Sezonowość, zmienne stacjonarne i zmienne zintegrowane

Stanisław Cichocki
Natalia Nehrebecka

Wykład 2

Plan wykładu

- 1. Szereg czasowy
- 2. Sezonowość
- 3. Zmienne stacjonarne
- 4. Zmienne zintegrowane
- 5. Regresja pozorna

Plan wykładu

- 1. Szereg czasowy
- 2. Sezonowość
- 3. Zmienne stacjonarne
- 4. Zmienne zintegrowane
- 5. Regresja pozorna

Szereg czasowy

- Szereg czasowy jest pojedynczą realizacją pewnego procesu stochastycznego.

Szereg czasowy

Kurs USD/PLN	Data
15.02.2016	3,922
16.02.2016	3,9404
17.02.2016	3,9431
18.02.2016	3,9416
19.02.2016	3,9531

Szereg czasowy

http://www.bankier.pl/inwestowanie/profile/quote.html?symbol=ROPA

Plan wykładu

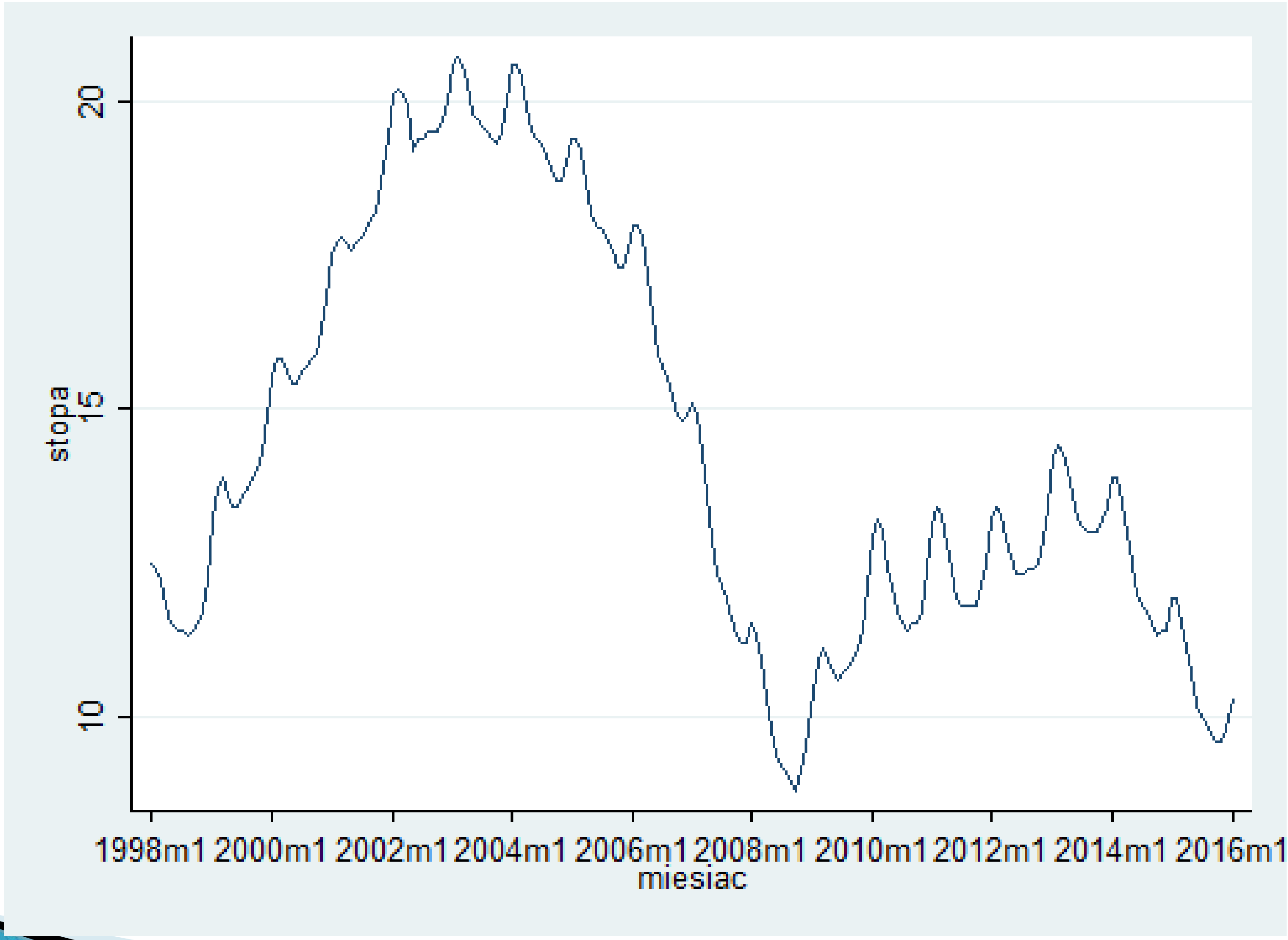
- 1. Szereg czasowy
- 2. Sezonowość
- 3. Zmienne stacjonarne
- 4. Zmienne zintegrowane
- 5. Regresja pozorna

Sezonowość

- O sezonowości mówimy wtedy gdy zmienna zmienia się w pewnym cyklu, zwykle związanym z cyklem kalendarzowym.
  - Np. zmienne kwartalne charakteryzują się sezonowością kwartalną a zmienne miesięczne charakteryzują się sezonowością miesięczną

- Sezonowość w danych może pojawiać się z rożnych powodów:
  - czynniki klimatyczne (spadek wartości dodanej w budownictwie w okresie zimowym);
  - czynniki kulturowe (wzrost wartości sprzedaży w okresie świąt).

Sezonowość



stopa
20
15
10
1998m1
2000m1
2002m1
2004m1
2006m1
2008m1
2010m1
2012m1
2014m1
2016m1
miesiac

Sezonowość

- Sezonowości należy uwzględnić w modelu jeśli ma ona wpływ na związek między zmienną objaśniającą a objaśnianą:
  - jeśli w modelu nie zostanie uwzględniona sezonowość to pojawi się ona w resztach, które nie będą spełniały założeń KMRL.

