

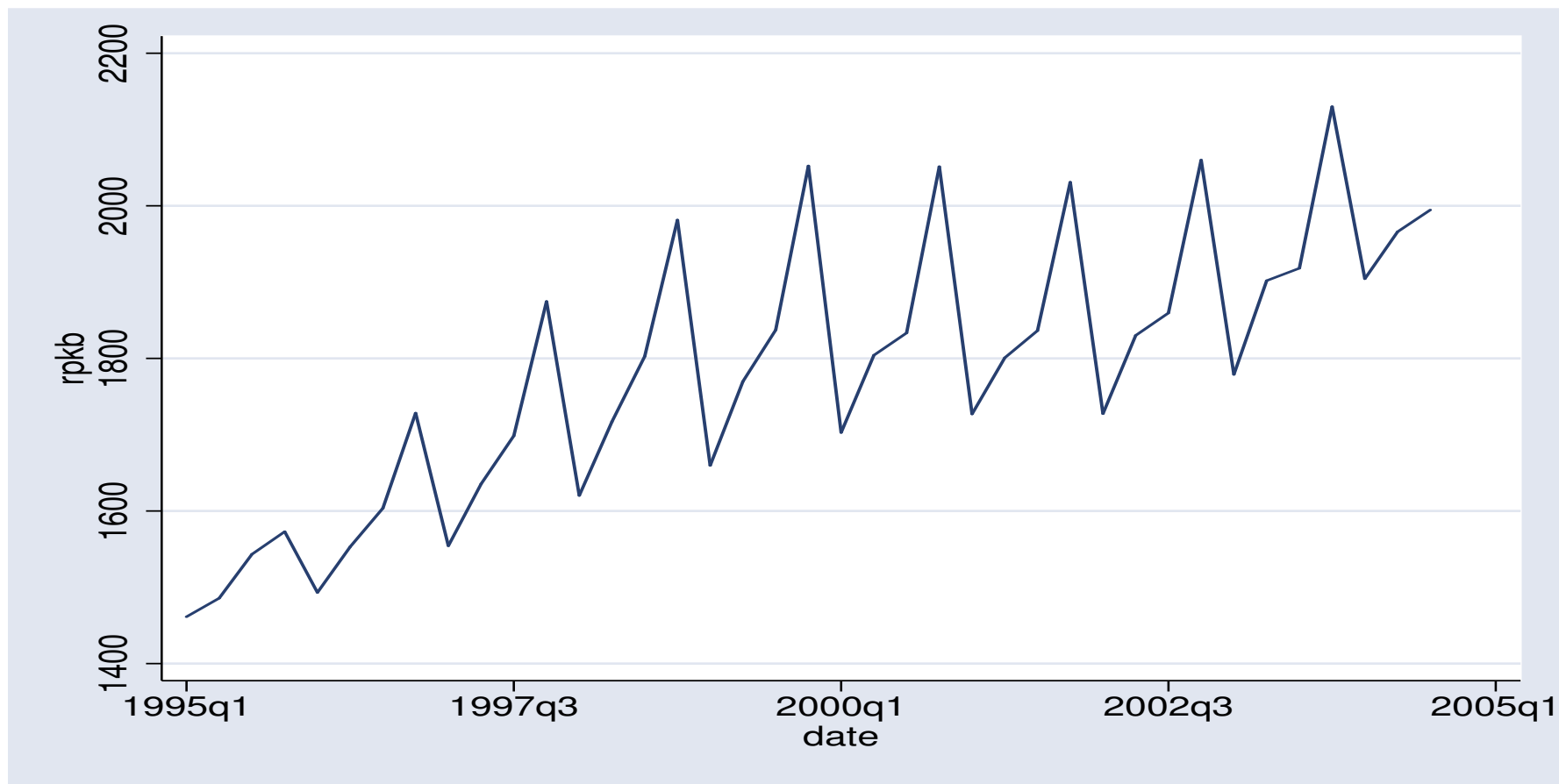
Sezonowość

- O sezonowości mówimy wtedy, gdy zmienna zmienia się w pewnym cyklu zwykle związanym z cyklem rocznym
- Na przykład zmienne kwartalne charakteryzują się zwykle sezonowością kwartalną a zmienne miesięczne sezonowością roczną
- Sezonowość można uwzględnić w modelu na dwa

sposoby:

- wstawiając do modelu zmienne zero jedynkowe związane z każdym kwartałem (miesiącem)
- stosując różnicowanie sezonowe: na przykład w przypadku danych miesięcznych $\Delta_4 y_t = y_t - y_{t-4}$

Przykład Poziom *PKB* w Polsce w latach 1996-2004. Zmiany sezonowe mają w tym przypadku głównie związek z cyklem inwestycyjnym - największe nakłady inwestycyjne księgowane są w czwartym kwartale.



