Moc testów

Wykład 4

Piotr Dybka

Ekonometria praktyczna

Czym jest stacjonarność? – przypomnienie

- W ścisłym sensie stacjonarność oznacza, że w procesie stochastycznym y_t łączny rozkład prawdopodobieństwa dla $\{y_t, y_{t+1}, \ldots, y_T\}$ jest taki sam dla dowolnych obserwacji i ich przesunięcia w czasie $\{y_{t+m}, y_{t+m+1}, \ldots, y_{T+m}\}$.
- Problemem jest to, że tak zdefiniowanej stacjonarności (*silnej* stacjonarności) nie da się zweryfikować na podstawie obserwowanych danych. Dlatego empirycznie testowana jest *słaba* stacjonarność, która występuje, gdy wartość oczekiwana, wariancja oraz kowariancja procesu są stałe w czasie:

$$\mu_t = \mu$$

$$\sigma_t^2 = \sigma^2$$

$$cov(y_t, y_{t+s})_t = cov(y_t, y_{t+s}) \quad \text{dla } s = 1, 2, \ldots$$

Stopień zintegrowania szeregu czasowego

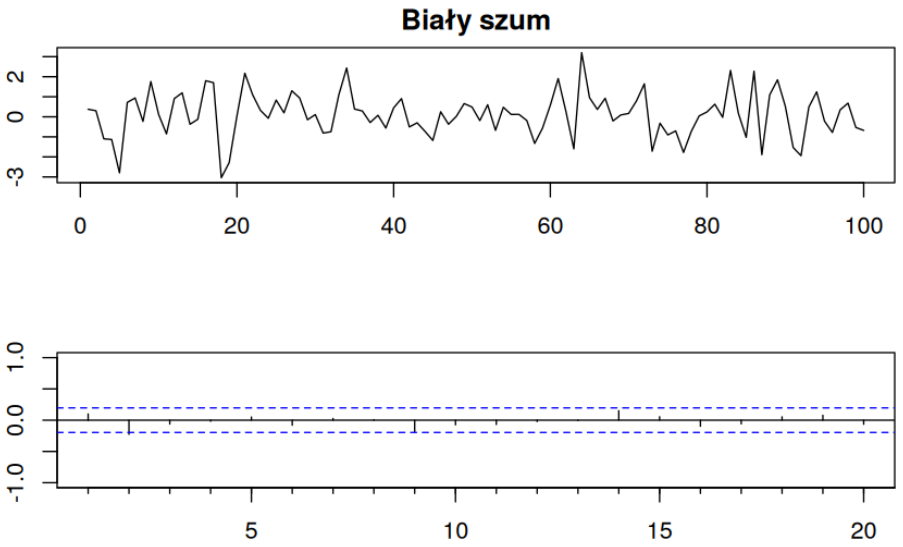
Stopień zintegrowania oznaczamy jako $I(K)$. Aby określić stopień zintegrowania szeregu:

1. Testujemy stacjonarność y_t. W przypadku gdy szereg jest stacjonarny, to mamy $y_t \sim I(0)$. W przypadku niestacjonarności analizujemy stacjonarność Δy_t, itd.
2. Jeżeli dla $k = 1, 2, \ldots, K-1$ zmienna $\Delta^k y_t$ jest niestacjonarna, a zmienna $\Delta^K y_t$ jest stacjonarna, to szereg y_t jest niestacjonarny, zintegrowany w stopniu K, co zapisujemy $y_t \sim I(K)$.

Biały szum

Biały szum (*white noise*, WN):