Sezonowość

- Uwzględnienie problemu sezonowości w procesie estymacji:
  a) posłużenie się danymi wyrównanymi sezonowo (publikowane przez urzędy statystyczne; samodzielnie można usunąć sezonowość z danych np. korzystając z programu TRAMO/SEATS);

  b) dodanie do modelu zmiennych zerojedynkowych związanych z poszczególnymi miesiącami/kwartałami;

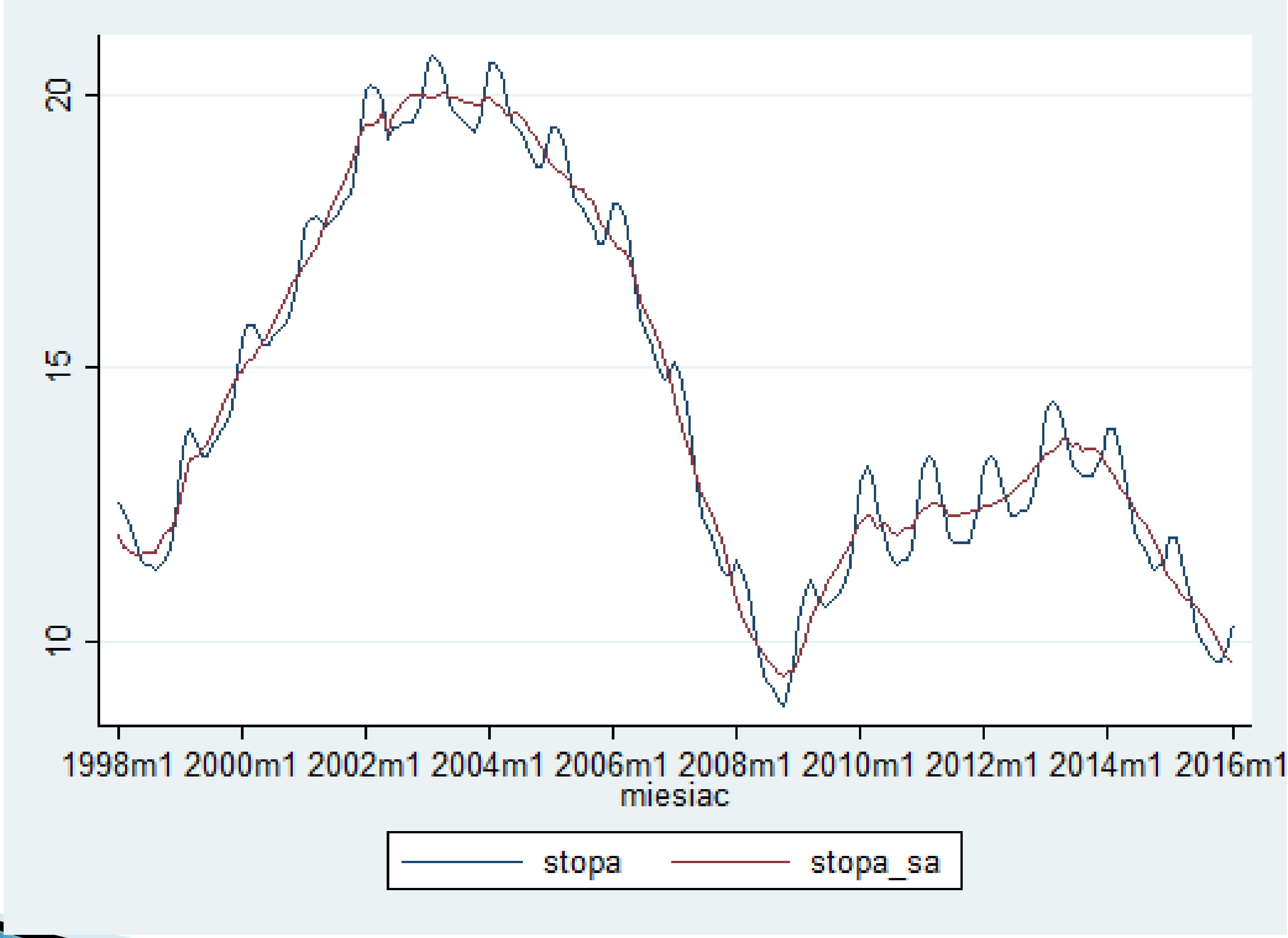
Sezonowość

- Uwzględnienie problemu sezonowości w procesie estymacji:
  c) zastosowanie różnicowania sezonowego: zamiast pierwotnych zmiennych stosujemy różnice miedzy tymi zmiennymi a wartościami tych samych zmiennych sprzed roku:

