

Rozdział 3. Stacjonarność

MODELOWANIE POLSKIEJ GOSPODARKI z R

Silna stacjonarność

O procesie stochastycznym $\{Y_t\}$ mówimy, że jest **stacjonarny w ścisłym sensie** (silna stacjonarność), jeżeli łączny rozkład prawdopodobieństwa dla $\{Y_{t_1}, Y_{t_2}, \dots, Y_{t_n}\}$ jest taki sam jak dla $\{Y_{t_1+m}, Y_{t_2+m}, \dots, Y_{t_n+m}\}$ dla dowolnych obserwacji t_1, t_2, \dots, t_n oraz przesunięcia w czasie opisywanego parametrem m .

Przykładem procesu stacjonarnego w ścisłym sensie jest np. wielokrotne losowanie z niezależnego rozkładu $N(0, 1)$.

Silnej stacjonarności nie można zweryfikować empirycznie

Słaba stacjonarność

Empirycznie można testować **stacjonarność w szerszym sensie** (słaba stacjonarność), która występuje gdy wartość oczekiwana, wariancja oraz kowariancja procesu są stałe w czasie:

$$\mu_t = \mu$$

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

$$\gamma_{s,t} = \gamma_s \text{ dla } s = 1, 2, \dots$$

Stopień zintegrowania

Schemat wnioskowania na temat stopnia zintegrowania zmiennej y_t :

- 1 Test dla stacjonarności y_t . W przypadku stacjonarności: $y_t \sim I(0)$. W przypadku niestacjonarności, analizujemy stacjonarność Δy_t , itd.
- 2 Jeżeli dla $k = 1, 2, \dots, d - 1$ zmienna $\Delta^k y_t$ jest niestacjonarna, zaś zmienna $\Delta^d y_t$ jest stacjonarna, to o szeregu y_t mówimy, że jest niestacjonarny, zintegrowany w stopniu d : $y_t \sim I(d)$.

Procesy białego szumu i błędzenia losowego

Biały szum (white noise, WN):

$\{Y_t\}$ jest białym szumem, jeżeli dla dowolnych $s \neq t$, Y_t i Y_s są względem siebie niezależne, mają jednakowy rozkład prawdopodobieństwa oraz ich wartości oczekiwane są równa zero.

Błądzenie losowe (random walk, RW):

$\{X_t\}$ jest błędzeniem losowym, jeżeli:

$$X_t = X_{t-1} + Y_t, Y_t \sim WN$$

Procesy białego szumu i błędzenia losowego

Ponieważ:

$$X_t = \sum_{s=1}^t Y_s + X_0.$$

To:

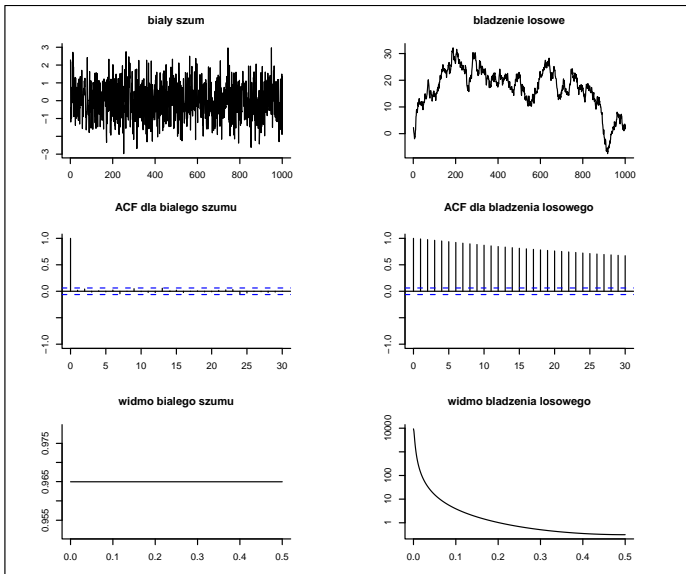
$$E(X_t) = X_0$$

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

$$\text{Cov}(X_t, X_{t+s}) = t\sigma^2$$

Wariancja oraz kowariancja zależy od momentu obserwacji t , proces błędzenia losowego jest procesem niestacjonarnym.

Procesy WN i RW - przykład



Wariancja długookresowa

Wzór:

$$\gamma_{\infty} = \gamma_0 + 2 \sum_{s=1}^{\infty} \gamma_s.$$

Uwaga: wariancja długookresowa jest równa wartości widma w zerze pomnożonej przez 2π

Zastosowanie: wnioskowanie na temat istotności zmiennych (m.in. w modelu liniowym lub w testach pierwiastka jednostkowego) w przypadku występowania autokorelacji

Wariancja długookresowa - ilustracja

Niech Y_t będzie stacjonarnym procesem o zerowej autokorelacji, wartości oczekiwanej μ oraz wariancji γ_0 . Dla statystyki $\bar{Y} = \frac{1}{T} \sum_{t=1}^T Y_t$:

$$E(\bar{Y}) = \mu$$
$$\text{Var}(\bar{Y}) = \frac{\gamma_0}{T}$$

W przypadku autokorelacji Y_t :

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

gdzie $\gamma_k = \text{Cov}(Y_t, Y_{t-k})$.

