

ACF, PACF, badanie stacjonarności

Stanisław Cichocki
Natalia Nehrebecka

Wykład 3

Plan wykładu

- ▶ 1. Regresja pozorna
- ▶ 2. Funkcje ACF i PACF
- ▶ 3. Badanie stacjonarności
 - Test Dickey-Fullera (DF)
 - Rozszerzony test Dickey-Fullera (ADF)
 - Test KPSS

Plan wykładu

- ▶ 1. Regresja pozorna
- ▶ 2. Funkcje ACF i PACF
- ▶ 3. Badanie stacjonarności
 - Test Dickey-Fullera (DF)
 - Rozszerzony test Dickey-Fullera (ADF)
 - Test KPSS

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:

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

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

$$\varepsilon_t \sim N(0,1)$$

$$u_t \sim N(0,1)$$

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

Plan wykładu

- ▶ 1. Regresja pozorna
- ▶ 2. Funkcje ACF i PACF
- ▶ 3. Badanie stacjonarności
 - Test Dickey-Fullera (DF)
 - Rozszerzony test Dickey-Fullera (ADF)
 - Test KPSS

Funkcja ACF

- ▶ Funkcja autokorelacji (Autocorrelation Function) to współczynnik korelacji między dwoma realizacjami y_t oddalonymi w czasie o k okresów.

$$\rho_k = \frac{\text{Cov}(y_t, y_{t-k})}{\text{Var}(y_t)}$$

$$\rho \in [-1,1]$$

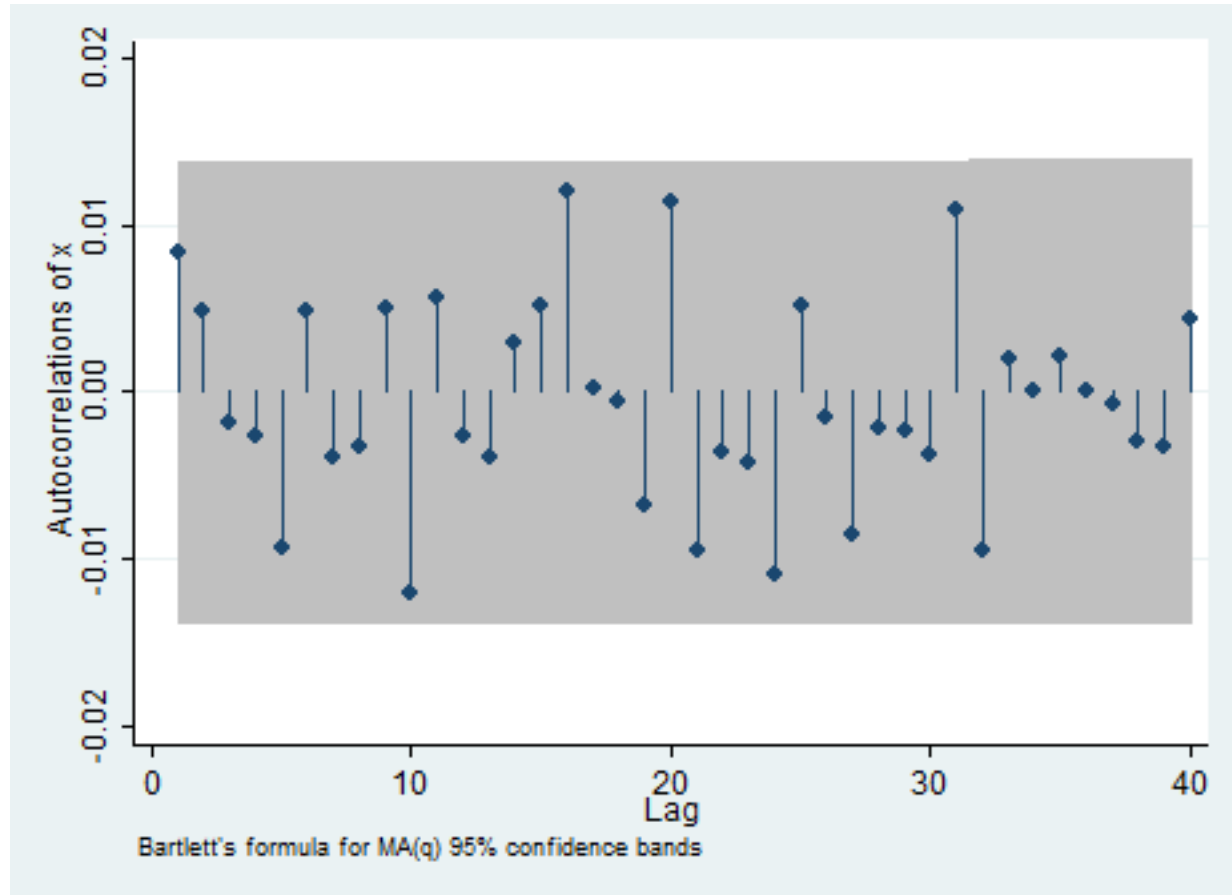
Funkcja PACF

- ▶ Funkcja autokorelacji cząstkowej (Partial Autocorrelation Function) mierzy korelację między obserwacjami y_t oddalonymi od siebie o k okresów bez uwzględnienia wpływu $y_{t-k-1}, y_{t-k-2}, \dots, y_{t-1}$
- ▶ Funkcja ta jest równa wyestymowanemu współczynnikowi α_k w modelu autoregresyjnym k tego rzędu:

$$y_t = \mu + \alpha_1 y_{t-1} + \dots + \alpha_k y_{t-k} + \varepsilon_t$$

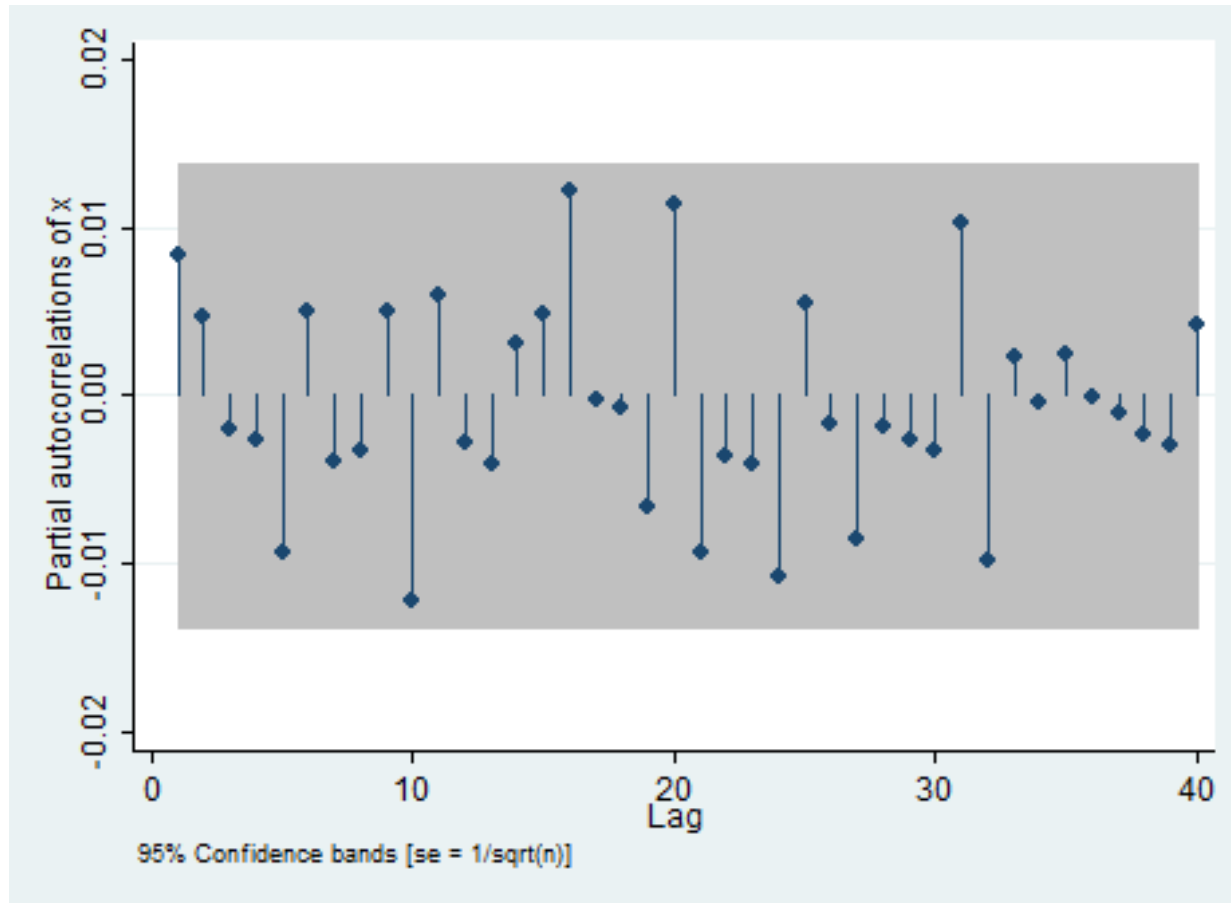
Funkcja ACF

- ▶ ACF dla białego szumu



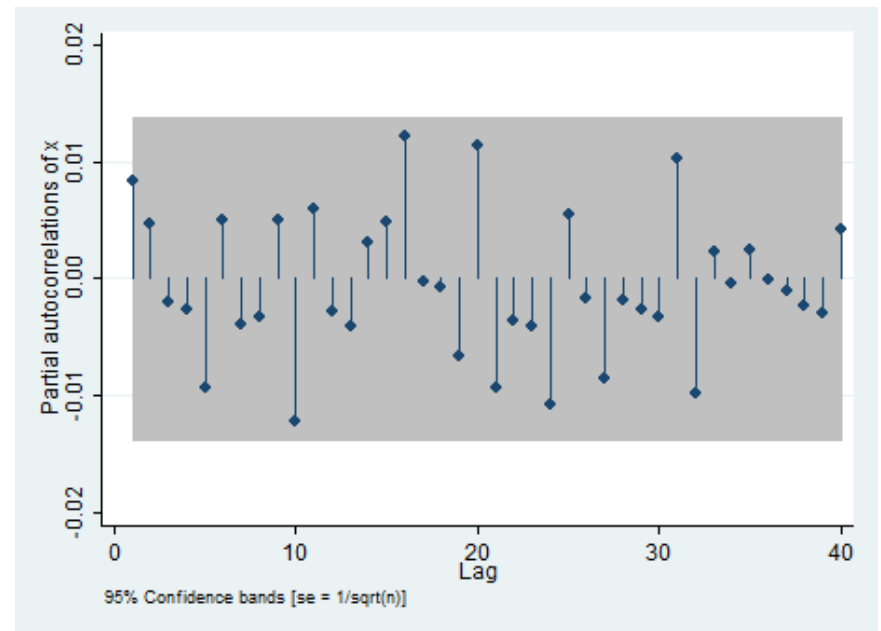
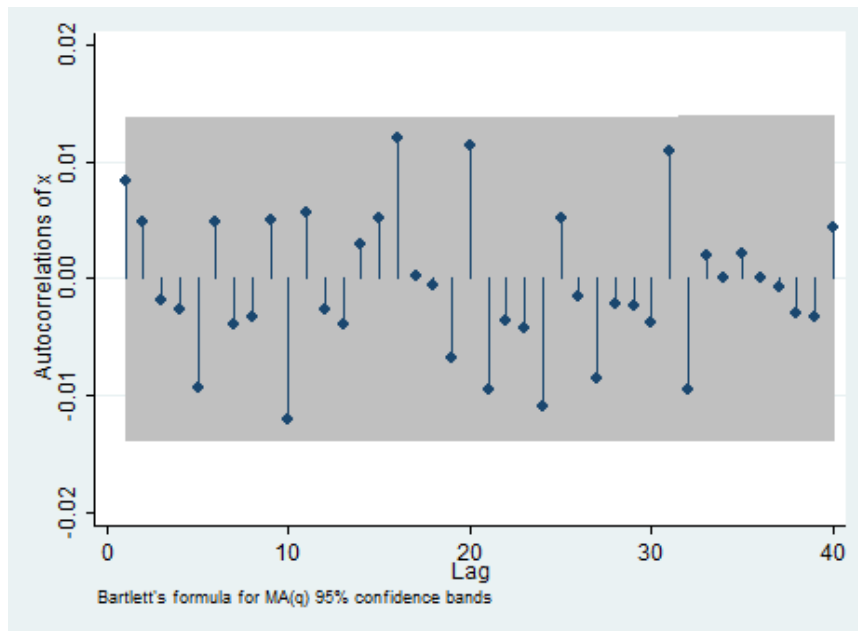
Funkcja PACF