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

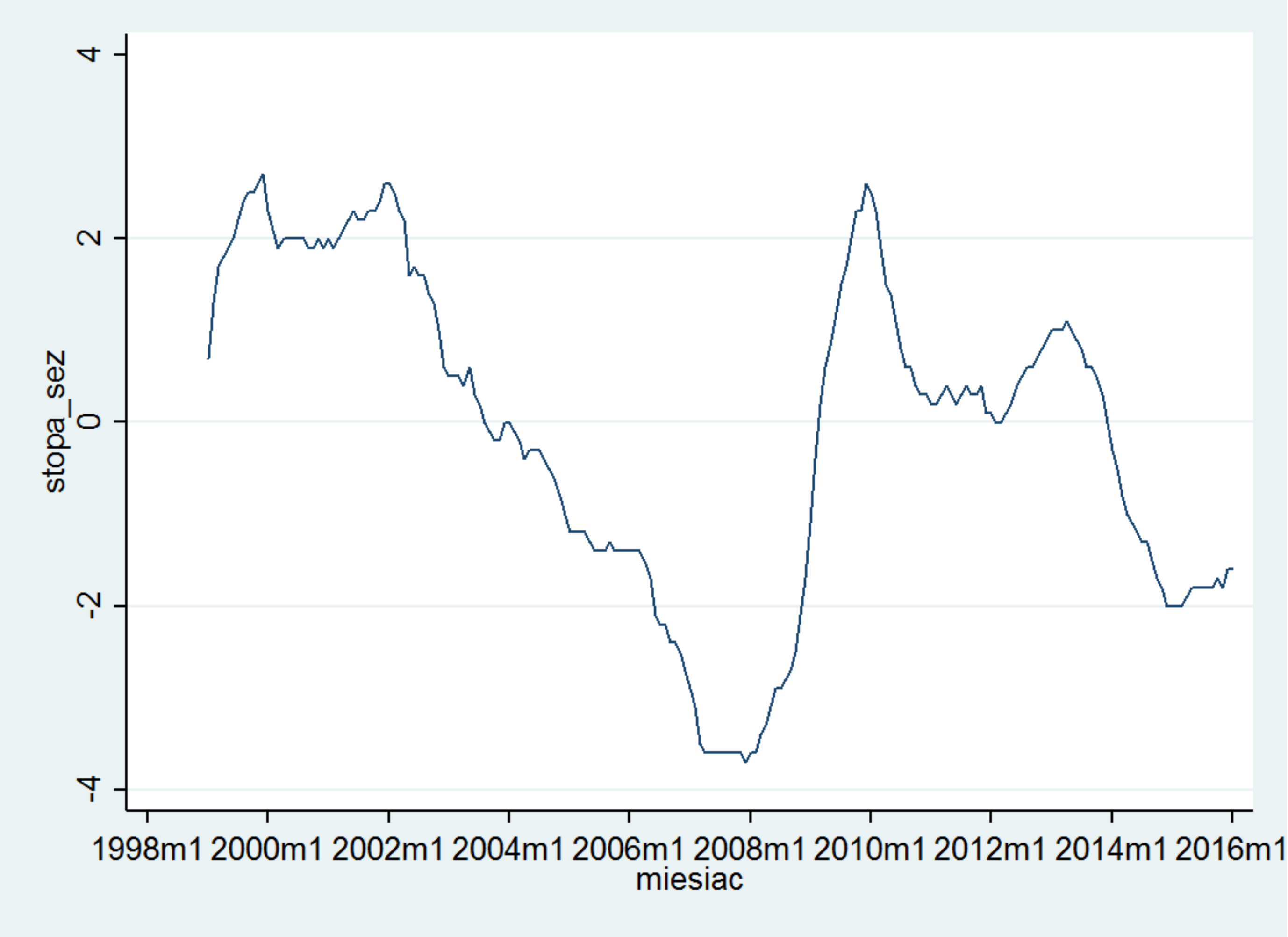
gdzie: s=4 dla zmiennych kwartalnych
s=12 dla zmiennych miesięcznych itd.

Sezonowość



20
15
10
1998m1 2000m1 2002m1 2004m1 2006m1 2008m1 2010m1 2012m1 2014m1 2016m1
miesiac
stopa
stopa_sa

Sezonowość



4
2
0
-2
-4
stopa_sez
1998m1 2000m1 2002m1 2004m1 2006m1 2008m1 2010m1 2012m1 2014m1 2016m1
miesiac

Plan wykładu

- 1. Szereg czasowy
- 2. Sezonowość
- 3. Zmienne stacjonarne
- 4. Zmienne zintegrowane
- 5. Regresja pozorna

Zmienne stacjonarne

- Zmienna jest stacjonarna w sensie słabym (stacjonarność kowariancyjna) jeśli:
  - $E(y_t) = \mu < \infty$ - wartość oczekiwania jest skończona i stała w czasie
  - $Var(y_t) = \sigma^2 < \infty$ - wariancja jest skończona i stała w czasie
  - $Cov(y_{t1}, y_{t1+h}) = Cov(y_{t2}, y_{t2+h}) = \gamma_h$ - dla dowolnych $t1, t2, h$

    kowariancje między realizacjami y_t zależą jedynie od dystansu w czasie h

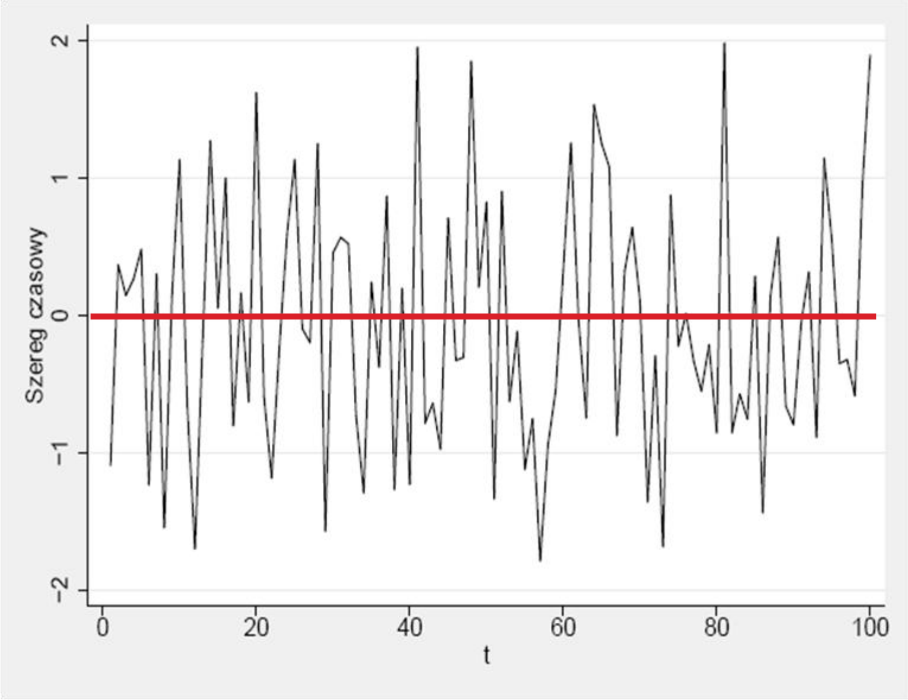
Intuicyjne: zmienna stacjonarna to zmienna, której własności nie zmieniają się w czasie.

