterministycznych zawartych w modelu (stała, trend linowy, zero-jedynkowe zmienne sezonowe etc.)
- \mathbf{u}_t jest $p \times 1$ wektorem błędów losowych
- $\boldsymbol{\Sigma}$ jest $p \times p$ macierzą wariancji wektora błędów losowych \mathbf{u}_t

VAR (Vector Autoregressive Model): forma strukturalna

- Zmiennymi objaśniającymi w modelu są opóźnione zmienne zależne w modelu - w tym sensie model *VAR* jest wielowymiarowym uogólnieniem modelu skalarnego modelu *AR*.
- W strukturalnej formie *VAR* zależności natychmiastowe oraz opóźnione między zmiennymi są *explicite* ujawnione w modelu
- Model *VAR* w formie strukturalnej jest często trudny do zidentyfikowania, ponieważ trudno jest znaleźć w ekonomii teorię dynamicznych zależności między zmiennymi (w szczególności makroekonomicznymi).

VAR forma zredukowana (standardowa)

- Mnożąc strukturalną formę VAR obustronnie przez \mathbf{A}^{-1} uzyskujemy:

$$\mathbf{x}_t = \mathbf{A}^{-1} \mathbf{B}_1 \mathbf{x}_{t-1} + \dots + \mathbf{A}^{-1} \mathbf{B}_k \mathbf{x}_{t-k} + \mathbf{A}^{-1} \boldsymbol{\Psi} \mathbf{D}_t + \boldsymbol{\varepsilon}_t$$

- Zmieniając notację:

$$\mathbf{x}_t = \boldsymbol{\Pi}_1 \mathbf{x}_{t-1} + \dots + \boldsymbol{\Pi}_k \mathbf{x}_{t-k} + \boldsymbol{\Phi} \mathbf{D}_t + \boldsymbol{\varepsilon}_t,$$

$$\boldsymbol{\varepsilon}_t \sim IID(0, \boldsymbol{\Omega})$$

gdzie $\boldsymbol{\Pi}_i = \mathbf{A}^{-1} \mathbf{B}_i$, $\boldsymbol{\Phi} = \mathbf{A}^{-1} \boldsymbol{\Psi}$, $\boldsymbol{\varepsilon}_t = \mathbf{A}^{-1} \mathbf{u}_t$, $\boldsymbol{\Omega} = \mathbf{A}^{-1} \boldsymbol{\Sigma} \mathbf{A}'^{-1}$

- Zmienne endogeniczne w zredukowanej formie VAR zależą wyłącznie od opóźnionych wartości zmiennych objaśnianych
- Nie znaczy to jednak, że estymując VAR w formie zredukowanej wykluczamy możliwość występowania natychmiastowych zależności między zmiennymi - zazwyczaj takie założenie jest nierealistyczne
- W przypadku VAR w formie zredukowanej macierze współczynników Π_i nie mają interpretacji strukturalnej - mogą być interpretowane jedynie jako mnożniki a nie współczynniki w zależnościach behawioralnych

Estymacja modelu VAR

- Model VAR w formie zredukowanej może być efektywnie oszacowany przy użyciu *MNK* zastosowanego osobno do każdego z równań modelu pod warunkiem, że
 - nie występuje autokorelacja korelacja błędów losowych
- Jeżeli narzucone są ograniczenia na parametry modelu, to asymptotycznie bardziej efektywna niż *MNK* jest *MNW*.
 - zastosowanie *MNW* wymaga jednak założenia, że
$$\varepsilon_t \sim NID(\mathbf{0}, \Omega)$$
- Występowanie autokorelacji błędów losowych powoduje niezgodność standardowych estymatorów parametrów modeli VAR z powodu występowania endogeniczności
- Zwykle autokorelację można wyeliminować dodając do modelu wystarczającą liczbę opóźnień (poprzez wybranie odpowiedniego k)
- Występowanie autokorelacji może być przetestowane za pomocą testu Breuscha-Godfrey'a zastosowanego osobno do każdego z równań lub łącznie do wszystkich równań