- *PKB* w wyrażeniu realnym - deflator *CPI*

Zmienne stacjonarne i niestacjonarne

- Zmienna stacjonarna:

1. Wartość oczekiwana równa 0

$$E(y_t) = 0$$

2. Wariancja skończona

$$\text{Var}(y_t) = \sigma^2 < \infty$$

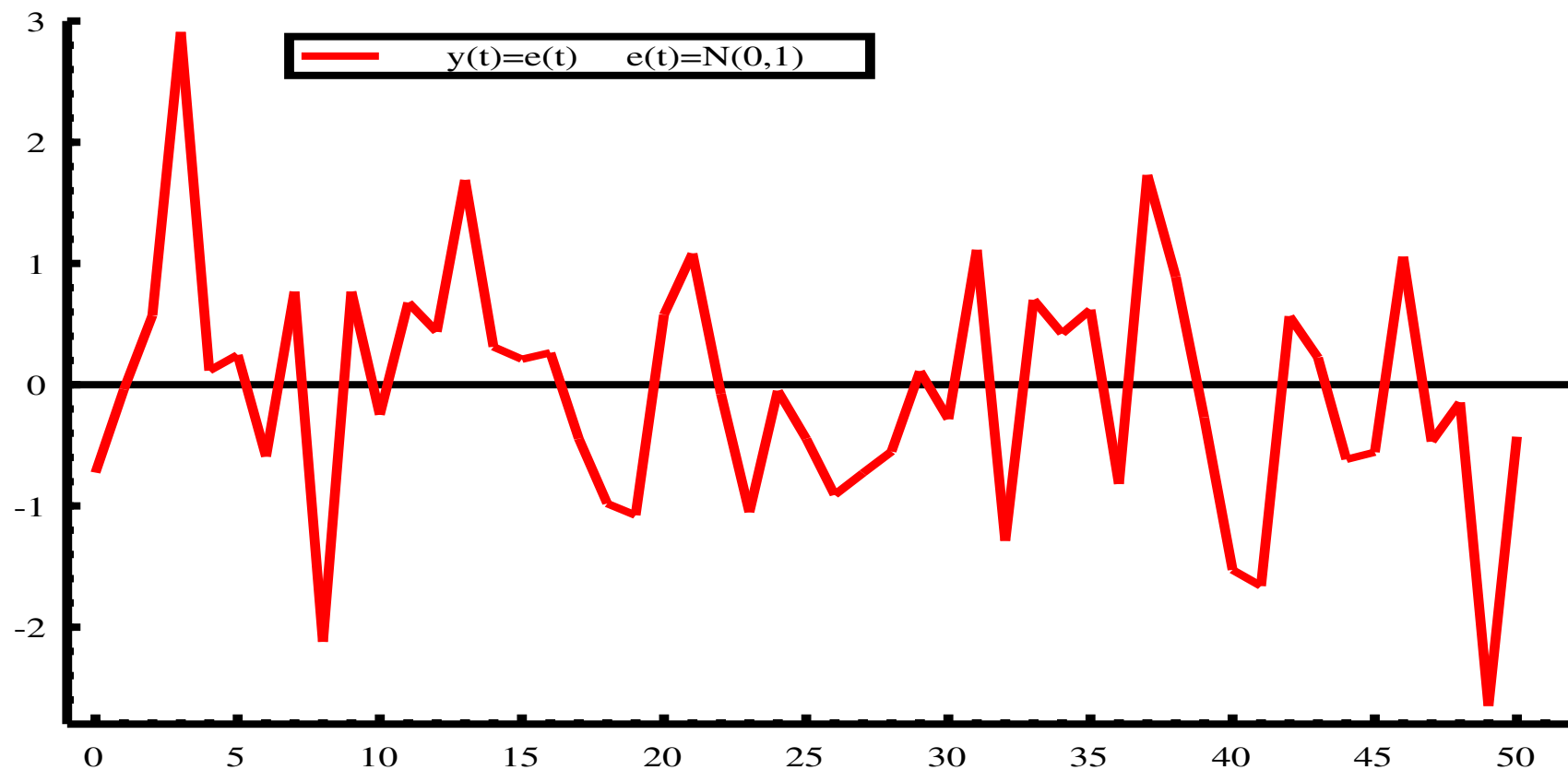
3. Kowariancje zależą jedynie od h - dystansu między realizacjami