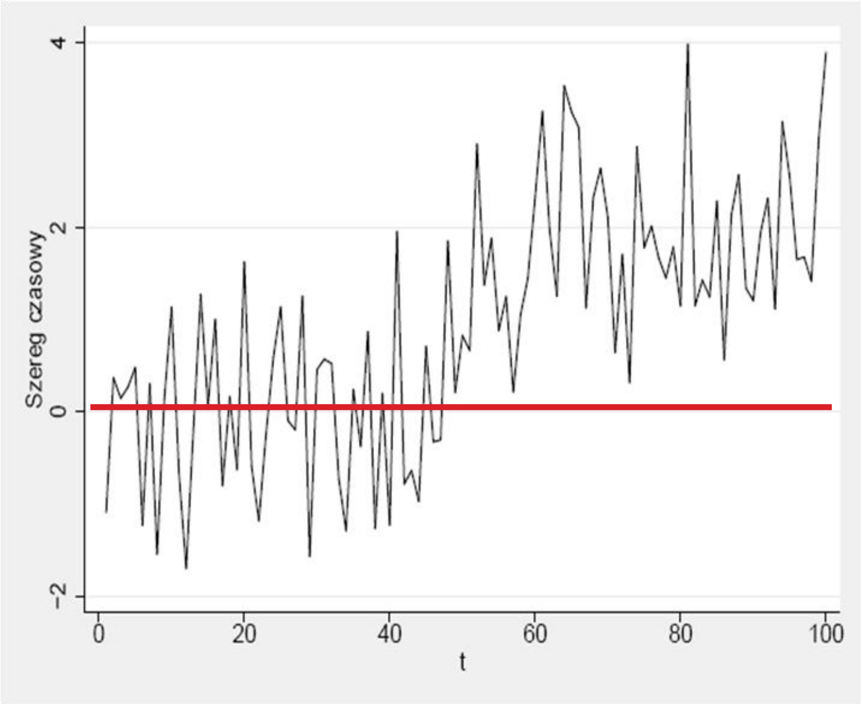
Zmienne stacjonarne

- Jeśli któryś z warunków nie jest spełniony: zmienna niestacjonarna.

Zmienne stacjonarne

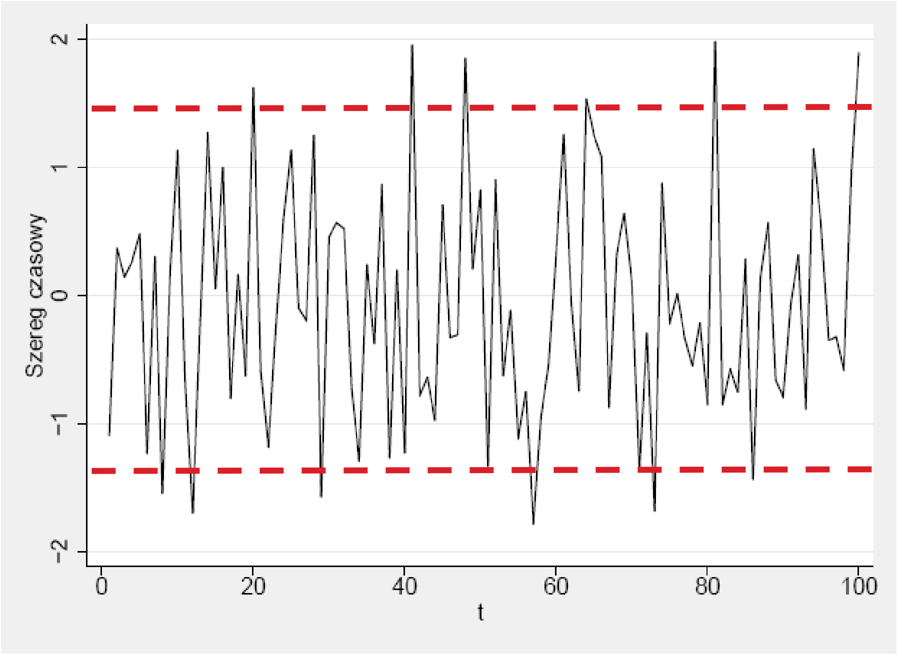
$E(y_t) = \mu < \infty$ - wartość oczekiwania jest skończona i stała w czasie

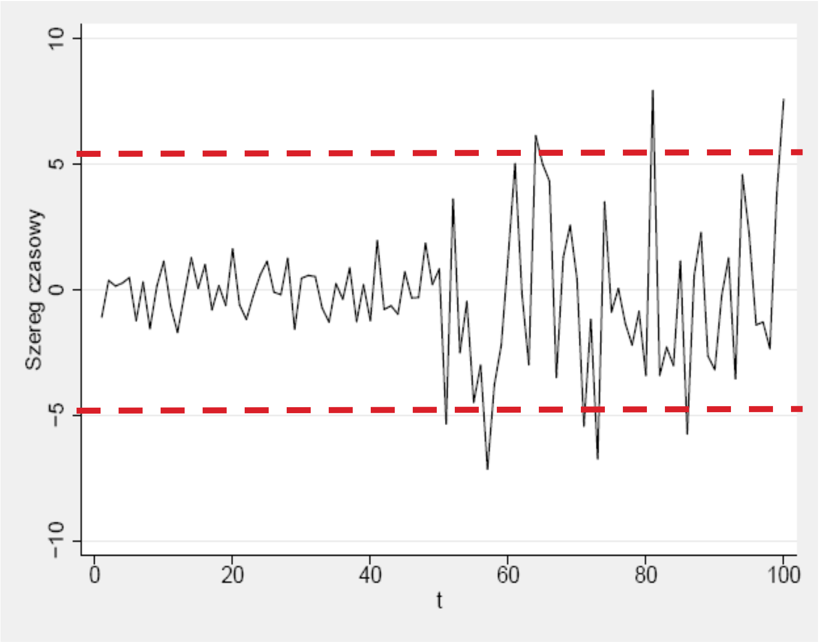




Zmienne stacjonarne

$Var(y_t) = \sigma^2 < \infty$ - wariancja jest skończona i stała w czasie





Zmienne stacjonarne

- Przykład zmiennej stacjonarnej: biały szum (white noise):

$$x_t \sim \mathrm{IID}(0, \sigma^2)$$

IID (Independently and Identically Distributed) – realizacje x_t są niezależne i mają identyczne rozkłady.