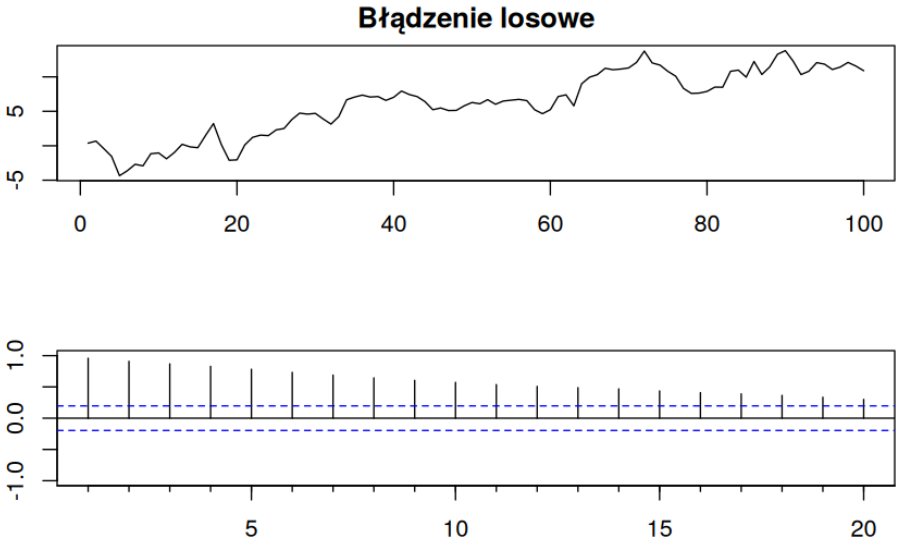
- Dla dowolnych $s \neq t$ obserwacje y_t oraz y_s są od siebie niezależne, mają jednakowy rozkład prawdopodobieństwa oraz ich wartości oczekiwane są równe 0.
- Przykład: wielokrotne losowanie z rozkładu normalnego.
- Biały szum jest stacjonarny.



Definicja powyżej to *silny* (i.i.d.) biały szum. W praktyce często wystarcza *słaba* definicja: $E(y_t) = 0$, $Var(y_t) = \sigma^2$, $cov(y_t, y_s) = 0$ dla $s \neq t$ – bez pełnej niezależności czy jednakowego rozkładu.

Błądzenie losowe

$$y_t = y_{t-1} + \epsilon_t, \qquad \epsilon_t \sim WN$$



- Proces błądzenia losowego jest niestacjonarny:

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

- Zatem

$$E(y_t) = y_0$$

$$Var(y_t) = t\sigma^2$$

$$Cov(y_t, y_{t+s}) = t\sigma^2$$

Szereg o wysokiej persystencji

- Załóżmy, że mamy szereg

$$y_t = \rho y_{t-1} + \epsilon_t, \qquad \epsilon_t \sim WN,\ \rho \in (-1, 1)$$

- Jest to szereg, który także jest stacjonarny – cechuje się powrotem do średniej (wartości oczekiwanej).
- Jednak dla dużych wartości $0.9 < |\rho| < 1$ cechuje się **wysoką persystencją** (powolnym powrotem do średniej).
- Taki szereg jest stacjonarny, jednak formalne testy statystyczne mają często problem z poprawną identyfikacją stacjonarności – zobaczymy to na konkretnych liczbach w dalszej części wykładu.

Półokres wygasania

- Dla stacjonarnego procesu $(y_t - \mu) = \rho(y_{t-1} - \mu) + \epsilon_t$, czyli dla $\rho \in (-1, 1)$, y_t powraca do swojej wartości oczekiwanej μ.
- Tempo powrotu do średniej jest opisane przez parametr $\delta = \rho - 1$.
- Przykładowo dla $\rho = 0{,}9$ (czyli $\delta = -0{,}1$) odchylenie od średniej wygasa w tempie 10% w ciągu jednego okresu.
- Półokres wygasania (*half-life*; hl) określa, po jakim czasie następuje wygaśnięcie połowy odchylenia zmiennej od poziomu równowagi:

$$hl = \frac{\ln(0{,}5)}{\ln(\rho)}$$

- Dla $\rho = 0{,}9$ wartość hl wynosi ok. 6,5 okresu.

Błąd I i II rodzaju

Przypomnienie ze statystyki:

- Błąd pierwszego rodzaju to sytuacja, gdy odrzucamy H_0 mimo że jest ona prawdziwa.
- Błąd drugiego rodzaju to sytuacja, gdy nie odrzucamy H_0 mimo że jest ona fałszywa.

		Stan faktyczny	
		H_0 prawdziwa	H_0 fałszywa
Decyzja na próbie	Brak odrzucenia H_0	decyzja prawidłowa	błąd II rodzaju
	Odrzucenie H_0	błąd I rodzaju	decyzja prawidłowa

Uwaga: prawdopodobieństwo błędu II rodzaju oznaczamy β. **Moc testu to nie to samo, co błąd II rodzaju** – moc to $1 - \beta$, czyli prawdopodobieństwo poprawnego odrzucenia fałszywej H_0. Im wyższa moc, tym rzadziej popełniamy błąd II rodzaju.