$$\text{Cov}(y_{t_1}, y_{t_1+h}) = \text{Cov}(y_{t_2}, y_{t_2+h}) = \gamma_h$$

dla dowolnych t, t_1, t_2 i h .

- Zmienną nazywamy trendostacjonarną jeśli $y_t - E(y_t)$ jest stacjonarne
- Często rozpatrywanym przykładem zmiennej stacjonarnej jest biały szum:

1. y_t ma rozkład stały w czasie i o wartości oczekiwanej 0 i skończonej wariancji σ^2
2. realizacje y_t są niezależne



- biały szum (zmienna $I(0)$)

Procesy zintegrowane

- Szczególnym rodzajem zmiennej stacjonarnej jest zmienna $x_t \sim I(0)$. Zmienną $I(0)$ definiuje się jako zmienną stacjonarną, którą można przedstawić jako $x_t = \sum_{i=0}^{\infty} \theta_i \varepsilon_{t-i}$, gdzie ε_i i ε_j są nieskorelowane.
- Zmienna zintegrowana rzędu d oznaczana jako $x_t \sim I(d)$, to taka zmienna, która po zastosowaniu d -tych różnic staje się zmienną stacjonarną $\Delta^d x_t \sim I(0)$.

Przykład Przykładem zmiennej niestacjonarnej zmienna zachowująca się zgodnie z modelem błędzenia przypadkowego

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

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

Jeśli podstawimy w tym wzorze za $y_{t-1} = y_{t-2} + \varepsilon_{t-1}$, to uzyskamy

$$y_t = y_{t-2} + \varepsilon_{t-1} + \varepsilon_t$$

Postępując rekurencyjnie w ten sposób uzyskamy:

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

Jeśli założymy, że $y_0 = 0$, to

$$E(y_t) = 0$$

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

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

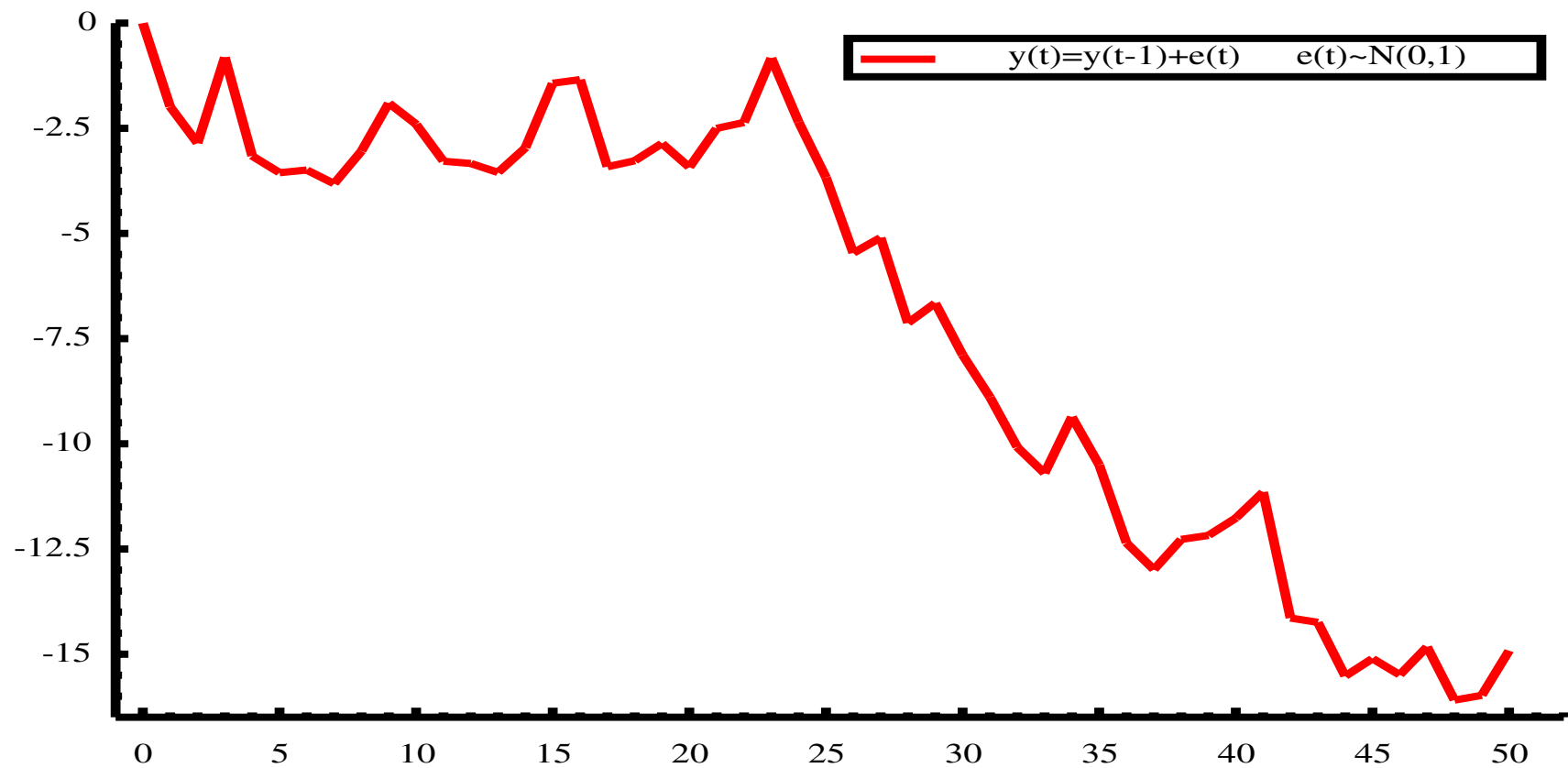
Wariancja i kowariancje błędzenia przypadkowego zależą od czasu - zmienna ta jest niestacjonarna! Za-