Wariancja długookresowa jest granicą powyższego wyrażenia dla $T \rightarrow \infty$

Wariancja długookresowa - estymacja

Estymacja długookresowej wariancji na podstawie obserwacji $\{y_t : t = 1, 2, \dots, T\}$ dokonywana jest najczęściej metodami nieparametrycznymi wg. wzoru:

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

gdzie $\kappa_{s,L} : 0 < s \leq L$ określa okno opóźnień (lag window), zaś L jest maksymalnym opóźnieniem.

Newley i West proponują stosowanie okna Bartletta, dla którego:

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

oraz maksymalnego opóźnienia:

$$L = 4(T/100)^{2/9}$$

Test Dickeya-Fullera

Najpopularniejszy test stacjonarności szeregu czasowego o zespole hipotez:

H_0 : y_t jest niestacjonarny

H_1 : y_t jest stacjonarny,

Test Dickeya-Fullera

Idea testu to spostrzeżenie, że proces $y_t = \rho y_{t-1} + \epsilon_t$ jest stacjonarny gdy $|\rho| < 1$.

Schemat postępowania:

- 1 Oszacuj MNK parametry modelu:

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

- 2 Dokonaj weryfikacji zespołu hipotez:

$$H_0 : \delta = 0$$

$$H_1 : \delta < 0$$

na podstawie statystyki:

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

która przy prawdziwości H_0 ma rozkład testu DF.

Test Dickeya-Fullera

Dla niestacjonarnych procesów $AR(P)$ stosuje się rozszerzony test DF (Augmented Dickey-Fuller, ADF):

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

Dla szeregów o wartości oczekiwanej różnej od zera ($E(y_t) \neq 0$):

$$\Delta y_t = \mu + \delta y_{t-1} + \sum_{p=1}^P \rho_p \Delta y_{t-p} + \epsilon_t.$$

Jeżeli H_1 wskazuje na trendostacjonarność to:

$$\Delta y_t = \mu_0 + \mu_1 t + \delta y_{t-1} + \sum_{p=1}^P \rho_p \Delta y_{t-p} + \epsilon_t.$$

Półokres wygasania

- Dla stacjonarnego procesu $(y_t - \mu) = \rho(y_{t-1} - \mu) + \epsilon_t$, tj. jeżeli $|\rho| < 1$, y_t powraca do swojej wartości oczekiwanej μ
- Tempo powrotu jest opisane przez parametr $\delta = \rho - 1$ (np. dla $\rho = 0.95$ i $\delta = -0.05$ odchylenie od średniej wygasa w tempie 5% w ciągu jednego okresu)
- Półokres wygasania (half-life, hl) - określa czas wygaśnięcia połowy odchylenia zmiennej od poziomu równowagi:

$$\rho^{hl} = \frac{1}{2} \Leftrightarrow hl = \frac{\ln 0,5}{\ln \rho}$$

- Dla $\rho = 0.95$ wartość hl wynosi 13,5 okresu.

Test ADF - Przykład

Wyniki testu ADF dla logarytmu PKB:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	1.5936617	0.8128333	1.961	0.0552	.
z.lag.1	-0.1330983	0.0684871	-1.943	0.0573	.
tt	0.0013020	0.0007067	1.842	0.0710	.
z.diff.lag1	-0.1078035	0.1347278	-0.800	0.4272	
z.diff.lag2	0.0632721	0.1324959	0.478	0.6349	

Wartosc statystyki testu ADF: -1.9434

Wartosci krytyczne

	1pct	5pct	10pct
ADF*	-4.04	-3.45	-3.15

Test ADF - Przykład

Wyniki testu ADF dla tempa wzrostu PKB:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	0.010159	0.003143	3.232	0.002114	**
z.lag.1	-0.976116	0.253240	-3.855	0.000315	***
z.diff.lag1	-0.198996	0.202951	-0.981	0.331288	
z.diff.lag2	-0.148250	0.133373	-1.112	0.271350	

Wartosc statystyki testu ADF: -3.8545

Wartosci krytyczne

	1pct	5pct	10pct
ADF*	-3.51	-2.89	-2.58

Test Phillipsa-Perrona

- Nieparametryczna korekta testu DF o autokorelację lub heteroskedastyczność składnika losowego ϵ_t regresji testu DF
- Wykorzystanie informacji o długookresowej wariancji składnika losowego ϵ_t
- W teście ADF korekta ma charakter parametryczny i polega na dodaniu opóźnień zmiennej Δy_t do regresji testu DF