Moc testu – definicja formalna

- **Poziom istotności** $\alpha = P(\text{odrzucenie } H_0 \mid H_0 \text{ prawdziwa})$ – ustalamy go z góry (zwykle 0,05 lub 0,01); to prawdopodobieństwo błędu I rodzaju.
- **Moc testu** $1 - \beta = P(\text{odrzucenie } H_0 \mid H_0 \text{ fałszywa})$ – to *nie* jest coś, co ustalamy z góry: zależy od konkretnej wartości parametru w ramach H_1, od liczebności próby N oraz od specyfikacji samego testu.
- W kontekście testów stacjonarności: moc testu ADF to prawdopodobieństwo poprawnego odrzucenia H_0 (niestacjonarność), gdy szereg *naprawdę* jest stacjonarny, dla **konkretnej** wartości $\rho < 1$.
- Moc **nie jest jedną liczbą** dla danego testu – to *funkcja*: zależy od tego, jak bardzo H_1 (tu: ρ) odbiega od H_0 (tu: $\rho = 1$), od N i od specyfikacji regresji testowej.
- W tym wykładzie sprawdzimy moc testów ADF i KPSS **empirycznie**, za pomocą symulacji Monte Carlo – wygenerujemy wiele sztucznych szeregów o znanym ρ i sprawdzimy, jak często test podejmuje poprawną decyzję.

Test Dickeya-Fullera

Najpopularniejszy test stacjonarności o układzie hipotez:

$$H_0: \textit{ szereg jest niestacjonarny} \qquad H_1: \textit{ szereg jest stacjonarny}$$

- W teście DF zakładamy, że y_t jest generowany przez proces AR(1):

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

- Jednak oszacowanie parametru ρ na podstawie powyższego równania może prowadzić do nieprawidłowych wniosków ze względu na występowanie regresji pozornej, dlatego szacowane jest równanie:

$$\Delta y_t = \delta y_{t-1} + \epsilon_t$$

- Gdzie $\delta = (\rho - 1)$.

$$H_0: \ \delta = 0 \iff \rho = 1 \qquad H_1: \ \delta < 0 \iff \rho < 1$$

Test Dickeya-Fullera

- Aby uniknąć ryzyka autokorelacji składnika losowego w regresji testowej (które prowadziłoby do obciążenia estymatora), można przekształcić test DF do rozszerzonego testu ADF:

$$\Delta y_t = \delta y_{t-1} + \sum_{i=1}^{P} \beta_i \, \Delta y_{t-i} + \epsilon_t$$

- Statystyka testowa DF (ADF) jest dana wzorem

$$DF = \frac{\hat{\delta}}{S_\delta}$$

- Statystyka ta nie ma standardowego rozkładu t-Studenta – pod H_0 (przy $\rho = 1$) jej rozkład jest funkcją procesu Wienera i nie ma prostej postaci analitycznej. Wartości krytyczne wyznacza się metodami numerycznymi (symulacyjnie) i tabelaryzuje (Dickey, Fuller 1979; MacKinnon 1996).

Wariancja długookresowa

Dla y_t będącego stacjonarnym procesem, dla statystyki $\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$ wartość oczekiwana wynosi $E(\bar{y}) = \mu$, a wariancję $\hat{\gamma}_0$ możemy policzyć jako:

$$Var(\bar{y}) = \frac{1}{T}\left(\gamma_0 + 2\sum_{s=1}^{T}\left(1 - \frac{s}{T}\right)\gamma_s\right)$$

Gdzie $\gamma_s = cov(y_t, y_{t-s})$. Przy zerowej autokorelacji wzór upraszcza się do $Var(\bar{y}) = \frac{\gamma_0}{T}$. Wariancja długookresowa jest granicą powyższego wyrażenia dla $T \to \infty$.

Dla skończonej próby można ją oszacować metodą Neweya-Westa:

$$\hat{\gamma}_\infty = \hat{\gamma}_0 + 2\sum_{s=1}^{L}\kappa_{s,L}\hat{\gamma}_s$$

$$\kappa_{s,L} = 1 - \frac{s}{L+1}$$

L to **maksymalna liczba uwzględnianych opóźnień** (tzw. parametr obcięcia, *bandwidth*) – nie należy go mylić z operatorem opóźnień z notacji ARMA.