uważmy, że kowariancje między realizacjami zmiennej maleją bardzo wolno - w przybliżeniu liniowo.

Odejmując od obu stron y_{t-1} otrzymujemy:

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

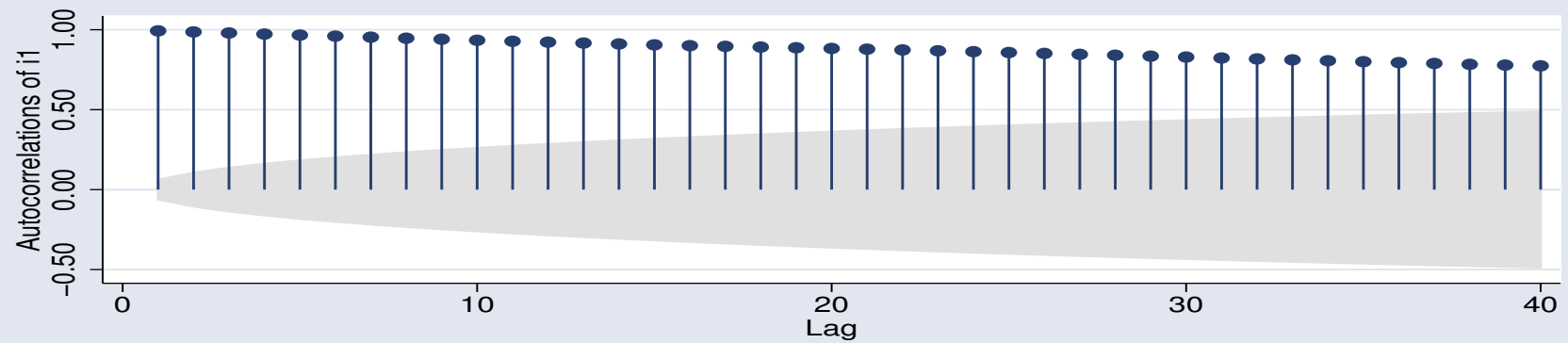
Po zastosowaniu pierwszych różnic otrzymaliśmy zmienną zachowującą się zgodnie z modelem białego szumu - zmienną $I(0)$. Wnioskujemy z tego, że błądzenie przypadkowe jest $I(1)$