- Modelowanie VAR wymaga dużej liczby obserwacji. Z reguły niemożliwe jest oszacowanie modeli VAR z dużą liczbą zmiennych i dużą liczbą opóźnień z racji na niewystarczającą liczbę stopni swobody.
- W przypadku estymowania modelu VAR ze zbyt małą liczbą stopni swobody łatwo może pojawić się problem nadmiernego dopasowania (*overfit*). Problem ten pojawia się, gdy liczba parametrów do oszacowania jest zbyt duża w stosunku do liczby obserwacji, taki model wydaje się dobrze dopasowany ale w rzeczywistości ma słabe własności statystyczne i prognostyczne
- Testowanie przyczynowości w sensie Grangera jest w przypadku modeli VAR szczególnie łatwe: testujemy łączną istotność wszystkich opóźnień zmiennej, która jest potencjalną przyczyną

Przykład: jedno z równań modelu VAR

- Przykład: VAR ze 4 zmiennymi, 4 opóźnieniami i stałą.
Pierwsze równanie modelu:

$$\begin{aligned}x_{1t} = & \pi_{1,11}x_{1,t-1} + \pi_{2,11}x_{1,t-2} + \pi_{3,11}x_{1,t-3} + \pi_{4,11}x_{1,t-4} \\ & + \pi_{1,12}x_{2,t-1} + \pi_{1,12}x_{2,t-2} + \pi_{3,12}x_{2,t-3} + \pi_{4,12}x_{2,t-4} \\ & + \pi_{1,13}x_{3,t-1} + \pi_{2,13}x_{3,t-2} + \pi_{3,13}x_{3,t-3} + \pi_{4,13}x_{3,t-4} \\ & + \pi_{1,14}x_{4,t-1} + \pi_{2,14}x_{4,t-2} + \pi_{3,14}x_{4,t-3} + \pi_{4,14}x_{4,t-4} \\ & + \phi_1 + \varepsilon_{1t}\end{aligned}$$

- W każdym równaniu mamy 17 parametrów. W całym modelu liczba parametrów do oszacowania: 68!
- Przetestowanie czy x_2 jest przyczyną w sensie Grangera x_2 przetestowania hipotezy łącznej
 $H_0 : \pi_{1,12} = \pi_{1,12} = \pi_{3,12} = \pi_{4,12} = 0.$

- W modelu VAR, \mathbf{x}_t zależy od ε_t i $\mathbf{x}_{t-1}, \mathbf{x}_{t-2}, \dots$

$$\mathbf{x}_t = \Pi_1 \mathbf{x}_{t-1} + \dots + \Pi_k \mathbf{x}_{t-k} + (\Phi \mathbf{D}_t + \varepsilon_t)$$

- Taki model może być zapisany za pomocą wektorowego wielomianu operatora opóźnień:

$$\mathbf{A}(L) \mathbf{x}_t = \varepsilon_t$$

gdzie $\mathbf{A}(L) = I - \Pi_1 L - \dots - \Pi_k L^k$

- Jeśli $\mathbf{A}(L)$ można odwrócić to \mathbf{x}_t można wyrazić jako funkcję szoków i elementów deterministycznych (model w formie MA)

$$\mathbf{x}_t = \mathbf{A}^{-1}(L) (\Phi \mathbf{D}_t + \varepsilon_t) = \sum_{i=0}^{\infty} \Psi_i L^i (\varepsilon_t + \Phi \mathbf{D}_t)$$

Warunki odwracalności wielomianu $\mathbf{A}(L)$

Warunek odwracalności wielomianu $\mathbf{A}(L)$:

Wszystkie pierwiastki wielomianu $\det(\mathbf{A}(\mu)) = 0$ są *poza* kołem jednostkowym

- Jeśli $\mathbf{A}(L)$ jest odwracalny to wpływ szoku ε_t maleje wraz z upływem czasu
- Jeśli $\mathbf{A}(L)$ nie jest odwracalny, to wpływ szoku nie maleje wraz z czasem.
- Warunek odwracalności $\mathbf{A}(L)$ jest więc dostatecznym warunkiem stabilności modelu *VAR*.
- Warunki stabilności można zweryfikować badając wartości własne macierzy zwanej macierzą towarzyszącą.
- Jeśli wszystkie wartości własne macierzy towarzyszącej leżą w kole jednostkowym (mają moduły *mniejsze* niż 1), to rozwiązania wielomianu $\mathbf{A}(\mu)$ leżą *poza* kołem jednostkowym i model *VAR* jest stabilny