Test KPSS

Układ hipotez (uwaga – jest inny niż w poprzednich testach!):

$$H_0: \textit{ szereg jest stacjonarny} \qquad H_1: \textit{ szereg jest niestacjonarny}$$

- Odwrócony układ hipotez względem DF/ADF.
- Wykorzystuje informacje o długookresowej wariancji składnika losowego.

1. Szacujemy model regresji y_t względem stałej (ewentualnie trendu) i zapisujemy reszty e_t.
2. Liczymy sumy reszt $S_t = \sum_{i=1}^{t} e_i$ dla każdego okresu ($t = 1, 2, \ldots, T$).
3. Metodą Neweya-Westa szacujemy długookresową wariancję reszt $\hat{\gamma}_\infty$.
4. Obliczamy statystykę testu:

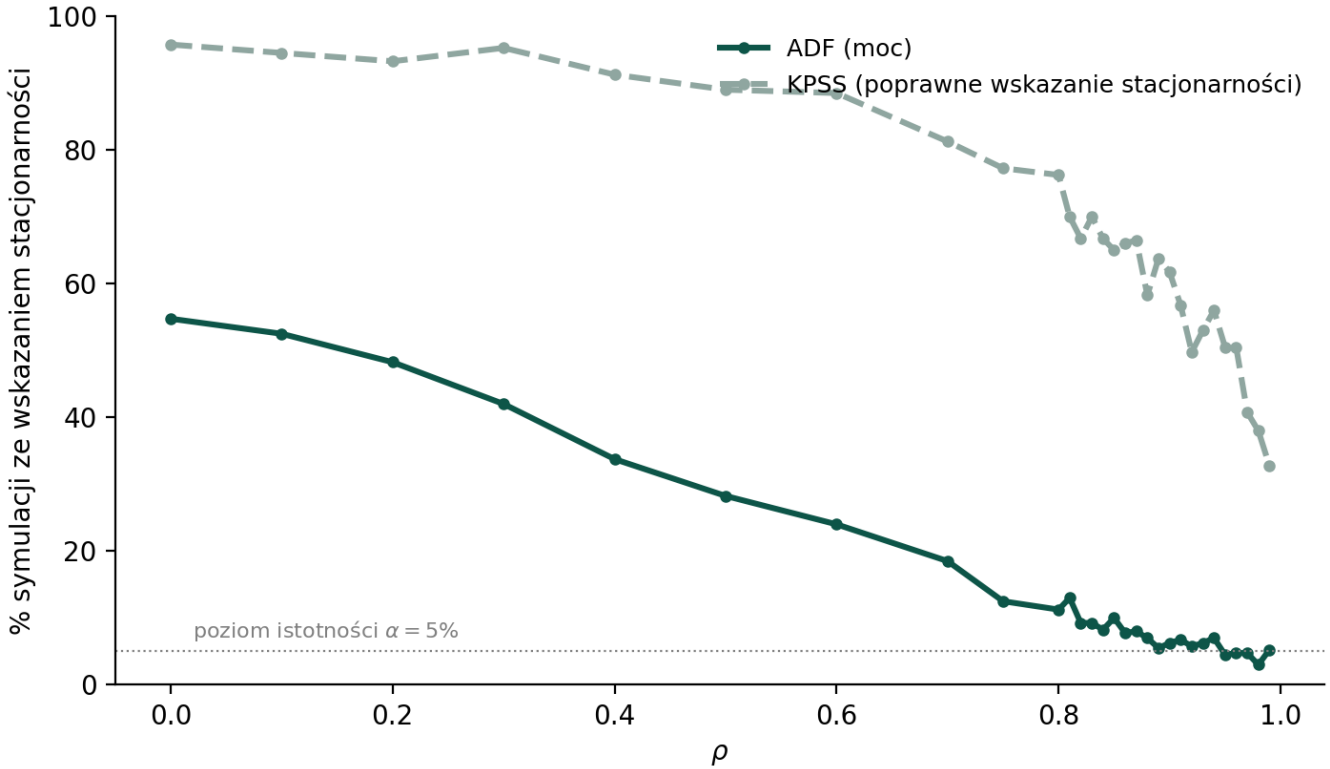
$$KPSS = \frac{1}{T^2} \frac{\sum_{i=1}^{T} S_i^2}{\hat{\gamma}_\infty}$$

Symulacja Monte Carlo mocy testu

Jak zmierzyć moc testu w praktyce, skoro zależy ona od nieznanego w rzeczywistych danych ρ? **Odpowiedź**: symulujemy dane o znanym ρ i sprawdzamy, jak często test podejmuje poprawną decyzję.