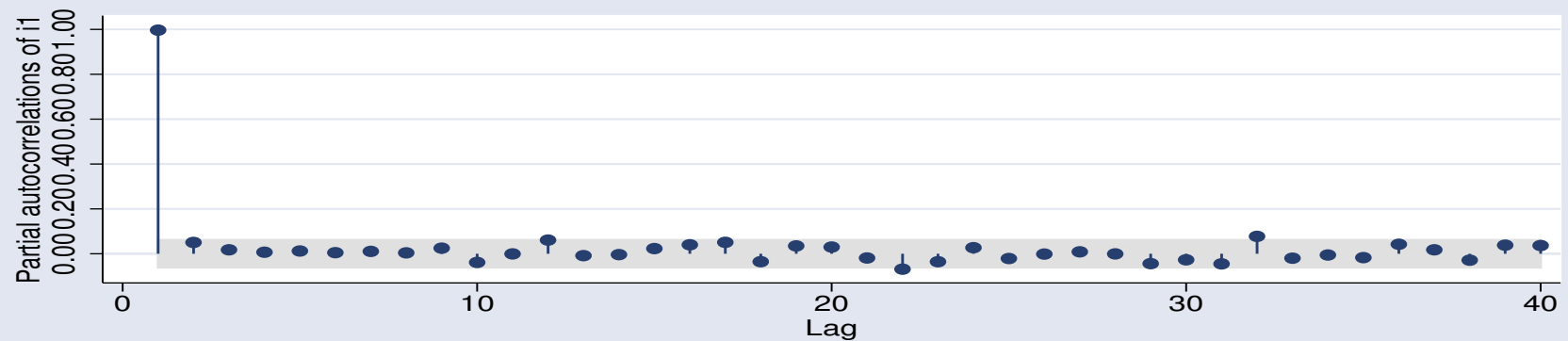
- błądzenie przypadkowe (zmienna $I(1)$)

- W ekonomii ważne są jedynie zmienne $I(1)$ i $I(2)$
- Wiele procesów w ekonomii wygląda tak na zmienne $I(1)$.
- W znanym artykule Nelson i Plosser (1982) pokazali, że znacząca część szeregów makroekonomicznych wygląda na szeregi $I(1)$
- W kontekście modelowania za pomocą modeli $ARMA$ problem niestacjonarności wykrywany był za pomocą funkcji ACF i $PACF$.

- Zmienne niestacjonarne charakteryzują się
 - wysoką korelacją między realizacjami - wolno malejąca funkcja ACF
 - bliską jedności wartością współczynniki przy pierwszym opóźnieniu - $PACF$ dla pierwszego opóźnienia bliskie jedności



Bartlett's formula for MA(q) 95% confidence bands



95% Confidence bands [se = 1/sqrt(n)]

- ACF i $PACF$ dla błędzenia przypadkowego $I(1)$

- Przy modelowaniu $ARMA$ w razie wykrycia niestacjonarności powinno się szereg zróżnicować i modelować pierwsze różnice.
- W przypadku, kiedy nawet pierwsze różnice wydają się niestacjonarne należy dalej różnicować - aż do skutku.
- Model $ARMA(p, q)$ dla zmiennej zróżnicowanej d razy oznacza się jako $ARIMA(p, d, q)$ - **AutoRegressive Integrated Moving Avarage**.

- Rząd integracji zmiennej można przetestować za pomocą jednego z testów stacjonarności/niestacjonarności (testów na istnienie pierwiastka jednostkowego)

Test Dickey-Fullera (DF)

- Badamy, czy zmienna jest niestacjonarna
- Najprostszy przypadek

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

- $H_0 : \beta = 1$, wtedy y_t jest niestacjonarna (jest błędzeniem przypadkowym - procesem $I(1)$)

- $H_1 : |\beta| < 1$, wtedy y_t jest stacjonarna
- Przekształcając

$$\begin{aligned}\Delta y_t &= (\beta - 1) y_{t-1} + \varepsilon_t \\ &= \rho y_{t-1} + \varepsilon_t\end{aligned}$$

- $H_0 : \rho = 0$, wtedy y_t jest niestacjonarna
- $H_0 : \rho \in (-2, 0)$, wtedy y_t jest niestacjonarna
- Hipotezy, że $\rho = 0$ nie można badać za pomocą

standardowych tablic - używamy do tego specjalnych tablic dla testu DF

- Sposób przeprowadzania testu:
 1. Przeprowadzamy regresję Δy_t na y_{t-1}
 2. Porównujemy statystykę t dla y_{t-1} z wartościami krytycznymi testu DF . Jeśli współczynnik jest mniejszy od wartości krytycznej odrzucamy H_0 .

