



AKADEMIA GÓRNICZO-HUTNICZA IM. STANISŁAWA STASZICA W KRAKOWIE  
Wydział Zarządzania

*Projekt EFID - Projekt 5A*

Autorzy: *Patrycja Piła, Mateusz Strojek, Julia Szutka, Magdalena Wnuk*

Kierunek studiów: *Informatyka i Ekonometria*

Prowadzący: *dr hab. Tomasz Wójtowicz*

Kraków, 2025

**Wersja E** (działanie różnych wersji testu ADF) Dla ustalonej długości danych  $N = 100$  i dla  $\phi_0 = 1$  i  $\phi_1 = 0,9$  oraz  $\sigma^2 = 1$  przeprowadzić następującą symulację:

1. Wygenerować  $N$  wartości modelu AR (1):  $X_t = \phi_0 + \phi_1 X_{t-1} + \varepsilon_t$ ,  $\varepsilon_t \sim iid (0, \sigma^2)$ .
2. Za pomocą różnych wersji testu ADF (bez stałej, ze stałą, z trendem liniowym) zbadać stacjonarność wygenerowanych danych.
3. Zapamiętać p-value przeprowadzonych testów.
4. Kroki 1-3 powtórzyć 1000 razy.

Obliczyć statystyki opisowe rozkładów uzyskanych wartości p-value. Przedstawić graficznie ich te rozkłady. Obliczyć odsetek przypadków odrzuceń hipotezy głównej w teście ADF. Ocenić jaki działanie i wnioskowanie z różnych postaci testów ADF w zależności od wartości parametrów procesu AR (1).

## Wstęp

Celem niniejszego zadania jest zbadanie działania różnych wersji testu ADF (Augmented Dickey-Fuller) w kontekście procesu autoregresyjnego rzędu pierwszego - AR(1). Test ADF jest jednym z podstawowych narzędzi służących do oceny stacjonarności szeregów czasowych, co stanowi kluczowy etap w analizie i modelowaniu danych czasowych.

## Przebieg badania

W ramach eksperymentu symulacyjnego analizowany jest model:

$$X_t = \phi_0 + \phi_1 X_{t-1} + \varepsilon_t$$

gdzie  $\varepsilon_t \sim iid(0, \sigma^2)$ , przy  $\sigma^2 = 1$ .

W badaniu użyto parametrów:

- $\phi_0 = 1$
- $\phi_1 = 0,9$

Dla podanych parametrów generowany jest zbiór 100 obserwacji z modelu AR(1). Aby zapewnić stacjonarność procesu pierwsza wartość procesu została obliczona jako  $X_0 = \frac{\phi_0}{1-\phi_1}$ , co odpowiada wartości oczekiwanej stacjonarnego modelu AR(1), a następnie wygenerowano dodatkowe 1000 obserwacji początkowych, które odrzucono przed analizą. Następnie przeprowadzane są trzy wersje testu ADF:

1. bez stałej,
2. ze stałą,
3. ze stałą i trendem liniowym.

Wartości p uzyskane z testów są zapisywane, a procedura powtarzana jest 1000 razy. Następnie obliczane są statystyki opisowe, wizualizowane rozkłady oraz wyznaczany jest odsetek przypadków, w których hipoteza zerowa testu została odrzucona.

Hipotezy testu ADF można sformułować następująco:

- $H_0: X_t \sim I(1)$ ,
- $H_1: X_t \sim I(0)$ .

Analiza uzyskanych wyników pozwala ocenić skuteczność i zachowanie testu ADF w różnych warunkach, w szczególności dla wartości  $\phi_1$  bliskich jedności, gdzie proces staje się niemal niestacjonarny. Ponadto umożliwia porównanie wpływu dołączenia stałej i trendu liniowego do modelu testowego na decyzje statystyczne testu.

## Użyte biblioteki

W badaniu wykorzystano funkcję **statsmodels.tsa.stattools.adfuller** do przeprowadzenia rozszerzonego testu pierwiastka jednostkowego Dickeya-Fullera. Rozszerzony test Dickeya-Fullera służy do badania stacjonarności szeregu czasowego poprzez testowanie obecności pierwiastka jednostkowego.

## Wyniki badania

$\Phi_0$	$\Phi_1$	Wersja ADF	Średnia wartość p	Medianą wartość p	Odch. std. Wartość p	Odsetek odrzuceń $H_0$ (%)
1	0,9	Bez stałej, bez trendu	0,5059	0,5089	0,1428	0,0
1	0,9	Ze stałą	0,147	0,0935	0,1629	35,5
1	0,9	Z trendem liniowym i stałą	0,2243	0,1614	0,2204	27,5

Tabela 1. Statystyki opisowe testu ADF dla procesu AR(1)

Tabela przedstawia wyniki symulacji testu Augmented Dickey-Fullera (ADF) dla procesu AR(1) o parametrach  $\Phi_0 = 1$  oraz  $\Phi_1 = 0,9$ , na podstawie 1000 prób o długości  $N = 100$ . Ujęto trzy wersje testu: bez stałej i trendu, ze stałą oraz ze stałą i trendem liniowym. Zawarte są wartości średnie, mediany oraz odchylenia standardowego wartości p, a także odsetek odrzuceń hipotezy zerowej o niestacjonarności.