- Zauważmy, że model

$$\begin{bmatrix} \mathbf{y}_t \\ \mathbf{y}_{t-1} \\ \vdots \\ \mathbf{y}_{t-k+1} \end{bmatrix} = \begin{bmatrix} \mathbf{\Pi}_1 & \mathbf{\Pi}_2 & \cdots & \mathbf{\Pi}_k \\ \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} \\ & \ddots & \ddots & \vdots \\ \mathbf{0} & & \mathbf{I} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{y}_{t-1} \\ \mathbf{y}_{t-2} \\ \vdots \\ \mathbf{y}_{t-k} \end{bmatrix} + \Phi \mathbf{D}_t + \varepsilon_t$$

można zapisać jako

$$\mathbf{Y}_t = \mathbb{A} \mathbf{Y}_{t-1} + \mathbf{1} \Phi \mathbf{D}_t + \mathbf{1} \varepsilon_t$$

gdzie $\mathbf{Y}_t = [\mathbf{y}'_t, \mathbf{y}'_{t-1}, \dots, \mathbf{y}'_{t-k+1}]'$, $\mathbf{1} = [\mathbf{I}, \mathbf{0}, \dots, \mathbf{0}]'$.

- Wykorzystując rekurencyjną substytucję i oraz to dekompozycję SVD $\mathbb{A} = \mathbf{Q}\mathbf{\Lambda}^i\mathbf{Q}^{-1}$ otrzymujemy ($\mathbf{\Lambda}$ diagonalna):

$$\mathbf{Y}_t = \sum_{i=0}^{\infty} \mathbb{A}^i (\mathbf{1}\Phi\mathbf{D}_t + \mathbf{1}\varepsilon_t) = \sum_{i=0}^{\infty} \mathbf{Q}\mathbf{\Lambda}^i\mathbf{Q}^{-1} (\mathbf{1}\Phi\mathbf{D}_t + \mathbf{1}\varepsilon_t)$$

- Zauważmy, że współczynniki w reprezentacji *MA* modelu *VAR* są równe $\Psi_i = \mathbf{1}'\mathbf{Q}\mathbf{\Lambda}^i\mathbf{Q}^{-1}\mathbf{1}$
- Warunkiem stabilności jest więc, by wszystkie elementy diagonalne $\mathbf{\Lambda}$ (wartości własne macierz towarzyszącej \mathbb{A}) miały moduły mniejsze od 1

$$\det(\mathbb{A} - \mathbf{I}\lambda) = 0$$

- Pierwiastki macierzy towarzyszącej \mathbb{A} są odwrotnościami pierwiastków równania $\mathbf{A}(\mu) = \mathbf{0}$.

- Wartość oczekiwane \mathbf{x}_t dla stabilnego VAR

$$E(\mathbf{x}_t) = E\left(\sum_{i=0}^{\infty} \boldsymbol{\Psi}_i L^i (\boldsymbol{\Phi} \mathbf{D}_{t-i} + \boldsymbol{\varepsilon}_{t-i})\right) = \sum_{i=0}^{\infty} \boldsymbol{\Psi}_i \boldsymbol{\Phi} \mathbf{D}_{t-i}$$

- Jeżeli jedynym elementem deterministycznym jest stała, to $\boldsymbol{\Phi} \mathbf{D}_{t-i} = \boldsymbol{\mu}$

$$E(\mathbf{x}_t) = \sum_{i=0}^{\infty} \boldsymbol{\Psi}_i \boldsymbol{\mu} = \mathbf{A}^{-1}(\mathbf{1}) \boldsymbol{\mu} = (\mathbf{I} - \boldsymbol{\Pi}_1 - \dots - \boldsymbol{\Pi}_k)^{-1} \boldsymbol{\mu}$$

- Równowaga długookresowo może być w tym przypadku interpretowana jako wielkość, wokół której oscylują \mathbf{x}_t na skutek wpływu zaburzeń losowych $\boldsymbol{\varepsilon}_t$

- Prognozowanie za pomocą VAR
 - 1 prognozy jednookresowe

$$\hat{\mathbf{x}}_{T+1} = \Pi_1 \mathbf{x}_T + \dots + \Pi_k \mathbf{x}_{T-k+1}$$

- 2 prognozy dwuokresowe

$$\hat{\mathbf{x}}_{T+2} = \Pi_1 \hat{\mathbf{x}}_{T+1} + \dots + \Pi_k \mathbf{x}_{T-k+2}$$