Uwaga: W teście Dickey Fullera:

- H_0 zmienna jest *niestacjonarna*

- H_1 zmienna jest stacjonarna

Uwaga: W niektórych tablicach testu DF przed wartościami krytycznymi pomijany jest znak $-$.

Rozszerzony test Dickey-Fullera (ADF)

- Często w regresji $\Delta y_t = \rho y_{t-1} + \varepsilon_t$ występuje autokorelacja
- Autokorelacja w regresjach z opóźnioną zmienną zależną jest bardzo niekorzystna - prowadzi do wystąpienia problemu równoczesności.
- Stosujemy wtedy test DF z rozszerzeniem (test

ADF - Augmented Dickey Fuller)

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

- Ilość opóźnień k dobieramy w taki sposób, by w modelu nie występowała autokorelacja.
- Sposób uwzględnienia stałej lub trendu w modelu:

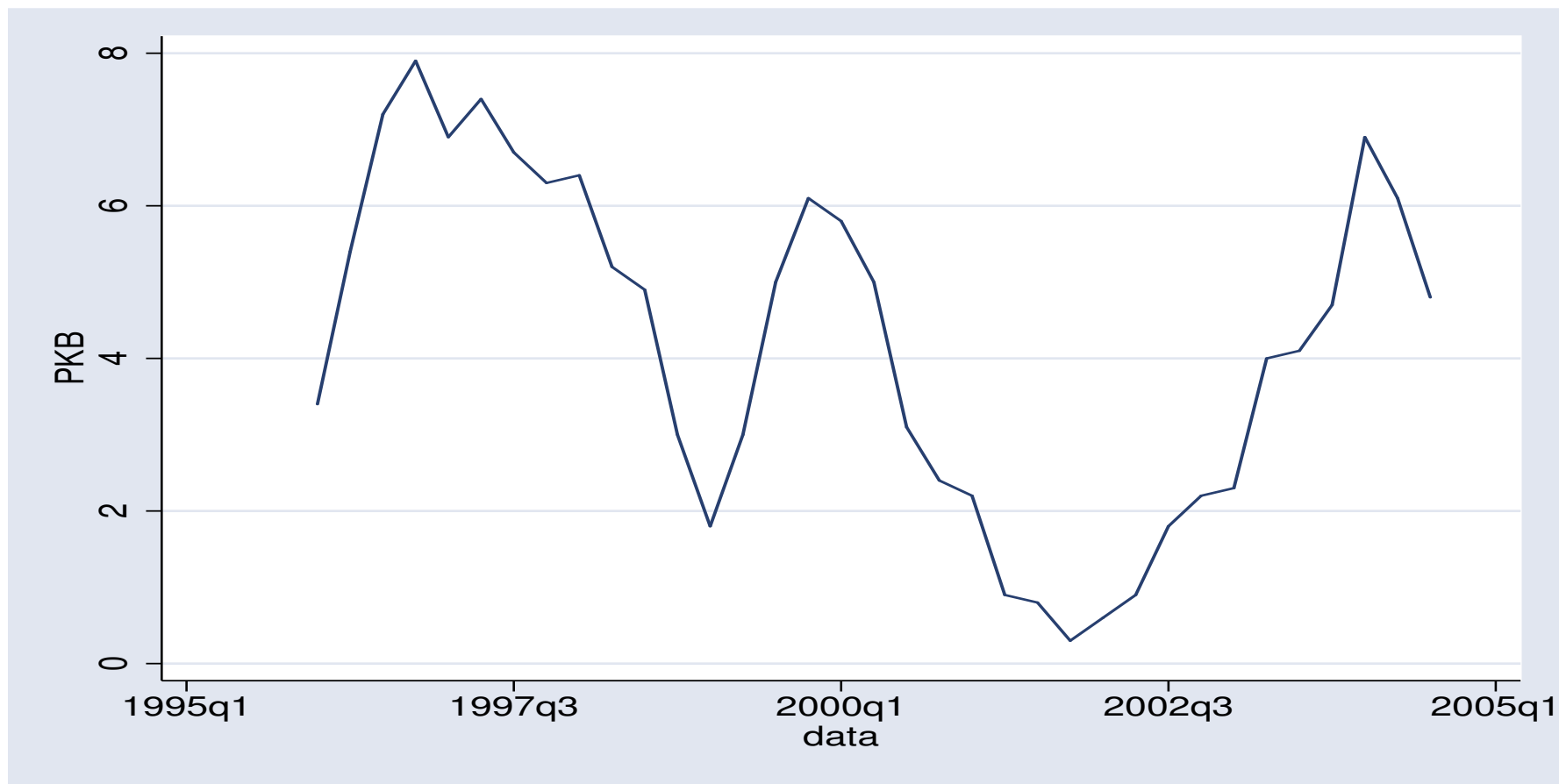
1. $\Delta y_t = \rho y_{t-1} + \sum_{i=1}^k \gamma_i \Delta y_{t-i} + \varepsilon_t$