Zmienne stacjonarne

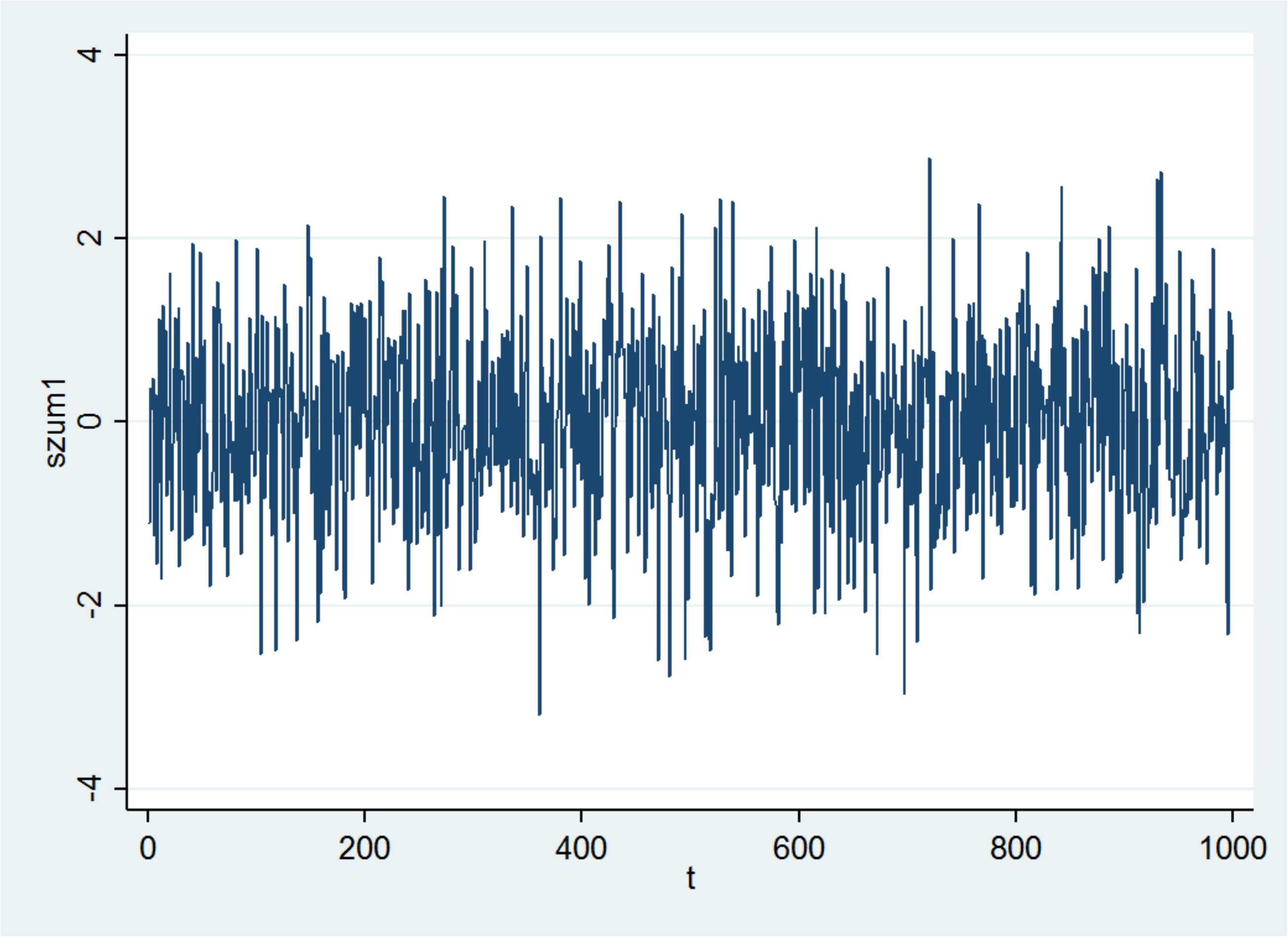
- Dla białego szumu:

$$E(x_t) = 0 < \infty$$

$$Var(x_t) = \sigma^2 < \infty$$

$$Cov(x_t, x_s) = 0 \quad \text{dla} \quad t \neq s$$

Zmienne stacjonarne



Zmienne stacjonarne

- Przykład zmiennej stacjonarnej: model AR(1) dla $|\alpha| < 1$

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

$$\varepsilon_t \sim \mathrm{IID}(0, \sigma^2)$$

Zmienne stacjonarne

- Przykład zmiennej stacjonarnej: model AR(1) dla $|\alpha|<1$

$$E(y_t) = E(\sum_{i=0}^{\infty}\alpha^i \varepsilon_{t-i}) = \sum_{i=0}^{\infty}\alpha^i E(\varepsilon_{t-i}) = 0$$

$$Var(y_t) = Var(\sum_{i=0}^{\infty}\alpha^i \varepsilon_{t-i}) = \sum_{i=0}^{\infty}\alpha^{2i} Var(\varepsilon_{t-i}) = \frac{\sigma^2}{1-\alpha^2}$$