⋮

- s prognozy s -okresowe ($s > k + 1$)

$$\hat{\mathbf{x}}_{T+s+1} = \Pi_1 \hat{\mathbf{x}}_{T+s} + \dots + \Pi_k \hat{\mathbf{x}}_{T-k+s}$$

- Ponieważ $E(\hat{\mathbf{x}}_{T+1} - \mathbf{x}_{T+1}) = 0$, $E(\hat{\mathbf{x}}_{T+2} - \mathbf{x}_{T+2}) = 0$, ... więc takie prognozy są nieobciążone jeśli Π_i jest znane
- Dla $s \rightarrow \infty$ prognozy dążą do równowagi długookresowej.
- Jeśli Π_i jest oszacowana to prognozy są asymptotycznie nieobciążone

- Funkcje reakcji służą mierzeniu reakcji \mathbf{x}_t na szok ε_{t-i}
- Wiemy już, że

$$\mathbf{x}_t = \sum_{i=0}^{\infty} \boldsymbol{\Psi}_i L^i (\varepsilon_t + \boldsymbol{\Phi} \mathbf{D}_t)$$

- Wpływ na l -ty element \mathbf{x}_t jednostkowego szoku w k -tym elemencie ε_{t-i} , jest dany przez $\{\boldsymbol{\Psi}_i\}_{l,k}$.
- **Krytyka:** jaki jest sens analizy reakcji na szok jednostkowy dotyczący jednego elementu ε_t w sytuacji, kiedy elementy ε_t są ze sobą skorelowane ($\boldsymbol{\Omega}$ jest niediagonalna)?
 - jeśli szoki są ze sobą skorelowane, to szok dotyczący jednego elementu ε_t współwystępuje z szokiem w innych elementach ε_t - w takich przypadkach analiza szoku dotyczącego tylko jednego elementu ε_t wydaje się nie mieć sensu - taki scenariusz jest nierealistyczny

- Mnożąc standardową (zredukowaną) formę VAR przez pewną nieosobliwą macierz \mathbf{C} otrzymujemy:

$$\begin{aligned}\mathbf{C}\mathbf{x}_t &= \mathbf{C}\mathbf{\Pi}_1\mathbf{x}_{t-1} + \dots + \mathbf{C}\mathbf{\Pi}_k\mathbf{x}_{t-k} + \mathbf{C}\mathbf{\Phi}\mathbf{D}_t + \mathbf{C}\boldsymbol{\varepsilon}_t \\ &= \mathbf{\Gamma}_1\mathbf{x}_{t-1} + \dots + \mathbf{\Gamma}_k\mathbf{x}_{t-k} + \mathbf{\Upsilon}\mathbf{D}_t + \boldsymbol{\eta}_t\end{aligned}\quad (1)$$

gdzie $\mathbf{\Gamma}_i = \mathbf{C}\mathbf{\Pi}_i$, $\mathbf{\Upsilon} = \mathbf{C}\mathbf{\Phi}$, $\boldsymbol{\eta}_t = \mathbf{C}\boldsymbol{\varepsilon}_t \sim IID(0, \mathbf{C}\boldsymbol{\Omega}\mathbf{C}')$.

- Za macierz \mathbf{C} wybiera się zazwyczaj macierz Choleskiego:
 - macierz \mathbf{C} jest dolnotrójkątna z jedynkami na przekątnej
 - $\mathbf{C}\boldsymbol{\Omega}\mathbf{C}' = \boldsymbol{\Sigma}_D$
 - macierz $\boldsymbol{\Sigma}_D$ jest diagonalna

- W rezultacie

$$\boldsymbol{\eta}_t = \mathbf{C}\boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_D)$$

- **Wniosek:** transformacja modelu przy użyciu macierzy Choleskiego ortogonalizuje szoki - szoki są nieskorelowane, realistyczny jest szok wpływający na jedno równanie

