

Wprowadzenie

ASC 2023

Piotr Żoch

- Sprawy organizacyjne
- Szereg czasowy: sezonowość, stacjonarność, integracja.

Sprawy organizacyjne

- **Email:** p.zoch@uw.edu.pl.
- **Dyżur:** uzgadniany indywidualnie.
- **Nieobecności:** max. 3 nieusprawiedliwione.
- **Forma zajęć:** konwersatorium
- **Slajdy i inne materiały:** Moodle/Github
- **Zaliczenie:** projekt zaliczeniowy (50%) + egzamin (50%)

- Dokładne wytyczne i przykładowe projekty:
<http://www.ekonometria.wne.uw.edu.pl>
- Indywidualna analiza dwóch szeregów czasowych (dane miesięczne/kwartalne/roczne)
 - Szereg sezonowy.
 - Szereg niesezonowy.
- Dekompozycja szeregu, dobranie modelu ARIMA/SARIMA, prognozy z modelu.
- Termin: **18 czerwca, 23:59.**
- Wybrane szeregi należy ze mną skonsultować do **15 kwietnia.**
- Kto pierwszy, ten lepszy!

- Z USOSa:
 - Enders W. (2014), *Applied Econometric Time Series*, Wiley.
 - Pawełek B., Wanat S., Zeliaś A. (2003/2013), *Prognozowanie ekonomiczne – Teoria, przykłady, zadania*, PWN.
- Klasyk:
 - Hamilton, J. (1994), *Time Series Analysis*, Princeton University Press.

Szeregi czasowe

- **Dane przekrojowe** (*cross sectional data*) - wiele jednostek $n = 1, 2, \dots, N$ obserwowanych w jednym okresie
- **Szeregi czasowe** (*time series data*) - jedna jednostka obserwowana w wielu okresach $t = 1, 2, \dots, T$
- **Dane panelowe** (*panel data*) - wiele jednostek $n = 1, 2, \dots, N$ obserwowanych w wielu okresach $t = 1, 2, \dots, T$

- **Proces stochastyczny:** zbiór zmiennych losowych $\{Y_t\}_{t \in T}$, które przyjmują wartości w przestrzeni probabilistycznej (Ω, \mathcal{F}, P) i są indeksowane przez zbiór T .
 - Przykład: $Y_t \sim N(0, \sigma^2)$
 - Przykład: $Y_t \sim N(t, \sigma_t^2)$
- **Szereg czasowy:** pojedyncza realizacja pewnego procesu stochastycznego.

$$\{y_t\}_{t \in T}$$

- Ciąg obserwacji pokazujący kształtowanie się badanego zjawiska w kolejnych okresach (dniach, miesiącach, kwartałach, latach, itp.)

```
../../ASC/week1/plot1.png
```

- Po co?
 - Prognozowanie przyszłych wartości
 - Badanie dynamiki szeregów
 - Dodatkowo: pewne właściwości szeregów czasowych ważne nawet przy zwykłej regresji liniowej
- Co robić?
 - Analiza czasowa (ten kurs)
 - Analiza częstotliwościowa (może o tym wspomnimy)

- **Sezonowość:** zmienność związana z cyklem kalendarzowym:
 - Przykład: dla zmiennych miesięcznych — sezonowość miesięczna.



`../../ASC/week1/plot2.png`

- Jeśli sezonowość ma wpływ na związek między zmienną objaśniającą a objaśnianą i nie zostanie uwzględniona w modelu, to pojawi się w resztach (nie będą spełniały założeń klasycznego modelu regresji liniowej).
- Co robić?
 - Pracować z danymi wyrównanymi sezonowo (różne metody, które poznamy na tym kursie).
 - Potraktować sezonowość poważnie i uwzględnić ją w modelu.

```
../../ASC/week1/plot4.png
```

- **Różnicowanie sezonowe:** prosty sposób na pozbycie się sezonowości — zamiast pierwotnych wartości zmiennych możemy wykorzystać

$$\Delta_s y_t := y_t - y_{t-s}$$

gdzie $s = 4$ dla zmiennych kwartalnych, $s = 12$ dla zmiennych miesięcznych.



../../ASC/week1/plot5.png

- Zmienna jest **stacjonarna** w sensie **ściśłym** jeśli łączne rozkłady $\{Y_t, Y_{t+1}, \dots, Y_T\}$ i $\{Y_{t+\tau}, Y_{t+1+\tau}, \dots, Y_{T+\tau}\}$ są takie same dla wszystkich $T \geq t \geq 1, \tau \geq 0$.
- Inaczej: $F(\{Y_t, Y_{t+1}, \dots, Y_T\}) = F(\{Y_{t+\tau}, Y_{t+1+\tau}, \dots, Y_{T+\tau}\})$
- Rozkład nie zależy od czasu.

- Zmienna jest **stacjonarna** w sensie **słabym** (*stacjonarność kowariancyjna*) jeśli spełnione są trzy warunki:
 - $E[Y_t] = \mu < \infty$
 - $Var[Y_t] = \sigma^2 < \infty$
 - $Cov(Y_t, Y_{t+h}) = Cov(Y_s, Y_{s+h}) = \gamma_h$
- Intuicyjnie: właściwości zmiennej nie zmieniają się w czasie.
- Jeśli którykolwiek warunek nie jest spełniony, zmienna jest **niestacjonarna**.
- W praktyce będziemy badać stacjonarność w sensie słabym.