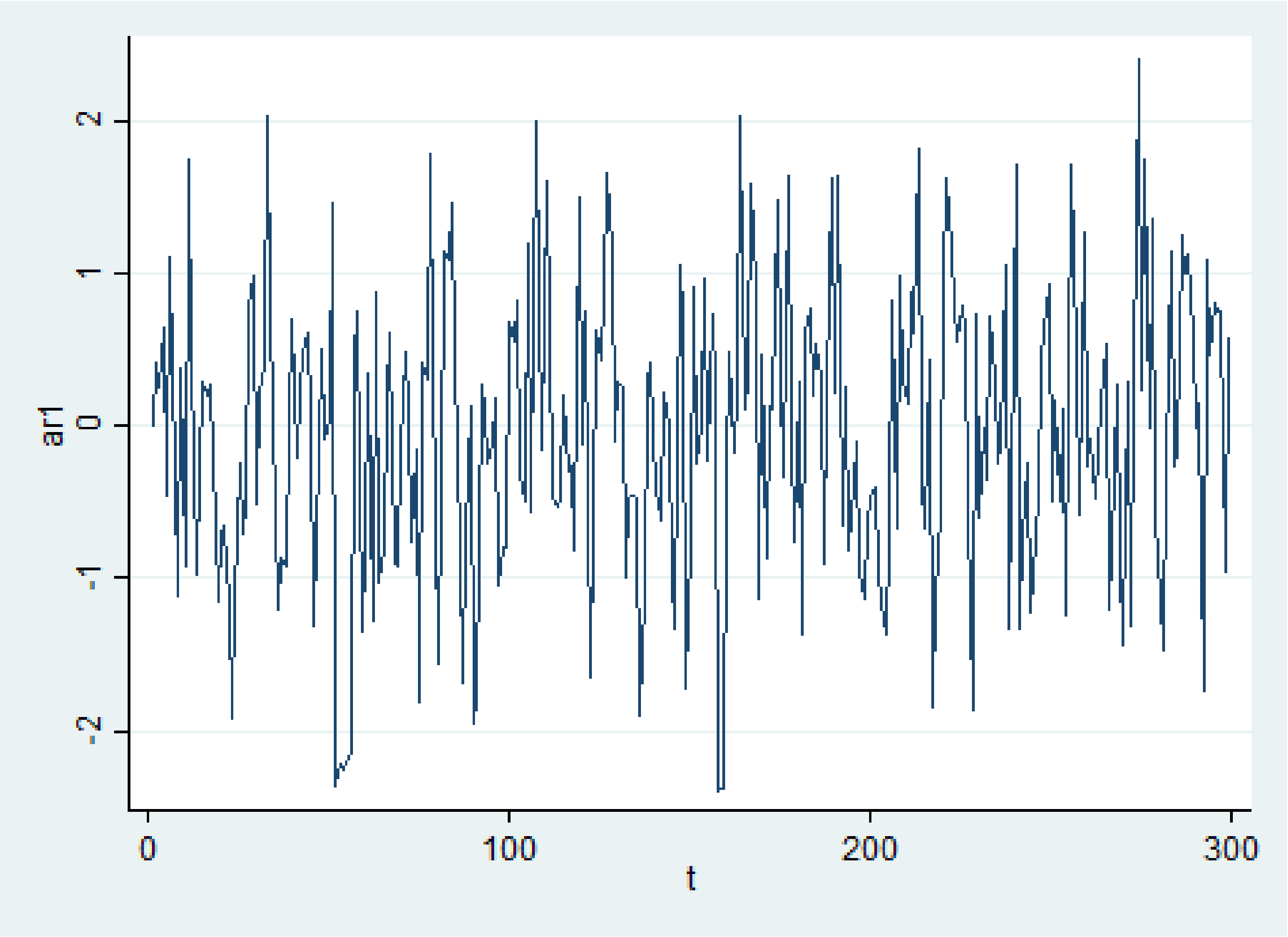
Zmienne stacjonarne

- Przykład zmiennej stacjonarnej: model AR(1) dla $|\alpha|<1$

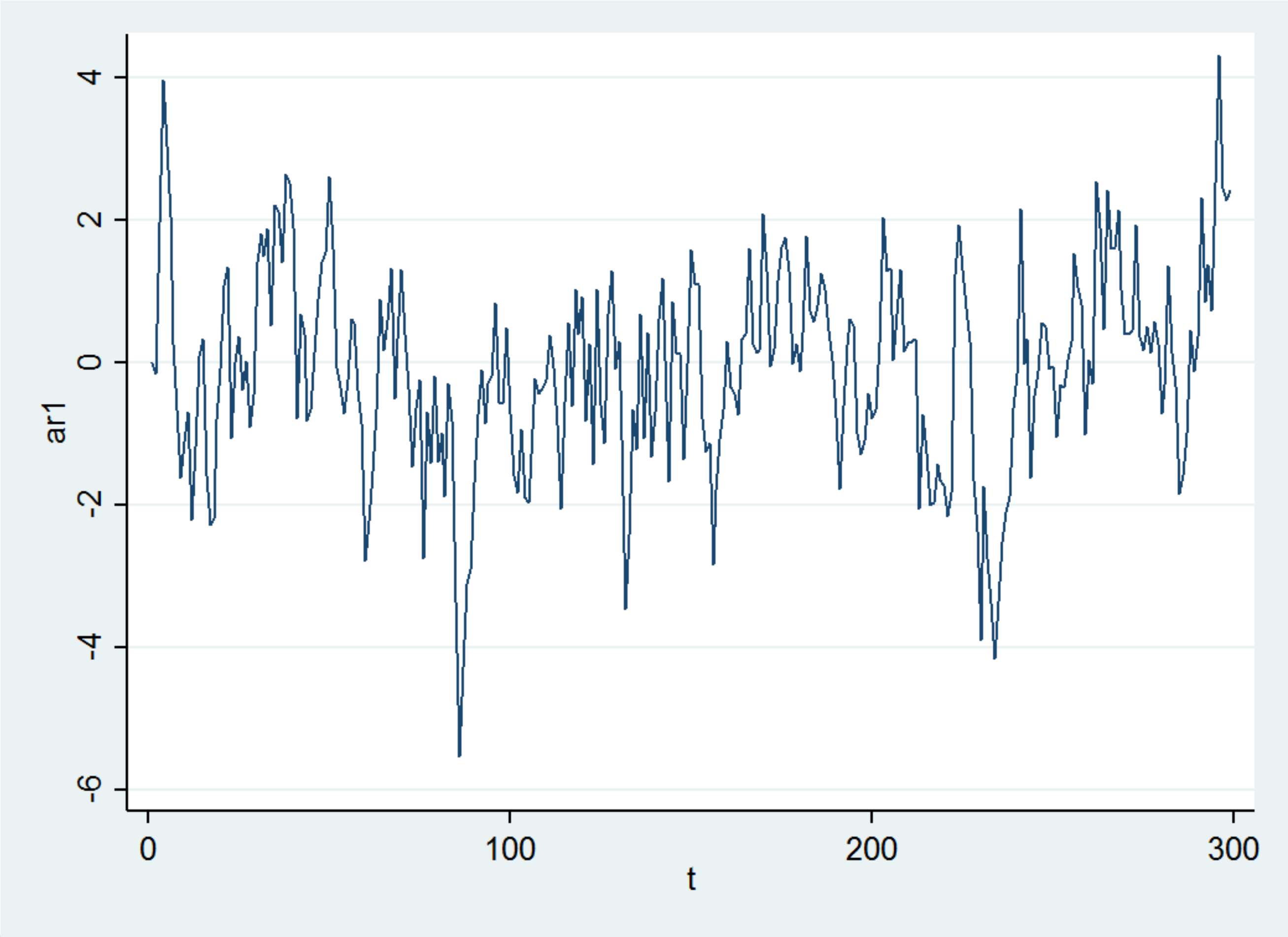
$$Cov(y_t, y_{t-h}) = Cov(\sum_{i=0}^{\infty}\alpha^i\varepsilon_{t-i}, \sum_{i=0}^{\infty}\alpha^i\varepsilon_{t-i-h}) =$$

$$Cov(\sum_{i=0}^{h-1}\alpha^i\varepsilon_{t-i} + \alpha^h\sum_{i=0}^{\infty}\alpha^i\varepsilon_{t-i-h}, \sum_{i=0}^{\infty}\alpha^i\varepsilon_{t-i-h}) = \alpha^h\sum_{i=h}^{\infty}\alpha^{2i}Var(\varepsilon_{t-i-h}) = \alpha^h\frac{\sigma^2}{1-\alpha^2}$$