- Model używany do uzyskania szoków zortogonalizowanych jest *de facto* jest modelem strukturalnym a restrykcje narzucone na elementy \mathbf{C} (jej dolnotrójkątność) identyfikują parametry modelu a pośrednio też szoki
- Definiujące funkcje reakcji macierz współczynników w reprezentacji MA można znaleźć na bazie relacji między macierzami współczynników oraz błędami w formie standardowej i przekształconej

$$\mathbf{x}_t = \sum_{i=0}^{\infty} \boldsymbol{\Psi}_i L^i (\boldsymbol{\Phi} \mathbf{D}_t + \boldsymbol{\varepsilon}_t) = \sum_{i=0}^{\infty} \boldsymbol{\Psi}_i \mathbf{C}^{-1} L^i (\boldsymbol{\Upsilon} \mathbf{D}_t + \boldsymbol{\eta}_t)$$

- Zortogonalizowane funkcje reakcji na szoki są prostymi liniowymi przekształceniami funkcji reakcji na szoki jednostkowe: $\mathbf{C}^{-1} \boldsymbol{\Psi}_i$

- Model strukturalny (1), ma następujące równania:

$$x_{1t} = \sum_{i=1}^k \gamma_{1i} x_{t-i} + \tau_i D_t + \eta_{1t}$$

$$x_{2t} = -c_{21} x_{1t} + \sum_{i=1}^k \gamma_{2i} x_{t-i} + \tau_i D_t + \eta_{2t}$$

...

$$x_{Gt} = - \sum_{i=1}^{G-1} c_{Gi} x_{it} + \sum_{i=1}^k \gamma_{Gi} x_{t-i} + \tau_i D_t + \eta_{Gt}$$

- Cechy takiego modelu strukturalnego (zależności równoczesne)
- szoki
 - x_{1t} zależy od η_{1t}
 - x_{2t} zależy od η_{1t}, η_{2t}
 - \vdots
 - x_{Gt} zależy od $\eta_{1t}, \dots, \eta_{Gt}$
- zmienne x_{it}
 - x_{1t} nie zależy od równoczesnych x_{it} ($j \neq i$)
 - x_{2t} zależy od x_{1t}
 - x_{3t} zależy od x_{1t}, x_{2t}
 - \vdots
 - x_{Gt} zależy od x_{1t}, \dots, x_{G-1t}

- Bardzo szczególna struktura przyczynowa - czy można ją uzasadnić na bazie teorii ekonomii?
- Możliwe rozwiązanie: analiza wrażliwości (sprawdzamy, czy kolejność zmiennych ma istotny wpływ na kształt funkcji reakcji)
- **Krytyka:** arbitralne restrykcje identyfikujących parametry modelu zostały zastąpione arbitralnymi restrykcjami identyfikującymi szoki

Wektorowy mechanizm korekty błędem (VECM)

- Rozważmy następujący model VAR

$$\mathbf{x}_t = \mathbf{\Pi}_1 \mathbf{x}_{t-1} + \dots + \mathbf{\Pi}_2 \mathbf{x}_{t-2} + \varepsilon_t$$

- Odejmując \mathbf{x}_{t-1} od obu stron i przeprowadzając dalsze przekształcenia uzyskujemy:

$$\begin{aligned}\Delta \mathbf{x}_t &= (\mathbf{\Pi}_1 - I) \mathbf{x}_{t-1} + \mathbf{\Pi}_2 \mathbf{x}_{t-2} + \varepsilon_t \\ &= (\mathbf{\Pi}_1 - I) \mathbf{x}_{t-1} + \mathbf{\Pi}_2 \mathbf{x}_{t-1} - \mathbf{\Pi}_2 \mathbf{x}_{t-1} + \mathbf{\Pi}_2 \mathbf{x}_{t-2} + \varepsilon_t \\ &= (\mathbf{\Pi}_1 + \mathbf{\Pi}_2 - I) \mathbf{x}_{t-1} - \mathbf{\Pi}_2 \Delta \mathbf{x}_{t-1} + \varepsilon_t \\ &= \mathbf{\Pi} \mathbf{x}_{t-1} + \mathbf{\Gamma}_1 \Delta \mathbf{x}_{t-1} + \varepsilon_t\end{aligned}$$

gdzie $\mathbf{\Pi} = \mathbf{\Pi}_1 + \mathbf{\Pi}_2 - I$ a $\mathbf{\Gamma}_1 = -\mathbf{\Pi}_2$