- ▶ PACF dla białego szumu



Funkcja ACF i PACF

- ▶ ACF i PACF dla białego szumu



Funkcja ACF i PACF

LAG	AC	PAC	Q	Prob>Q	-1	0	1	-1	0	1
					[Autocorrelation]			[Partial Autocor]		
1	0.0083	0.0083	1.3868	0.2389						
2	0.0048	0.0047	1.8458	0.3974						
3	-0.0019	-0.0020	1.9194	0.5893						
4	-0.0027	-0.0027	2.068	0.7233						
5	-0.0094	-0.0093	3.828	0.5744						
6	0.0048	0.0050	4.2914	0.6373						
7	-0.0039	-0.0039	4.599	0.7088						
8	-0.0033	-0.0034	4.823	0.7763						
9	0.0050	0.0050	5.3186	0.8057						
10	-0.0120	-0.0122	8.2129	0.6081						

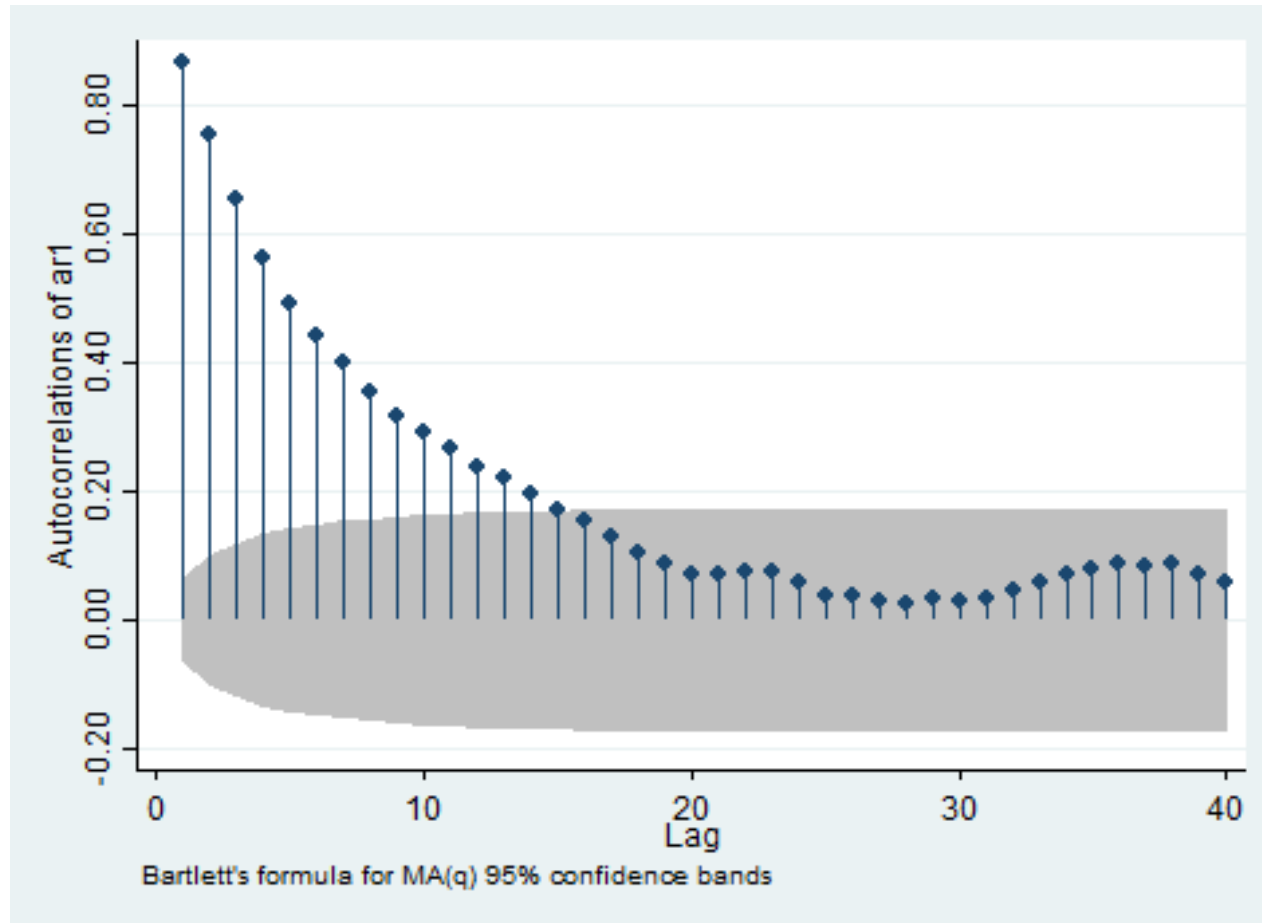
Statystyka Q Ljunga-Boxa:

$$Q = T(T + 2) \sum_{k=1}^m \frac{\hat{\rho}_k^2}{T - k} \sim \chi^2_m$$

H_0 proces jest białym szumem

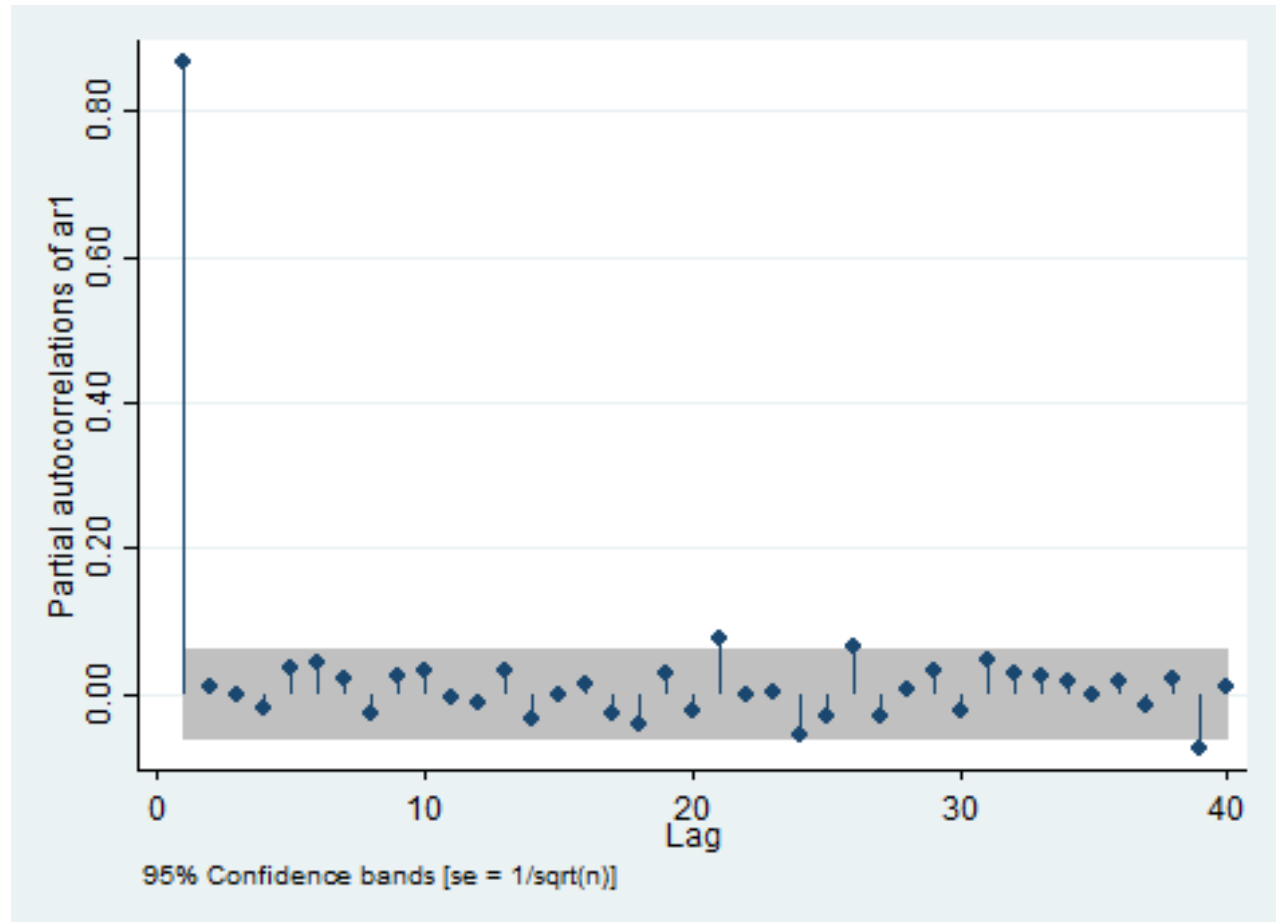
Funkcja ACF

- ▶ ACF dla AR(1) gdy $|\alpha| < 1$



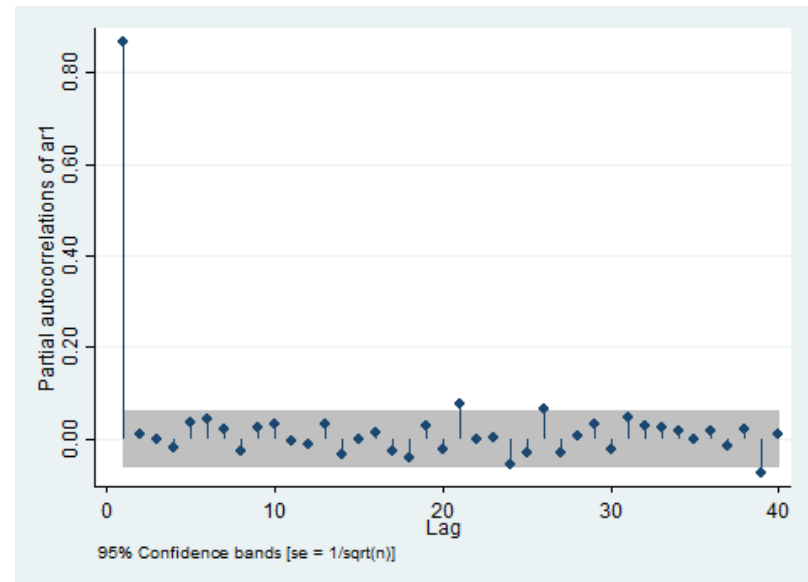
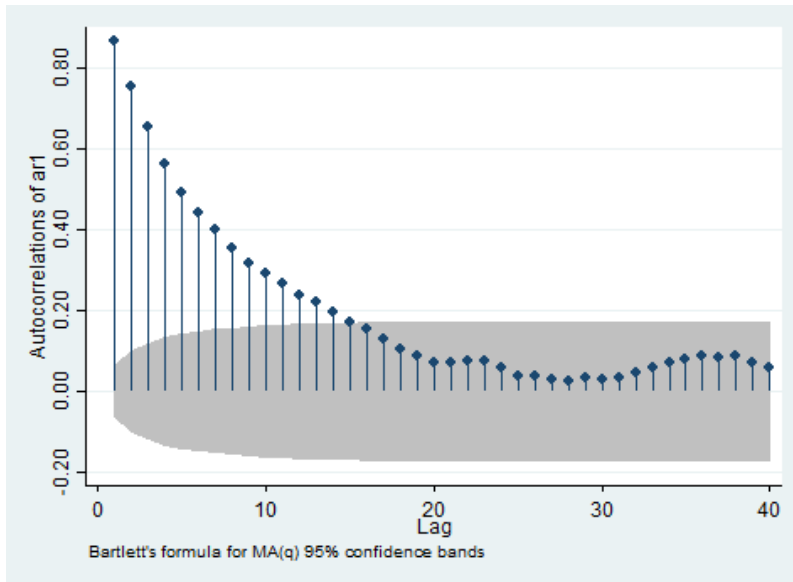
Funkcja PACF