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

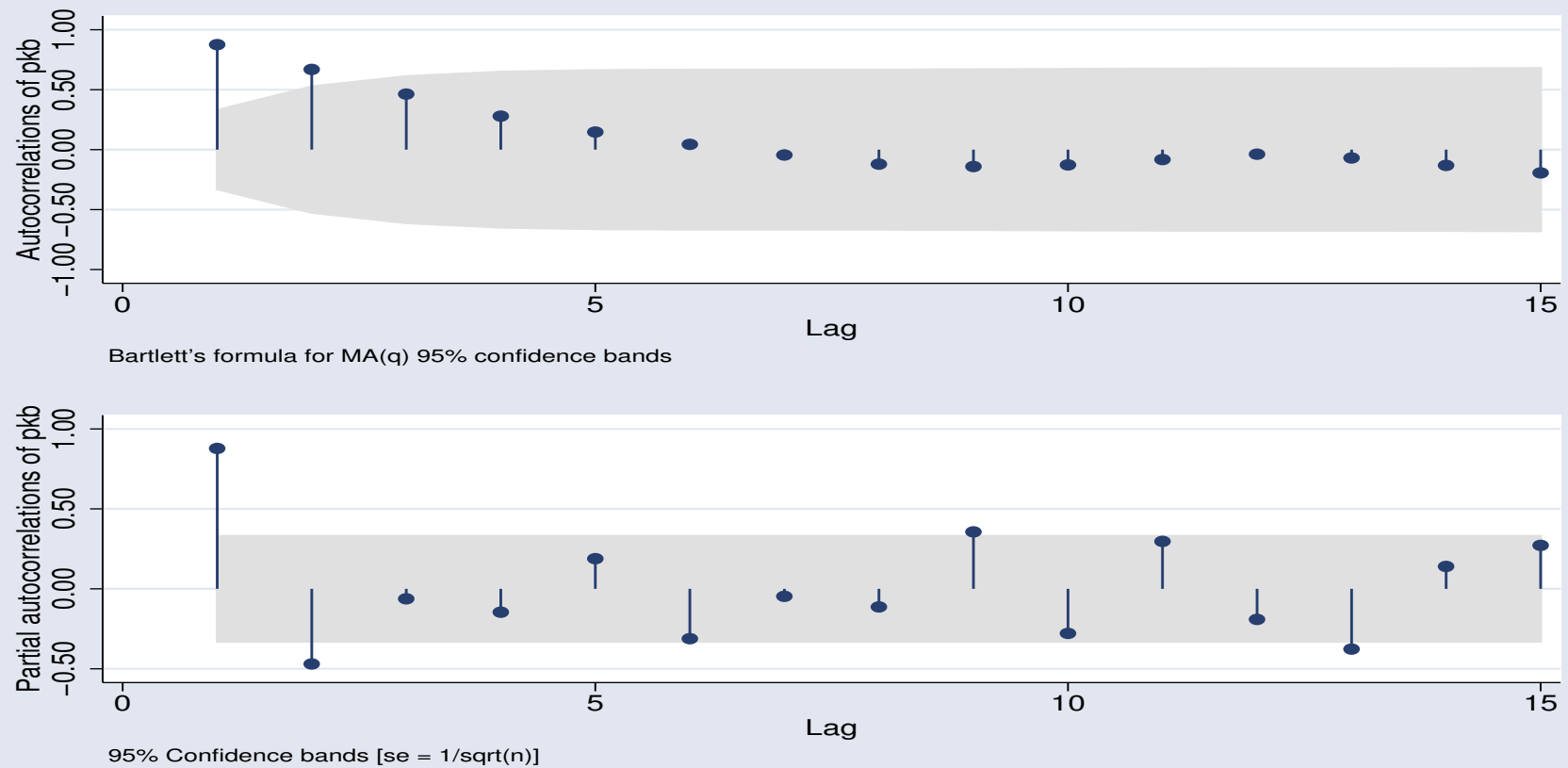
$$3. \Delta y_t = a + bt + \rho y_{t-1} + \sum_{i=1}^k \gamma_i \Delta y_{t-i} + \varepsilon_t$$