- W podobny sposób możemy przekształcić model VAR w równoważną formę typu $VECM$:

$$\Delta \mathbf{x}_t = \mathbf{\Pi} \mathbf{x}_{t-1} + \sum_{i=1}^{k-1} \mathbf{\Gamma}_i \Delta \mathbf{x}_{t-i} + \mathbf{\Phi} \mathbf{D}_t + \varepsilon_t, \quad t = 1 \dots T \quad (2)$$

gdzie $\mathbf{\Pi} = \sum_{i=1}^k \mathbf{\Pi}_i - I$ a $\mathbf{\Gamma}_i = -\sum_{j=i+1}^k \mathbf{\Pi}_j$

Wektorowy mechanizm korekty błędem (VECM)

- Kombinacja liniowa zmiennej niestacjonarnej i stacjonarnej jest niestacjonarna.
- W szczególności jeśli $\mathbf{x}_{1t} \sim I(1)$ i $\mathbf{x}_{2t} \sim I(0)$, to $\mathbf{x}_{1t} + \Theta \mathbf{x}_{2t} \sim I(1)$
- Załóżmy, że $\mathbf{x}_t \sim I(1)$ i przeanalizujemy stopnie integracji elementów modelu w formie *VECM*:

$$\underbrace{\Delta \mathbf{x}_t}_{I(0)} = \underbrace{\Pi \mathbf{x}_{t-1}}_{I(1)} + \sum_{i=1}^{k-1} \underbrace{\Gamma_i \Delta \mathbf{x}_{t-i}}_{I(0)} + \underbrace{\Phi \mathbf{D}_t + \varepsilon_t}_{I(0)}$$

- Stopień integracji lewej i prawej strony może być równa tylko wtedy, gdy $\Pi \mathbf{x}_{t-1} \sim I(0)$.

Wniosek

Wiersze macierzy Π są wektorami kointegrującymi

- Liczba liniowo niezależnych wierszy macierzy Π determinuje liczbę relacji kointegrujących między zmiennymi

- Wnioski z poprzedniego slajdu można sformułować w formie twierdzenia Grangera

Grangera (uproszczone)

Jeśli $\mathbf{x}_t \sim I(1)$, i istnieją macierze α i β o wymiarach $p \times r$ i $p \times r$ oraz rzędu r , takie, że w równaniu (2)

$$\Pi = \alpha\beta'$$

to \mathbf{x}_t jest skointegrowane i $\beta' \mathbf{x}_t$ jest $I(0)$

- Kolumny macierzy β są wektorami kointegrującymi
- Jeśli $r = p$, to macierz Π jest nieosobliwa i $\mathbf{x}_t \sim I(0)$

- Z twierdzenia Grangera wynika, że skointegrowany VAR można zapisać jako:

$$\Delta \mathbf{x}_t = \alpha \beta' \mathbf{x}_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta \mathbf{x}_{t-i} + \Phi \mathbf{D}_t + \varepsilon_t, \quad t = 1 \dots T$$

- Wyrażenie $\beta' \mathbf{x}_t$ można interpretować jako odchylenie od równowagi długookresowej (błąd)
- Macierz α zawiera współczynniki dostosowań
- Macierze Γ_i zawierają współczynniki wpływające na dynamikę krótkookresową
- Sam model jest wielorównaniowym uogólnieniem modelu *ECM* (Wektorowy Mechanizm Korekty Błędem *VECM*)

- Liczbę relacji kointegrujących ustalamy wykorzystując testy rzędu macierzy kointegrującej
- Najbardziej popularnym testem tego typu jest test Johansena (statystyka śladu - trace statistic)
- Test ten jest testem bazującym na statystyce LR dla $H_0 : r = s$ przy alternatywie $H_1 : r > s$ postaci:

$$\begin{aligned} LR_T(H_0) &= 2 \ln \left[\frac{L_{\max}(H(p))}{L_{\max}(H(r))} \right] \\ &= -T \left[\sum_{i=s+1}^p \ln(1 - \hat{\lambda}_i) \right] \end{aligned}$$

gdzie $1 > \lambda_1 > \lambda_2 > \dots > \lambda_p > 0$ i λ_i są wartościami własnymi pewnej macierzy

- Testowanie rzędu macierzy kointegrującej ma charakter sekwencyjny:
- 1 Zaczynamy od hipotezy $H_0 : r = 0$ przy $H_1 : r > 0$
 - 1 jeśli $H_0 : r = 0$ została odrzucona testujemy $H_0 : r = 1$ przy $H_1 : r > 1$
 - ⋮
 - 2 zatrzymujemy się, gdy $H_0 : r = s$ nie została odrzucona