- ▶ PACF dla AR(1) gdy $|\alpha| < 1$



Funkcja ACF i PACF

- ▶ ACF i PACF dla AR(1) gdy $|\alpha| < 1$

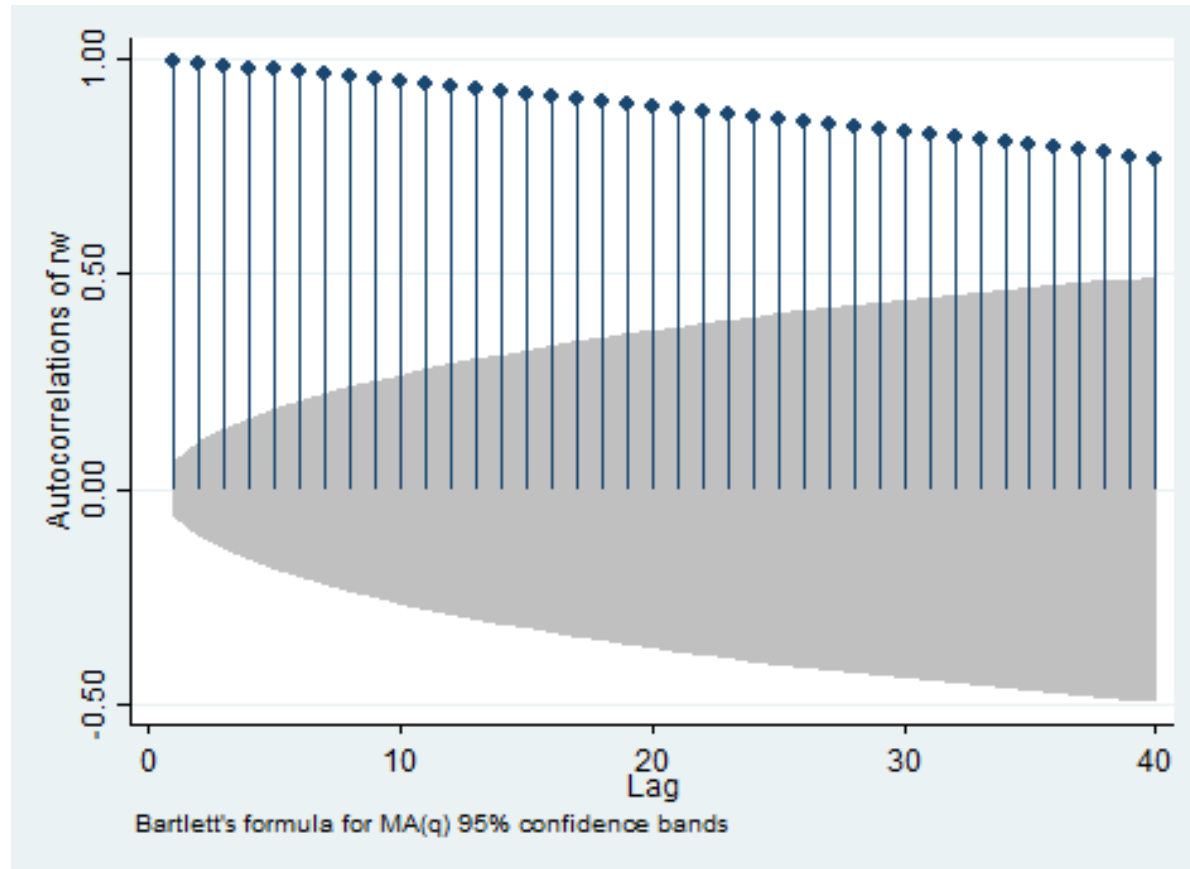


Funkcja ACF i PACF

LAG	AC	PAC	Q	Prob>Q	-1	0	1	-1	0	1	
					Prob>Q	[Autocorrelation]			[Partial Autocor]		
1	0.8668	0.8683	753.52	0.0000		-----		-----			
2	0.7536	0.0114	1323.6	0.0000		-----					
3	0.6538	-0.0012	1753.2	0.0000		-----					
4	0.5624	-0.0212	2071.4	0.0000		-----					
5	0.4921	0.0349	2315.3	0.0000		---					
6	0.4400	0.0419	2510.5	0.0000		---					
7	0.3974	0.0219	2669.9	0.0000		---					
8	0.3519	-0.0275	2795	0.0000		--					
9	0.3178	0.0228	2897.1	0.0000		--					
10	0.2931	0.0307	2984.1	0.0000		--					