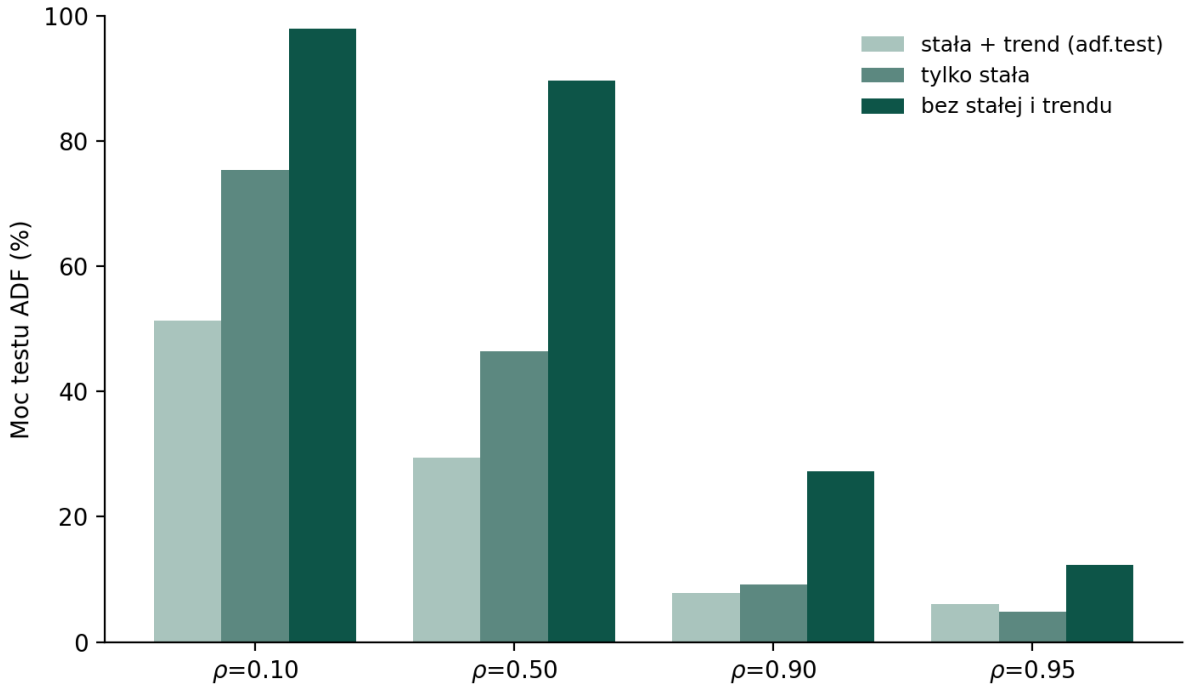
1. Ustalamy ρ (np. $\rho = 0{,}95$) oraz liczbę powtórzeń M (np. $M = 250$).
2. Dla każdego powtórzenia $m = 1, \ldots, M$: generujemy szereg $y_t = \rho y_{t-1} + \epsilon_t$ ($n = 50$ obserwacji), liczymy p-value testów ADF i KPSS dla tego szeregu.
3. Odsetek powtórzeń, w których ADF odrzuca H_0 (p-value $< 0{,}05$), to **oszacowanie mocy** testu ADF dla tego ρ i tego n.
4. Analogicznie dla KPSS, tyle że tutaj H_0 jest prawdziwe (proces jest stacjonarny) – odsetek, w którym KPSS *nie* odrzuca H_0, mówi nam o rzeczywistym rozmiarze testu, a nie o jego mocy w ścisłym sensie (o mocy KPSS moglibyśmy mówić dopiero, symulując dane naprawdę niestacjonarne).
5. Powtarzamy dla siatki wartości ρ (np. $\rho = 0{,}90, 0{,}91, \ldots, 0{,}99$), aby uzyskać **krzywą mocy**.