Test Johansena - model trendów

Rozkład testu Johansena zależy od przyjętego modelu trendów.
Najczęściej rozważa się następujące przypadki:

Poziomy \mathbf{x}_t	Model korekty błędem
ograniczona stała	$\Delta \mathbf{x}_t = \alpha \beta' \mathbf{x}_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta \mathbf{x}_{t-i} + \varepsilon_t$
stała	$\Delta \mathbf{x}_t = \alpha (\beta' \mathbf{x}_{t-1} + \mu) + \sum_{i=1}^{k-1} \Gamma_i \Delta \mathbf{x}_{t-i} + \varepsilon_t$
stała+ograniczony trend	$\Delta \mathbf{x}_t = \alpha \beta' \mathbf{x}_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta \mathbf{x}_{t-i} + \mu + \varepsilon_t$
trend	$\Delta \mathbf{x}_t = \alpha (\beta' \mathbf{x}_{t-1} + \xi t) + \sum_{i=1}^{k-1} \Gamma_i \Delta \mathbf{x}_{t-i} + \varepsilon_t$
ograniczony trend kwadratowy	$\Delta \mathbf{x}_t = \alpha \beta' \mathbf{x}_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta \mathbf{x}_{t-i} + \mu + \xi t + \varepsilon_t$

- Przed rozpoczęciem testowania należy ustalić właściwy model dla trendów

- W modelu *VECM*, Proces Generujący Dane (*DGP*) ma następującą formę

$$\Delta \mathbf{x}_t = \mathbf{\Pi} \mathbf{x}_{t-1} + \sum_{i=1}^{k-1} \mathbf{\Gamma}_i \Delta \mathbf{x}_{t-i} + \mathbf{\Phi} \mathbf{D}_t + \varepsilon_t,$$

- Jeśli występuje kointegracja, to $\mathbf{\Pi} = \alpha \beta'$
- **Jednak:** ten DGP nie zmieni się jeśli α i β wymienimy na macierze $\alpha^* = \alpha \mathbf{A}^{-1'}$ i $\beta^* = \beta \mathbf{A}$ dla dowolnej $r \times r$ i nieosobliwej macierzy \mathbf{A} , ponieważ $\mathbf{\Pi} = \alpha \beta' = \alpha^* \beta^*$
- **Wniosek:** macierze α i β nie są zidentyfikowane.

Identyfikacja macierzy kointegrującej β

- Celem zidentyfikowania macierzy α i β konieczne jest narzucenie na te macierzy liczby ograniczeń równej liczbie elementów \mathbf{A} . Potrzebujemy więc r^2 ograniczeń.
- Pakiety statystyczne często raportują β nawet jeśli nie jest ona zidentyfikowana. Oszacowania te uzyskiwane są na bazie arbitralnych, technicznych restrykcji n.p.

$$\beta = \begin{bmatrix} \mathbf{I} \\ \beta^* \end{bmatrix}$$

- Tego typu założenia nie mają zwykle interpretacji strukturalnej.
- Wektory kointegrujące mają interpretację strukturalną tylko wtedy, gdy ich identyfikacja opiera się na restrykcjach zakorzenionych w teorii ekonomii.
- **Uwaga:** zauważmy jednak, że liczbę wektorów kointegrujących (rzęd macierzy kointegrującej) można przetestować bez identyfikowania wektorów kointegrujących.

Identyfikacja macierzy kointegrującej β

- Identyfikacja opiera się zazwyczaj na układzie liniowych ograniczeń narzuconych na wektory β_i bazujących na teorii ekonomii:

$$R'_i \beta_i = 0, \quad i = 1, \dots, r^2$$

- Ograniczenia identyfikujące parametry nie mogą są nietestowalne ale w przypadku przeidentyfikowania modelu można ograniczenia można przetestować.
- Możliwa jest także identyfikacja macierzy parametrów α i β wykorzystująca ograniczenia narzucone na α .
- Zmienne egzogeniczne: możliwe jest umieszczenie w modelu *VECM* zmiennych egzogenicznych (*VECMX*).
- Zmienne egzogeniczne (jeśli są $I(1)$) mogą również pojawiać się w zależnościach kointegrujących. W takim przypadku należy jednak użyć specjalnych tablic przy testowaniu liczby wektorów kointegrujących.