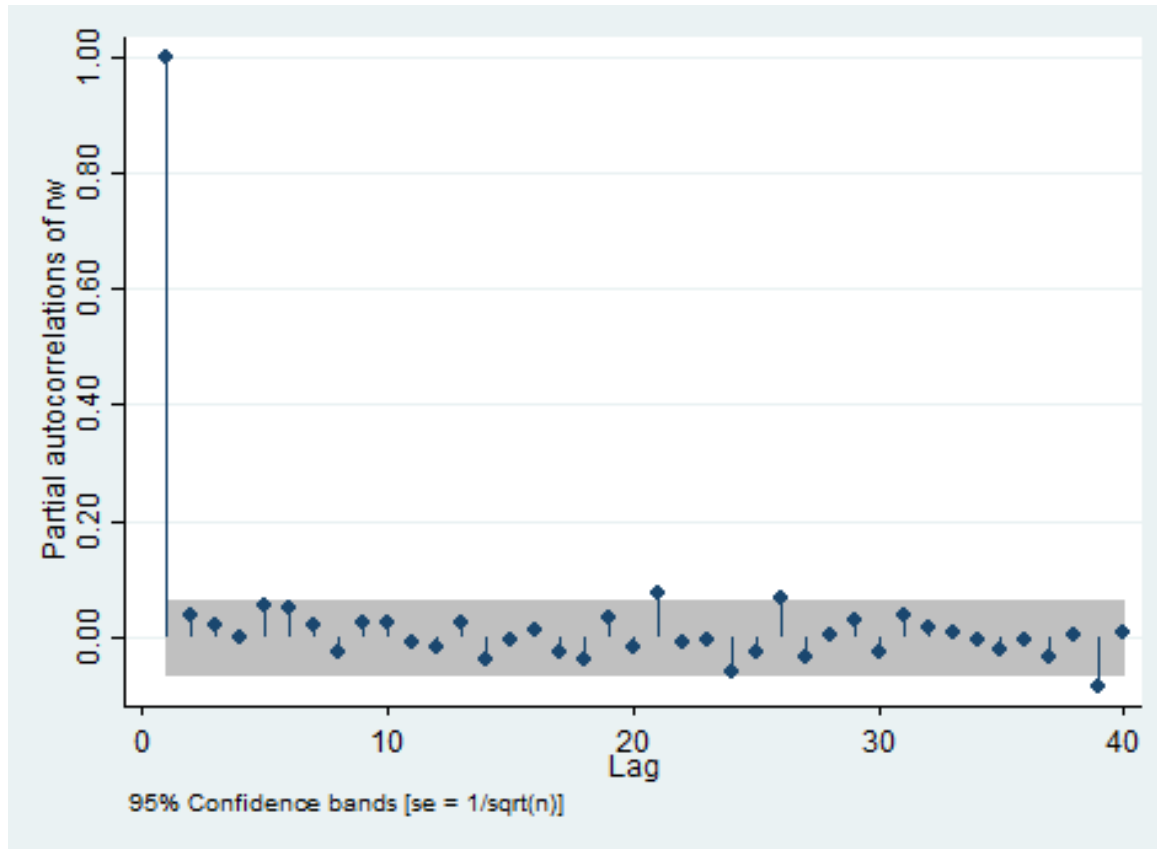
Funkcja ACF

- ▶ ACF dla błędzenia przypadkowego



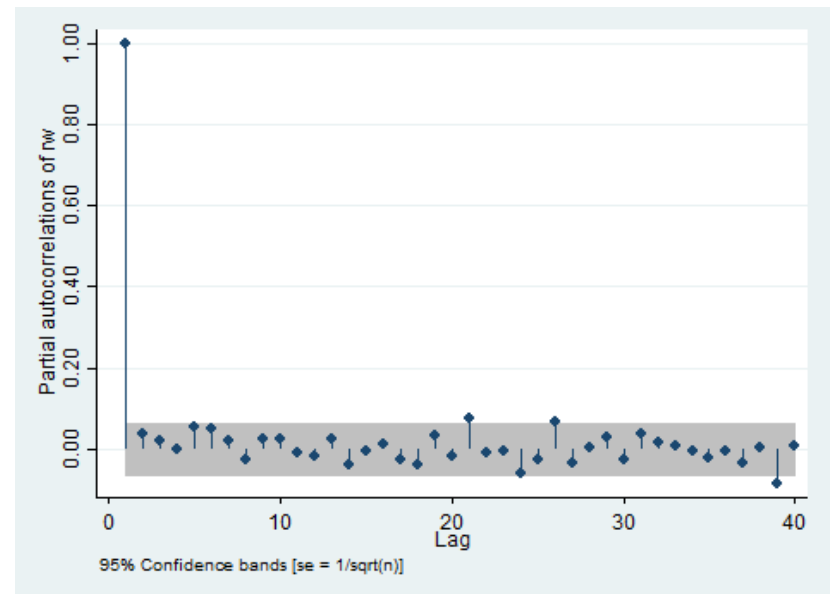
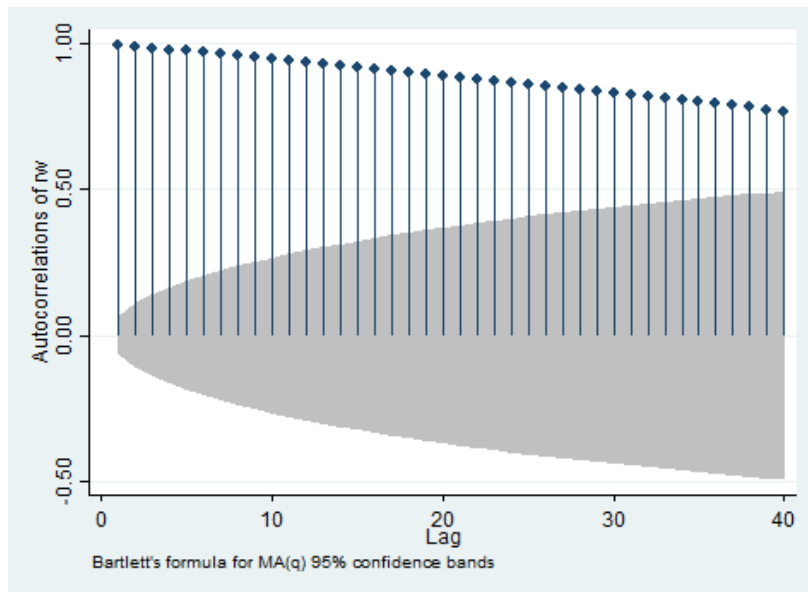
Funkcja PACF

- ▶ PACF dla błędzenia przypadkowego



Funkcja ACF i PACF

- ▶ ACF i PACF dla błędzenia przypadkowego



Funkcja ACF i PACF

LAG	AC	PAC	Q	Prob>Q	-1 0 1 [Autocorrelation]	-1 0 1 [Partial Autocor]
1	0.9941	0.9975	991.19	0.0000	-----	-----
2	0.9884	0.0369	1972	0.0000	-----	
3	0.9829	0.0208	2943	0.0000	-----	
4	0.9771	0.0006	3903.5	0.0000	-----	
5	0.9717	0.0528	4854.3	0.0000	-----	
6	0.9664	0.0503	5795.8	0.0000	-----	
7	0.9614	0.0228	6728.5	0.0000	-----	
8	0.9562	-0.0262	7652	0.0000	-----	
9	0.9508	0.0242	8566	0.0000	-----	
10	0.9456	0.0264	9470.9	0.0000	-----	

Plan wykładu

- ▶ 1. Regresja pozorna
- ▶ 2. Funkcje ACF i PACF
- ▶ 3. Badanie stacjonarności
 - Test Dickey-Fullera (DF)
 - Rozszerzony test Dickey-Fullera (ADF)
 - Test KPSS

Test Dickey-Fullera (DF)

- ▶ Najwcześniejszy i najpopularniejszy test, za pomocą którego badamy czy zmienna jest stacjonarna.
- ▶ Mamy model:

$$y_t = \beta y_{t-1} + \varepsilon_t$$
$$\varepsilon_t \sim IID(0, \sigma^2)$$

$H_0 : \beta = 1$ - y_t błądzenie przypadkowe (zmienna niestacjonarna)

$H_1 : |\beta| < 1$ - y_t jest procesem AR(1) (zmienna stacjonarna)

Test Dickey-Fullera (DF)

- ▶ Odejmując od obu stron y_{t-1} :

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

$$\Delta y_t = (\beta - 1) y_{t-1} + \varepsilon_t$$

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

$H_0 : \rho = 0$ - y_t jest niestacjonarne

$H_0 : \rho \in (-2, 0)$ - y_t jest stacjonarne

Test Dickey-Fullera (DF)

- ▶ Problem: nie można używać statystyki t do testowania istotności parametru ρ ponieważ rozkłady statystyk testowych są niestandardowe jeśli w modelu zmienne niestacjonarne.
- ▶ Specjalne tablice z wartościami krytycznymi dla testu DF.
- ▶ Uwaga techniczna: wielkości krytyczne rozkładu statystyki DF są zawsze ujemne.

Test Dickey-Fullera (DF)

- ▶ Test DF przeprowadzamy w następujący sposób:
 1. regresja Δy_t na y_{t-1}
 2. porównujemy statystykę t dla y_{t-1} z wartościami krytycznymi testu DF:
 - a) wartość statystyki jest mniejsza od wartości krytycznej - odrzucamy H_0 o niestacjonarności y_t i przyjmujemy H_1 o stacjonarności y_t ;
 - b) wartość statystyki jest większa od wartości krytycznej – nie ma podstaw do odrzucenia H_0

Test Dickey-Fullera (DF)

- ▶ Test DF dla białego szumu:

Dickey-Fuller test for unit root

Number of obs = 19999

----- Interpolated Dickey-Fuller -----						
Test	1% Critical		5% Critical		10% Critical	
Statistic	Value		Value		Value	
Z(t)	-140.235	-2.580	-1.950	-1.620		

D.x	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	

x						
L1.	-.9916464	.0070713	-140.24	0.000	-1.005507	-.9777861

Test Dickey-Fullera (DF)

- ▶ Test DF dla białego szumu:

Dickey-Fuller test for unit root

Number of obs = 19999

	Test Statistic	1% Critical Value	5% Critical Value	10% Critical Value
z(t)	-140.235	-2.580	-1.950	-1.620