Krzywa mocy: ADF i KPSS



Wyniki symulacji ($n = 50$, $M = 400$ dla każdego ρ). Moc ADF systematycznie spada wraz ze wzrostem ρ i przy $\rho \to 1$ zbliża się do poziomu istotności ($\alpha = 5\%$) – test praktycznie **traci zdolność odróżnienia** silnie persystentnego procesu stacjonarnego od niestacjonarnego. KPSS zachowuje wyższy odsetek poprawnych wskazań dłużej, ale też spada z ρ.

Specyfikacja deterministyczna a moc testu



`tseries::adf.test()` **zawsze** dopasowuje regresję ze stałą i trendem liniowym – nie da się tego zmienić. Nasz proces symulacyjny nie ma ani dryfu, ani trendu. Zbyt bogata specyfikacja **silnie obniża moc**: dla $\rho = 0{,}1$ moc spada z ok. 98% (bez stałej i trendu) do ok. 51% (ze stałą i trendem) – na tych samych danych! **Wniosek**: specyfikacja deterministyczna powinna odpowiadać naturze danych, a nie być wybierana automatycznie.

Testy stacjonarności dla danych panelowych

W przypadku danych panelowych testowanie stacjonarności jest nieco trudniejsze, ponieważ mamy zarówno wymiar czasowy, jak i przekrojowy. W praktyce jest wiele różnych testów o układzie hipotez:

$$H_0: \textit{ szeregi w panelu są niestacjonarne} \qquad H_1: \textit{ szeregi w panelu są stacjonarne}$$

Idea testów polega na przeprowadzeniu oddzielnych testów (np. ADF) dla każdej jednostki panelu i uśrednieniu wyniku (np. Im, Pesaran, Shin (2003)) lub połączeniu danych i wyliczeniu statystyki testowej Levin, Lin, Chu (2002):

1. Przeprowadzamy test ADF dla każdej jednostki panelu.
2. Zapisujemy reszty z modeli pomocniczych ADF.
3. Liczymy relację wariancji długookresowej do krótkookresowej dla każdej jednostki panelu.
4. Liczymy statystykę testową.

Test Im, Pesaran, Shin (2003)

Układ hipotez taki sam jak dla testu Levin, Lin, Chu (2002):

H_0 : *szeregi w panelu są niestacjonarne* H_1 : *szeregi w panelu są stacjonarne*

- **Kluczowa różnica** względem LLC nie sprowadza się tylko do sposobu uśredniania – dotyczy **postaci hipotezy alternatywnej**:
  - **LLC**: zakłada **wspólny** parametr ρ dla wszystkich jednostek panelu. H_1: *wszystkie* kraje mają szereg stacjonarny (z tym samym ρ) – założenie restrykcyjne i często nierealistyczne dla niejednorodnych paneli.
  - **IPS**: pozwala na **zróżnicowane** ρ_i w różnych jednostkach. H_1: *przynajmniej część* krajów ma szereg stacjonarny. Statystyką testową jest znormalizowana średnia indywidualnych statystyk ADF ($\bar{t}$), co i tak wymaga jednorodnej alternatywy tylko w ograniczonym sensie.
- Konsekwencja: gdy panel jest niejednorodny (część krajów stacjonarna, część nie), LLC i IPS mogą dawać **sprzeczne wnioski** – zobaczymy to na przykładzie.

Testy panelowe – przykład

Panel 28 krajów UE (w tym Wielka Brytania, przed brexitem), 1995–2017, sześć zmiennych makroekonomicznych (`pmax=4`, `exo="intercept"`):

Zmienna	LLC (staty.)	LLC (p)	IPS (staty.)	IPS (p)
PAFP	−6,28	< 0,001	−3,59	< 0,001
SLE	−7,26	< 0,001	−1,65	0,049
X1	−2,86	0,002	0,63	0,734
FDIYS	−4,39	< 0,001	−1,73	0,042
GOVC1	−6,80	< 0,001	−3,77	< 0,001
POT	1,92	0,973	−0,51	0,305

- Dla większości zmiennych oba testy się **zgadzają**.
- Dla **X1** testy dają **sprzeczne wnioski**: LLC odrzuca H_0 (p=0,002, wskazuje stacjonarność), IPS nie odrzuca (p=0,734, wskazuje niestacjonarność). To dokładnie sytuacja z poprzedniego slajdu: jeśli w panelu jest mieszanka krajów stacjonarnych i niestacjonarnych, założenie o wspólnym ρ (LLC) może prowadzić do odrzucenia H_0, mimo że część krajów faktycznie ma pierwiastek jednostkowy.