Zmienne stacjonarne



Zmienne stacjonarne



Zmienne stacjonarne

- Przykład zmiennej niestacjonarnej: błądzenie przypadkowe (random walk)

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

$$\varepsilon_t \sim \mathrm{IID}(0, \sigma^2)$$

Zmienne stacjonarne

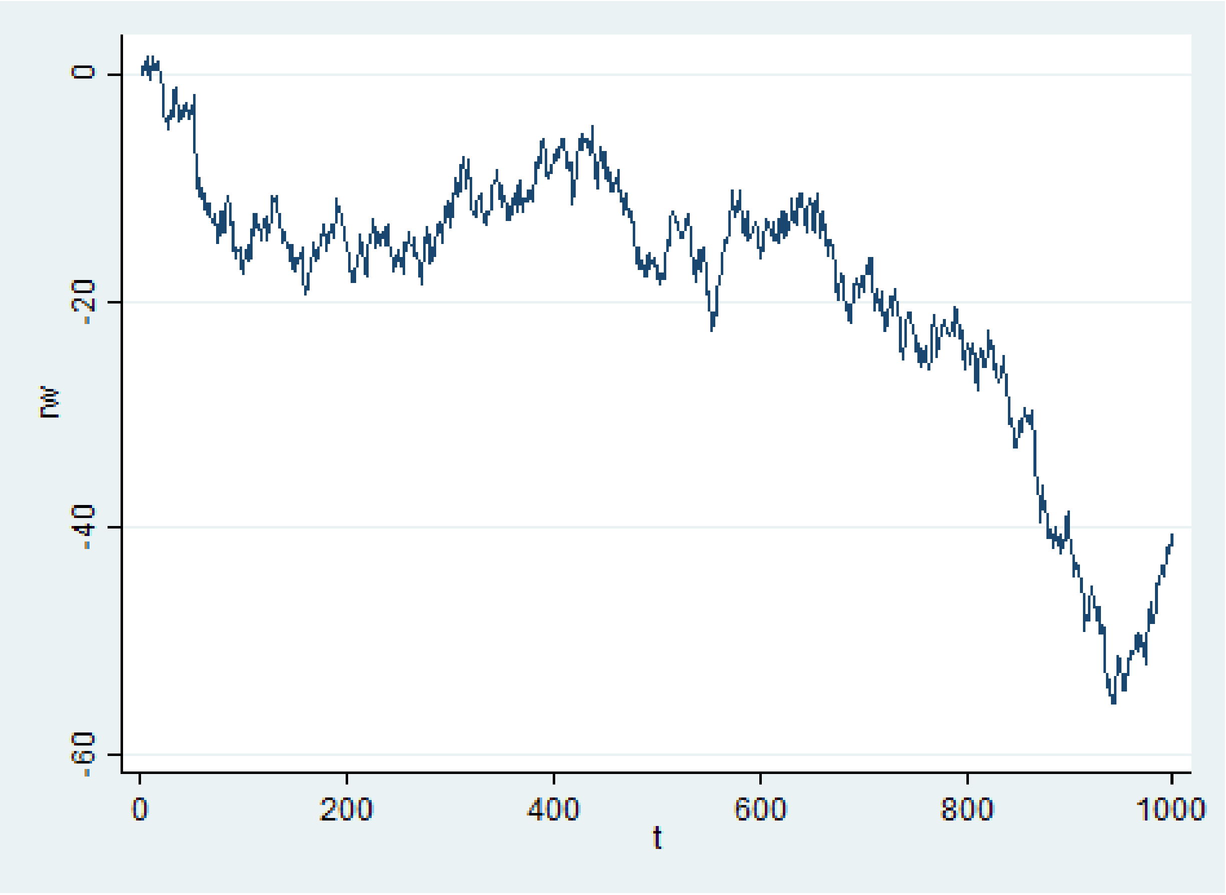
- Przykład zmiennej niestacjonarnej: błądzenie przypadkowe (random walk)