D.x	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
x					
L1.	-.9916464	.0070713	-140.24	0.000	-1.005507 - .9777861

Test Dickey-Fullera (DF)

- ▶ Test DF dla białego szumu:

Source	SS	df	MS			
Model	19948.5217	1	19948.5217	Number of obs =	19999	
Residual	20285.4259	19998	1.01437273	F(1, 19998) =	19665.87	
Total	40233.9476	19999	2.01179797	Prob > F =	0.0000	
				R-squared =	0.4958	
				Adj R-squared =	0.4958	
				Root MSE =	1.0072	

D.x	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
x						
L1.	-.9916464	.0070713	-140.24	0.000	-1.005507	-.9777861

Test Dickey-Fullera (DF)

- ▶ Test DF dla białego szumu:

Source	SS	df	MS	Number of obs	=	19999
Model	19948.5217	1	19948.5217	F(1, 19998)	=	19665.87
Residual	20285.4259	19998	1.01437273	Prob > F	=	0.0000
Total	40233.9476	19999	2.01179797	R-squared	=	0.4958
				Adj R-squared	=	0.4958
				Root MSE	=	1.0072

D.x	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
x					
L1.	-.9916464	.0070713	-140.24	0.000	-1.005507 - .9777861

Test Dickey-Fullera (DF)

Breusch-Godfrey LM test for autocorrelation

lags (p)	chi2	df	Prob > chi2
1	1.699	1	0.1925
2	1.699	2	0.4276
3	1.768	3	0.6220
4	1.896	4	0.7549

H0: no serial correlation

Test Dickey-Fullera (DF)

- ▶ Test DF dla błędzenia przypadkowego:

Dickey-Fuller test for unit root Number of obs = 999

		----- Interpolated Dickey-Fuller -----				
Test		1% Critical	5% Critical	10% Critical		
Statistic		Value	Value	Value		
Z(t)		-0.277	-2.580	-1.950	-1.620	

D.rw		Coef.	Std. Err.	t	P> t	[95% Conf. Interval]

rw						
L1.		-.0005213	.001879	-0.28	0.782	-.0042086 .003166

Test Dickey-Fullera (DF)

- ▶ Test DF dla błędzenia przypadkowego:

Dickey-Fuller test for unit root

Number of obs = 999

----- Interpolated Dickey-Fuller -----						
	Test		1% Critical	5% Critical	10% Critical	
	Statistic		Value	Value	Value	
Z(t)	-0.277		-2.580	-1.950	-1.620	
D.rw	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
rw						
L1.	-.0005213	.001879	-0.28	0.782	-.0042086	.003166

Test Dickey-Fullera (DF)

- ▶ Test DF dla błędzenia przypadkowego

Source	SS	df	MS	Number of obs = 999		
Model	.077155196	1	.077155196	F(1, 998) =	0.08	
Residual	1000.45476	998	1.00245968	Prob > F =	0.7815	
Total	1000.53191	999	1.00153345	R-squared =	0.0001	
				Adj R-squared =	-0.0009	
				Root MSE =	1.0012	

D.rw	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
rw						
L1.	-.0005213	.001879	-0.28	0.782	-.0042086	.003166

Test Dickey-Fullera (DF)

- ▶ Test DF dla błędzenia przypadkowego

Source	SS	df	MS	Number of obs = 999		
Model	.077155196	1	.077155196	F(1, 998) =	0.08	
Residual	1000.45476	998	1.00245968	Prob > F	= 0.7815	
Total	1000.53191	999	1.00153345	R-squared	= 0.0001	
				Adj R-squared	= -0.0009	
				Root MSE	= 1.0012	

D.rw	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
rw						
L1.	-.0005213	.001879	-0.28	0.782	-.0042086	.003166

Test Dickey-Fullera (DF)

Breusch-Godfrey LM test for autocorrelation

lags (p)	chi2	df	Prob > chi2
1	1.084	1	0.2977
2	1.477	2	0.4779
3	2.629	3	0.4524
4	2.639	4	0.6199

H0: no serial correlation

Test Dickey-Fullera (DF)

- ▶ Test DF dla zróżnicowanego błędzenia przypadkowego:

Dickey-Fuller test for unit root Number of obs = 998

		----- Interpolated Dickey-Fuller -----			
Test		1% Critical	5% Critical	10% Critical	
Statistic		Value	Value	Value	

Z(t)	-30.573	-2.580	-1.950	-1.620	

D.drw	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]

drw					
L1.	-.9675424	.0316471	-30.57	0.000	-1.029645 - .9054398

Test Dickey-Fullera (DF)

- ▶ Test DF dla zróżnicowanego błędzenia przypadkowego:

Dickey-Fuller test for unit root Number of obs = 998

	Test Statistic	1% Critical Value	5% Critical Value	10% Critical Value
Z(t)	-30.573	-2.580	-1.950	-1.620

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
D.drw					
drw					
L1.	-.9675424	.0316471	-30.57	0.000	-1.029645 - .9054398

Test Dickey-Fullera (DF)

- ▶ Test DF dla zróżnicowanego błędzenia przypadkowego

Source	SS	df	MS			
Model	936.633212	1	936.633212	Number of obs =	998	
Residual	999.063052	997	1.00206926	F(1, 997) =	934.70	
Total	1935.69626	998	1.93957541	Prob > F =	0.0000	
				R-squared =	0.4839	
				Adj R-squared =	0.4834	
				Root MSE =	1.001	

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
D.drw						
rw						
LD.	-.9675424	.0316471	-30.57	0.000	-1.029645	-.9054398

Test Dickey-Fullera (DF)

Breusch-Godfrey LM test for autocorrelation

lags (p)	chi2	df	Prob > chi2
1	0.015	1	0.9032
2	1.874	2	0.3919
3	3.245	3	0.3554
4	3.245	4	0.5176

H0: no serial correlation

Test Dickey-Fullera (DF)

- ▶ Równanie regresji:

$$\Delta y_t = \alpha_1 + \rho y_{t-1} + \varepsilon_t$$

$H_0 : \rho = 0, \alpha_1 = 0$ - błądzenie przypadkowe

$H_1 : \rho \in (-2, 0)$ - model AR(1) ze stałą

Test Dickey-Fullera (DF)

- ▶ Równanie regresji:

$$\Delta y_t = \alpha_1 + \alpha_2 t + \rho y_{t-1} + \varepsilon_t$$

$H_0 : \rho = 0, \alpha_2 = 0$ - błądzenie przypadkowe z dryfem

$H_1 : \rho \in (-2, 0)$ - model AR(1) z trendem liniowym

Rozszerzony test Dickey-Fullera (ADF)

- ▶ Często reszty z regresji:

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

wykazują silną autokorelację, co jest problemem.