Test Phillipsa-Perrona - algorytm postępowania

- 1 Oszacuj parametry regresji:

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

- 2 Oblicz reszty e_t , średni błąd szacunku S_ρ , oszacowanie wariancji składnika losowego $\hat{\sigma}^2$, autokowariancji $\{\hat{\gamma}_j : j = 0, 1, \dots, T - 1\}$ oraz długookresowej wariancji $\hat{\gamma}_\infty$ (metodą Newley'a i Westa)
- 3 Oblicz statystykę testu PP:

$$Z_\tau = \left(\frac{\hat{\gamma}_0}{\hat{\gamma}_\infty} \right)^{1/2} \left(\frac{\hat{\rho} - 1}{S_\rho} \right) - \frac{T}{2} (\hat{\gamma}_\infty - \hat{\gamma}_0) \left(\frac{S_\rho}{\sqrt{\hat{\sigma}^2 \hat{\gamma}_\infty}} \right).$$

która ma rozkład testu PP.

Test PP - Przykład

Wyniki testu PP dla poziomu PKB:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	1.8020955	0.7462397	2.415	0.0190	*
y.l1	0.8529679	0.0612579	13.924	<2e-16	***
trend	0.0014377	0.0006377	2.255	0.0280	*

Wartosc statystyki testu PP (Z-tau): -2.4173

Wartosci krytyczne:

	1pct	5pct	10pct
PP*	-4.11	-3.48	-3.17

Test PP - Przykład

Wyniki testu PP dla tempa wzrostu PKB:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	0.012443	0.002062	6.034	0.000000126	***
y.l1	-0.148614	0.130845	-1.136	0.261	

Wartosc statystyki testu PP (Z-tau): -8.7272

Wartosci krytyczne:

	1pct	5pct	10pct
PP*	-3.54	-2.91	-2.59

Test KPSS

Zespół hipotez testu KPSS:

H_0 : y_t jest stacjonarny

H_1 : y_t jest niestacjonarny

Idea testu:

$$y_t = x_t + z_t$$

$$x_t = x_{t-1} + v_t, \quad v_t \sim WN(0, \sigma_v^2)$$

$$z_t = \mu_0 + \epsilon_t,$$

Przy prawdziwości H_0 wariancja σ_v^2 jest zerowa, zaś wartości zmiennej x_t są stałe w czasie.

Test KPSS - schemat postępowania

- 1 Oblicz reszty e_t regresji y_t względem stałej (oraz trendu)
- 2 Oblicz sumy $S_t = \sum_{j=1}^t e_j$ dla $t = 1, 2, \dots, T$
- 3 Metodą Newley'a-Westa oblicz długookresową wariancję reszt $\hat{\gamma}_\infty$
- 4 Obliczyć statystykę:

$$KPSS = \frac{1}{T^2} \frac{\sum_{t=1}^T S_t^2}{\hat{\gamma}_\infty},$$

która ma rozkład testu KPSS.

Test KPSS - Przykład

Wyniki testu KPSS dla poziomu PKB:

Wartosc statystyki testu KPSS (z trendem): 0.1629

Wartosci krytyczne:

	10pct	5pct	2.5pct	1pct
KPSS*	0.119	0.146	0.176	0.216

Wyniki testu KPSS dla tempa wzrostu PKB:

Wartosc statystyki testu KPSS (bez trendu): 0.1786

Wartosci krytyczne:

	10pct	5pct	2.5pct	1pct
KPSS*	0.347	0.463	0.574	0.739

Moc testów pierwiastka jednostkowego

Moc testów ADF i PP jest niska, zaś test KPSS jest podatny na zaburzenie rozmiaru.

W celu ilustracji problemu przeprowadzimy symulacje Monte Carlo:

- Wygenerujemy $N = 100$ realizacji procesu stacjonarnego:

$$y_t = (1 - \rho) + \rho y_t - 1 + \epsilon_t, \quad \epsilon_t \sim N(0, 1)$$

dla różnych wartości parametru $\rho \in \{0, 90, 0, 93, 0, 96, 0, 99\}$

- Na podstawie realizacji procesu $\{y_t : t = 1, 2, \dots, T\}$ dokonamy analizy stacjonarności szeregu procesu y_t wykorzystując testy omówione w poprzednich punktach rozdziału
- Długość próby: $T = 64$, co odpowiada liczbie kwartałów z okresu 1995–2010r.

Moc testów pierwiastka jednostkowego

Wyniki symulacji Monte Carlo (T=64, N=100)

Liczba przypadków, w których test wskazuje na stacjonarnosc procesu:

	rho=0.9	rho=0.93	rho=0.96	rho=0.99
ADF	24	17	9	3
PP	28	22	12	5
KPSS	44	29	21	6