$$E(y_t) = 0$$

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

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

Zmienne stacjonarne



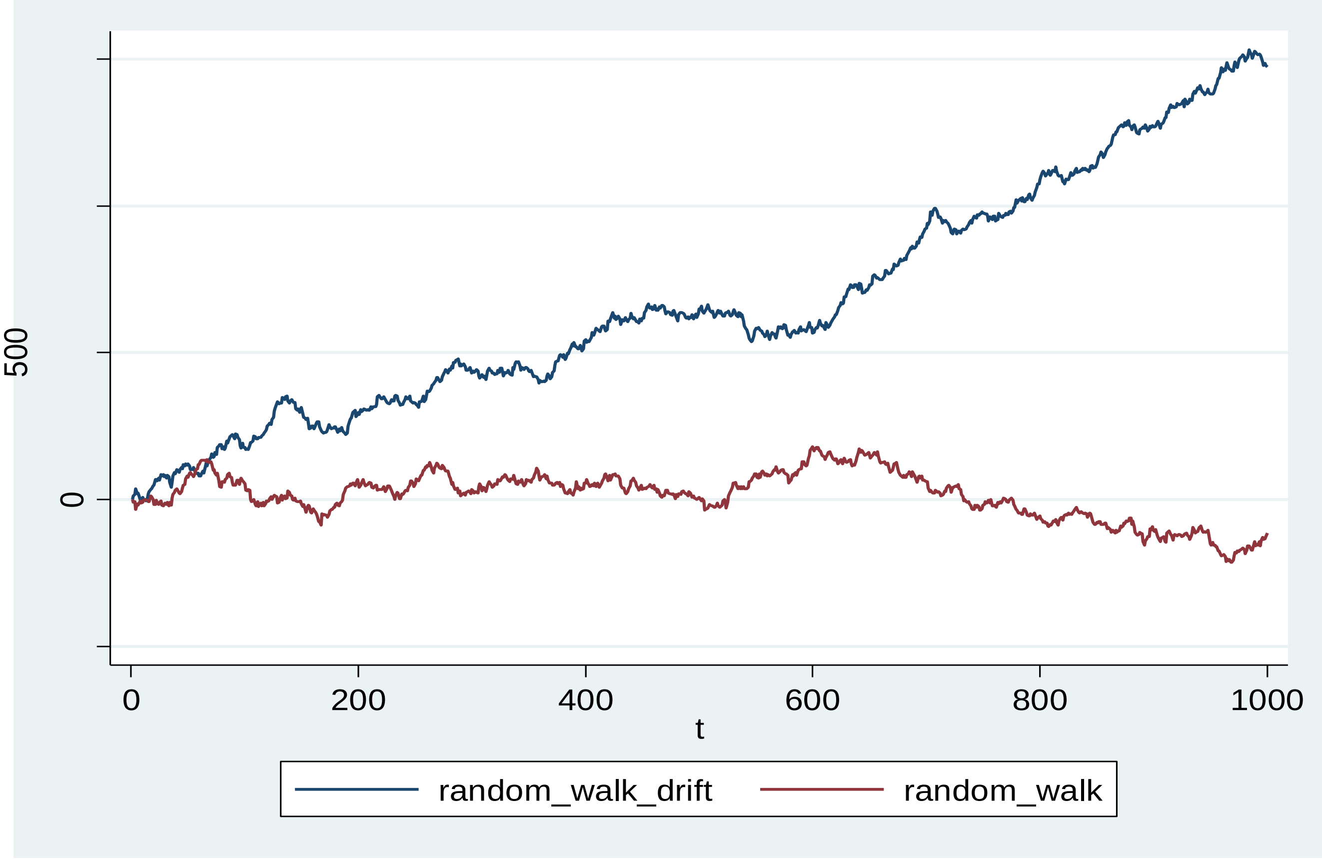
Zmienne stacjonarne

- Przykład zmiennej niestacjonarnej: błądzenie przypadkowe z dryfem

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

$$\varepsilon_t \sim \mathrm{IID}(0, \sigma^2)$$

Zmienne stacjonarne



Zmienne stacjonarne

- Standardowa definicja stacjonarności w wielu przypadkach okazuje się zbyt restrykcyjna: zmienne ekonomiczne oscylują nie tyle wokół stałej ale wokół pewnego trendu.

- Zmienna stacjonarna wokół trendu (trendostacjonarna) jeśli odchylenie od trendu:

$$y_t - E(y_t)$$

jest stacjonarne.

Zmienne stacjonarne

- Przykład zmiennej trendostacjonarnej: trend liniowy

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

$$E(y_t) = \alpha + \beta t$$

$$y_t - E(y_t) = \varepsilon_t$$

Plan wykładu

- 1. Szereg czasowy
- 2. Sezonowość
- 3. Zmienne stacjonarne
- 4. Zmienne zintegrowane
- 5. Regresja pozorna

Zmienne zintegrowane

- Zmienne zintegrowane: zmienne niestacjonarne, które można sprowadzić do stacjonarności poprzez różnicowanie.

- Zmienna, która po zastosowaniu d-tych różnic staje się zmienną stacjonarną oznaczamy jako:

$$y_t \sim I_d$$

- Mówimy, ze zmienna y_t jest zintegrowana rzędu d.
- Zmienne stacjonarne są zintegrowane rzędu 0:

$$y_t \sim I(0)$$

Zmienne zintegrowane

- Przykład zmiennej niestacjonarnej: błądzenie przypadkowe (random walk)

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

$$\varepsilon_t \sim \mathrm{IID}(0, \sigma^2)$$

- Różnicując zmienną y_t :

$$\Delta y_t = \varepsilon_t$$

Biały szum, zmienna I(0). Wobec tego błądzenie przypadkowe jest zmienną I(1)

Plan wykładu

- 1. Szereg czasowy
- 2. Sezonowość
- 3. Zmienne stacjonarne
- 4. Zmienne zintegrowane
- 5. Regresja pozorna

Regresja pozorna

- Dlaczego w ogóle zajmować się stacjonarnością zmiennych? Założenie o stacjonarności zmiennych w modelu jest niezbędne przy wyprowadzaniu rozkładów typowych statystyk testowych używanych przy testowaniu hipotez.

- Jeśli zmienne w modelu są niestacjonarne to rozkłady asymptotyczne statystyk testowych są niestandardowe, co może prowadzić do błędnych wyników wnioskowania statystycznego.

- Przykładem jest problem regresji pozornej.

Regresja pozorna

- Występuje gdy część zmiennych w modelu nie jest stacjonarna (najczęściej I(1)).
- W takim przypadku statystyki t dla zmiennych I(1) okazują się często istotne nawet jeśli między zmienną objaśnianą a zmiennymi objaśniającymi nie ma żadnego związku.

Regresja pozorna

- Generujemy obserwacje dwóch niezależnych zmiennych niestacjonarnych:

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

$$x_t = x_{t-1} + u_t$$

$$\varepsilon_t \sim \mathrm{N}(0,1)$$

$$\mathrm{u}_\mathrm{t} \sim \mathrm{N}(0,1)$$

Następnie przeprowadzamy regresję y_t na X_t, zapisujemy statystykę t i DW.

Regresja pozorna

- Następnie powatrzamy całość 1000 razy zapisując za każdym razem wynik. Mając serię statystyk t obliczamy średnią, odchylenie standardowe, skośność, kurtozę i porównujemy z parametrami testu t-Studenta. Przy poziomie istotności 5% powinniśmy uzyskać istotny wynik w mniej więcej 5% przypadków.

Regresja pozorna

	Teoretyczne	Regresja y na x
Średnia	0,000	0,048
Odch.stand.	1,021	4,849
Skośność	0,000	-0,214
Kurtoza	3,125	3,845
5% percentyl	1,677	8,294
% istotnych wyników	5,000	64,200
	Statystyka DW	
Średnia	2,000	0,313

Pytania teoretyczne

1. Wyjaśnić co to znaczy, że w danych występuje sezonowość i omówić sposoby uwzględniania sezonowości w procesie modelowania.
2. Podać definicję zmiennej stacjonarnej i trendostacjonarnej.
3. Wyjaśnić, co to są zmienne I(0) i I(1) i udowodnić, że biały szum jest zmienną I(0) a błądzenie przypadkowe zmienną I(1).
4. Wyjaśnić na czym polega zjawisko regresji pozornej.
5. Dlaczego przed przystąpieniem do weryfikacji hipotez o istotności zmiennych w modelu szacowanym na szeregu czasowym powinniśmy przetestować ich rząd integracji?

Dziękuję za uwagę