- ▶ Rozszerzony test Dickey-Fullera (test ADF) różni się od standardowego testu DF rozszerzeniem regresji o dodatkowe elementy, których celem jest eliminacja autokorelacji reszt.

Rozszerzony test Dickey-Fullera (ADF)

- ▶ Przeprowadzamy regresję:

$$\Delta y_t = \rho y_{t-1} + \sum_{i=1}^k \gamma_i \Delta y_{t-i} + \varepsilon_t$$

gdzie: $\gamma_i \Delta y_{t-i}$ - rozszerzenie

- ▶ Liczbę opóźnień k dobieramy tak aby z reszt wyeliminować autokorelację.

Rozszerzony test Dickey-Fullera (ADF)

- ▶ Statystyka testowa dla testu ADF : statystyka t policzona dla oszacowania parametru przy y_{t-1}
- ▶ Dla dużych prób tablice wartości krytycznych dla testu ADF są takie same jak w teście DF.
- ▶ Dla małych prób, małopróbkowe wartości krytyczne testu DF są jedynie aproksymacją prawdziwych wartości krytycznych testu ADF.

Rozszerzony test Dickey-Fullera (ADF)

- ▶ Test ADF przeprowadzamy w następujący sposób:
 1. regresja Δy_t na y_{t-1}
 2. badamy występowanie autokorelacji za pomocą testu Breuscha-Godfrey (nie Durbina-Watsona):
 - a) jeśli autokorelacja nie występuje to porównujemy statystykę t dla y_{t-1} z wartościami krytycznymi testu ADF:
 - i. wartość statystyki jest mniejsza od wartości krytycznej - odrzucamy H_0 o niestacjonarności y_t i przyjmujemy H_1 o stacjonarności y_t ;

Rozszerzony test Dickey-Fullera (ADF)

ii. wartość statystyki jest większa od wartości krytycznej –
nie ma podstaw do odrzucenia H_0

b) jeśli autokorelacja występuje to dodajemy stopniowo
rozszerzenia do momentu aż się jej pozbędziemy a następnie
porównujemy statystykę t dla y_{t-1} z wartościami krytycznymi
testu ADF

i. wartość statystyki jest mniejsza od wartości krytycznej
- odrzucamy H_0 o niestacjonarności y_t i przyjmujemy H_1
o stacjonarności y_t ;

Rozszerzony test Dickey-Fullera (ADF)

ii. wartość statystyki jest większa od wartości krytycznej –
nie ma podstaw do odrzucenia H_0

Rozszerzony test Dickey-Fullera (ADF)

- ▶ W przypadku błędzenia przypadkowego:

Source	SS	df	MS			
Model	.077155196	1	.077155196	Number of obs =	999	
Residual	1000.45476	998	1.00245968	F(1, 998) =	0.08	
Total	1000.53191	999	1.00153345	Prob > F =	0.7815	
				R-squared =	0.0001	
				Adj R-squared =	-0.0009	
				Root MSE =	1.0012	

D.rw	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
rw						
L1.	-.0005213	.001879	-0.28	0.782	-.0042086	.003166

Rozszerzony test Dickey-Fullera (ADF)

- ▶ W przypadku błędzenia przypadkowego:

Breusch-Godfrey LM test for autocorrelation

lags (p)		chi2	df	Prob > chi2
1		1.084	1	0.2977
2		1.477	2	0.4779
3		2.629	3	0.4524
4		2.639	4	0.6199

H0: no serial correlation

Test KPSS

- ▶ Test KPSS (Kwiatkowski, Philips, Schmidt, Shin) testuje hipotezę zerową o stacjonarności zmiennej.
- ▶ Oparty jest na następującym modelu:

$$y_t = \delta + \zeta_t + \varepsilon_t$$

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

$$\zeta_t = \zeta_{t-1} + u_t$$

$$u_t \sim IID(0, \sigma_u^2)$$

Test KPSS

$H_0 : \sigma_u^2 = 0$ - zmienna y_t jest stacjonarna

$H_1 : \sigma_u^2 > 0$ - zmienna y_t jest niestacjonarna

- Hipotezę zerową odrzucamy gdy statystyka testowa $>$ wartości krytycznej.
- Statystyka testowa dla testu KPSS zawsze >0 .

Test KPSS

▶ Test KPSS dla białego szumu:

Critical values for $H_0: x$ is level stationary

10%: 0.347 5% : 0.463 2.5%: 0.574 1% : 0.739

Lag order	Test statistic
0	.0629
1	.0656
2	.064
3	.0623
4	.0607
5	.0594
6	.0582
7	.0582
8	.0584
9	.059
10	.0596

Test KPSS

▶ Test KPSS dla białego szumu:

Critical values for H0: x is level stationary

10%: 0.347 5% : 0.463 2.5%: 0.574 1% : 0.739

Lag order	Test statistic
0	.0629
1	.0656
2	.064
3	.0623
4	.0607
5	.0594
6	.0582
7	.0582
8	.0584
9	.059
10	.0596

Test KPSS

▶ Test KPSS dla błędzenia przypadkowego

Critical values for H_0 : rw is level stationary

10%: 0.347 5% : 0.463 2.5%: 0.574 1% : 0.739

Lag order	Test statistic
0	55.9
1	28
2	18.7
3	14.1
4	11.3
5	9.42
6	8.09
7	7.09
8	6.31
9	5.69
10	5.18

Test KPSS

▶ Test KPSS dla błędzenia przypadkowego

Critical values for H0: rw is level stationary

10%: 0.347 5% : 0.463 2.5%: 0.574 1% : 0.739

Lag order	Test statistic
0	55.9
1	28
2	18.7
3	14.1
4	11.3
5	9.42
6	8.09
7	7.09
8	6.31
9	5.69
10	5.18

Pytania teoretyczne

1. Wyjaśnić na czym polega zjawisko regresji pozornej.
2. Dlaczego przed przystąpieniem do weryfikacji hipotez o istotności zmiennych w modelu szacowanym na szeregu czasowym powinniśmy przetestować ich rząd integracji?
3. Opisać procedurę testowania stacjonarności za pomocą rozszerzonego testu Dickey-Fullera (ADF).
4. Opisać jak dla różnych modeli trendów formułujemy hipotezę zerową i alternatywną w rozszerzonym teście Dickey-Fullera (ADF).
5. Na czym polega najważniejsza różnica między testowaniem stacjonarności za pomocą testu ADF i KPSS?

Dziękuję za uwagę