- Przykład: biały szum (*white noise*): Y_t jest i.i.d. z wartością oczekiwaną równą 0 i wariancją σ^2 .
 - *independent and identically distributed*
- Stacjonarność
 - $E[Y_t] = 0 < \infty$
 - $\text{Var}[Y_t] = \sigma^2 < \infty$
 - $\text{Cov}(Y_t, Y_s) = 0$ dla $t \neq s$.
- Tak.

- Przykład: model AR(1)

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

$$\varepsilon_t \sim iid(0, \sigma_\varepsilon^2)$$

$$|\rho| < 1$$

- Stacjonarność? Sprawdźmy trzy warunki (słabej) stacjonarności.

- Warunek pierwszy: $E[y_t] = \mu < \infty$.
- Zauważmy, że

$$\begin{aligned} E[y_t] &= E \left[\sum_{i=0}^{\infty} \rho^i \varepsilon_{t-i} \right] \\ &= \sum_{i=0}^{\infty} \rho^i E[\varepsilon_{t-i}] \\ &= 0 < \infty \end{aligned}$$

- Warunek drugi: $\text{Var}[y_t] = \sigma^2 < \infty$.
- Zauważmy, że

$$\begin{aligned}\text{Var}[y_t] &= \text{Var}\left[\sum_{i=0}^{\infty} \rho^i \varepsilon_{t-i}\right] \\ &= \sum_{i=0}^{\infty} \rho^{2i} \text{Var}[\varepsilon_{t-i}] \\ &= \frac{\sigma_{\varepsilon}^2}{1 - \rho^2} < \infty\end{aligned}$$

- Warunek drugi: $\text{Cov}(Y_t, Y_{t+h}) = \text{Cov}(Y_s, Y_{s+h}) = \gamma_h$.

$$\begin{aligned}\text{Cov}[y_t, y_{t-h}] &= \text{Cov}\left[\sum_{i=0}^{\infty} \rho^i \varepsilon_{t-i}, \sum_{i=0}^{\infty} \rho^i \varepsilon_{t-h-i}\right] \\&= \text{Cov}\left[\sum_{i=0}^{h-1} \rho^i \varepsilon_{t-i} + \rho^h \sum_{i=0}^{\infty} \rho^i \varepsilon_{t-i-h}, \sum_{i=0}^{\infty} \rho^i \varepsilon_{t-h-i}\right] \\&= \rho^h \sum_{i=0}^{\infty} \rho^{2i} \text{Var}[\varepsilon_{t-i-h}] \\&= \rho^h \frac{\sigma_{\varepsilon}^2}{1 - \rho^2}\end{aligned}$$

- Wniosek?

- Przykład: błędzenie losowe (*random walk*)

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

$$\varepsilon_t \sim iid(0, \sigma_\varepsilon^2)$$

- Podobne do AR(1), ale z $\rho = 1$.
- Stacjonarność?

- Zauważmy, że:

$$E[y_t] = 0$$

$$\begin{aligned} \text{Var}[y_t] &= \sum_{s=1}^t \text{Var}[\varepsilon_s] \\ &= t\sigma_\varepsilon^2 \end{aligned}$$

$$\begin{aligned} \text{Cov}[y_t, y_{t-h}] &= \sum_{s=1}^{t-h} \text{Var}[\varepsilon_s] \\ &= (t-h)\sigma_\varepsilon^2 \end{aligned}$$

- Stacjonarność?

- Przykład: błędzenie losowe z dryfem

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

$$\varepsilon_t \sim iid(0, \sigma_\varepsilon^2)$$

- Stacjonarność?

- Zmienna jest **stacjonarna wokół trendu** (*trendostacjonarna*) jeśli odchylenie od trendu

$$y_t - E[y_t]$$

jest stacjonarne

- Przykład: trend liniowy

$$y_t = \alpha + \beta t + \varepsilon_t$$

$$E[y_t] = \alpha + \beta t$$

$$y_t - E[y_t] = \varepsilon_t$$

Zmienna zintegrowana

- Zmienna jest **zintegrowana**, jeśli jest zmienną niestacjonarną, której d -te różnice są stacjonarne.
 - d -te różnice:

$$\Delta_1 y_t := y_t - y_{t-1}$$

$$\Delta_2 y_t := \Delta_1 (\Delta_1 y_t)$$

$$\Delta_d y_t := \Delta_1 (\Delta_{d-1} y_t)$$

- Mówimy, że zmienna y_t jest zintegrowana stopnia d .

$$y_t \sim I(d)$$

- Zmienne stacjonarne są zintegrowane stopnia 0.

$$y_t \sim I(0)$$

- Przykład: błędzenie losowe (*random walk*)

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

$$\varepsilon_t \sim iid(0, \sigma_\varepsilon^2)$$

- Wiemy już, że y_t jest zmienną niestacjonarną.
- Czy pierwsze różnice są stacjonarne?

- W tym przypadku mamy

$$\begin{aligned}\Delta_1 y_t &= y_t - y_{t-1} \\ &= \varepsilon_t\end{aligned}$$

czyli $\Delta_1 y_t$ to biały szum, zmienna stacjonarna.

- Konkluzja: $y_t \sim I(1)$.