W przedstawionej tabeli średnie wartości p w wersji testu bez stałej i trendu są wysokie (0,5059), test błędnie uznaje proces za niestacjonarny. Po dodaniu stałej średnia wartość p znaczco spada (0,147), co skutkuje odrzuceniem  $H_0$  w 35,5% przypadków. W wersji ze stałą i trendem liniowym wartości p są nieco wyższe (0,2243), a liczba odrzuceń  $H_0$  spada do 27,5%.

Medianą wartości p pokazuje środkową wartość rozkładu wyników testu. Dla wersji bez stałej i trendu (0,5089) większość wartości p jest wysoka, więc test nie

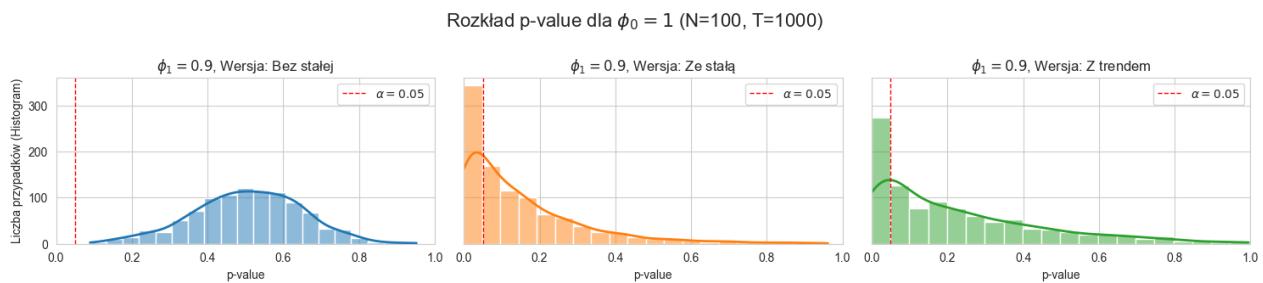
odrzuca  $H_0$ . Po dodaniu stałej i trendu mediany spadają (kolejno 0,0935 i 0,1614), co oznacza większą skłonność testu do wykrywania stacjonarności.

Odchylenie standardowe wartości p pokazuje, jak bardzo wyniki testu różniły się między symulacjami. Większe wartości (0,1629-0,2204) przy testach ze stałą lub trendem wskazują na większą zmienność wyników, czyli test staje się bardziej wrażliwy na losowe fluktuacje danych.

Dla wartości  $\phi_1$  bliskich jedności pojawia się trudność w odróżnieniu procesów stacjonarnych od niestacjonarnych, ponieważ szereg zachowuje się podobnie do błędzenia losowego (tzw. proces z pierwiastkiem jednostkowym). Warunek stacjonarności dla modelu AR(1) to:

$$|\phi_1| < 1$$

Jeżeli  $|\phi_1| \geq 1$ , proces staje się niestacjonarny, co powoduje, że test ADF ma mniejszą moc i częściej nie odrzuca hipotezy zerowej nawet przy pozornie stacjonarnych danych.



Na wykresach przedstawiono rozkłady wartości p z testu Dickeya-Fullera dla 1000 symulacji (N=100 obserwacji każda) z parametrem  $\phi_1 = 0,9$  i  $\phi_0 = 1$ . Czerwona przerywana linia oznacza poziom istotności  $\alpha = 0,05$ .

- **Bez stałej**

Rozkład jest względnie równomierny z lekkim skupieniem w zakresie średnich wartości (0.4-0.6). Test rzadko odrzuca hipotezę zerową, traktując proces jako niestacjonarny.

- **Ze stałą**

Zdecydowana większość wartości p koncentruje się blisko zera. Test często odrzuca hipotezę zerową, poprawnie identyfikując proces jako stacjonarny.

- **Z trendem**

Podobnie jak w wersji ze stałą, rozkład jest silnie skośny w prawo z dominacją bardzo małych wartości p. Test często odrzuca  $H_0$ , wykrywając stacjonarność. Włączenie trendu liniowego powoduje zmniejszenie odsetka odrzuceń  $H_0$ .

Wersja ze stałą okazała się najbardziej odpowiednia dla procesu z  $\phi_0 = 1$  i  $\phi_1 = 0.9$ . Pominięcie ważnych elementów, jak stała, sprawia, że test traci swoją skuteczność,

natomiąst dodanie dodatkowych elementów, jak trend, nie jest najlepszym rozwiązaniem i nie przynosi korzyści.