- Dla każdej z tych możliwości stosuje się inną wersję tablic DF

Przykład Realny wzrost PKB w latach 1996.1-2004.3, indeks kwartalny - wzrost PKB kwartał do analogicznego kwartału poprzedniego roku. Zauważmy, że przy tak skonstruowanym indeksie (różnicowanie sezonowe), znika sezonowość.



- Wzrost *PKB* indeks kwartalny



- *ACF* i *PACF* dla indeksu *PKB*

D.pkb		Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
pkb							
	L1	-.2760446	.1350873	-2.04	0.056	-.5598524	.0077633
	LD	.5846543	.2200515	2.66	0.016	.1223433	1.046965
	L2D	-.0616403	.2488176	-0.25	0.807	-.5843866	.461106
	L3D	.3055767	.2634729	1.16	0.261	-.2479593	.8591128
	L4D	-.3489368	.2367631	-1.47	0.158	-.8463575	.148484
	L5D	.2966282	.2515727	1.18	0.254	-.2319064	.8251628
	L6D	-.0437733	.271078	-0.16	0.874	-.6132871	.5257405
	L7D	.1129747	.2414823	0.47	0.646	-.3943609	.6203102
_cons		.9466827	.54444	1.74	0.099	-.1971433	2.090509

Breusch-Godfrey LM test for autocorrelation

lags(p)	chi2	df	Prob > chi2
1	0.568	1	0.4511

H0: no serial correlation

Augmented Dickey-Fuller test for unit root Number of obs = 27

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

Z(t)	-2.043	-3.736	-2.994	-2.628

* MacKinnon approximate p-value for Z(t) = 0.2679

Test KPSS (Kwiatkowski, Philips, Schmidt, Shin)

- Możliwe jest także testowanie stacjonarności za pomocą testów, w których H_0 mówi, że zmienna jest $I(0)$
- Tego rodzaju testem test test $KPSS$
- W modelu

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

gdzie ε_t jest $I(0)$ oraz

$$\zeta_t = \zeta_{t-1} + u_t, \text{ i } u_t \sim IID(0, \sigma_u^2)$$

a więc ζ_t jest błędzeniem przypadkowym.

- Testujemy $H_0 : \sigma_u = 0$ co jest równoważnie hipotezie, że $\zeta_t = 0$.
- Jeśli H_0 jest prawdziwa, błędzenie przypadkowe nie odgrywa roli i y_t jest stacjonarne.
- Jeśli odrzucimy H_0 , to musimy przyjąć hipotezę al-

trenatywną, że proces jest niestacjonarny.

- Często się zdarza (szczególnie jeśli mamy mało obserwacji), że zarówno H_0 w teście DF (niestacjonarność) jak H_0 w teście $KPSS$ (stacjonarność), nie mogą być odrzucone.
- Sposób pozbywania się wpływu korelacji między ε_t jest w przypadku testu $KPSS$ dosyć skomplikowany.

Przykład (kontynuacja) Testowanie stacjonarności

Ści wzrostu PKB

KPSS test for pkb

Maxlag = 9 chosen by Schwert criterion

Autocovariances weighted by Bartlett kernel

Critical values for H_0 : pkb is level stationary

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

Lag order	Test statistic
0	1.14
1	.606
2	.435
3	.354
4	.308
5	.281
6	.263
7	.252
8	.245
9	.241

- Dla większej liczby opóźnień nie jesteśmy w stanie odrzucić hipotezy o stacjonarności wzrostu PKB !

KPSS test for D.pkb

Maxlag = 9 chosen by Schwert criterion

Autocovariances weighted by Bartlett kernel

Critical values for H_0 : D.pkb is level stationary

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

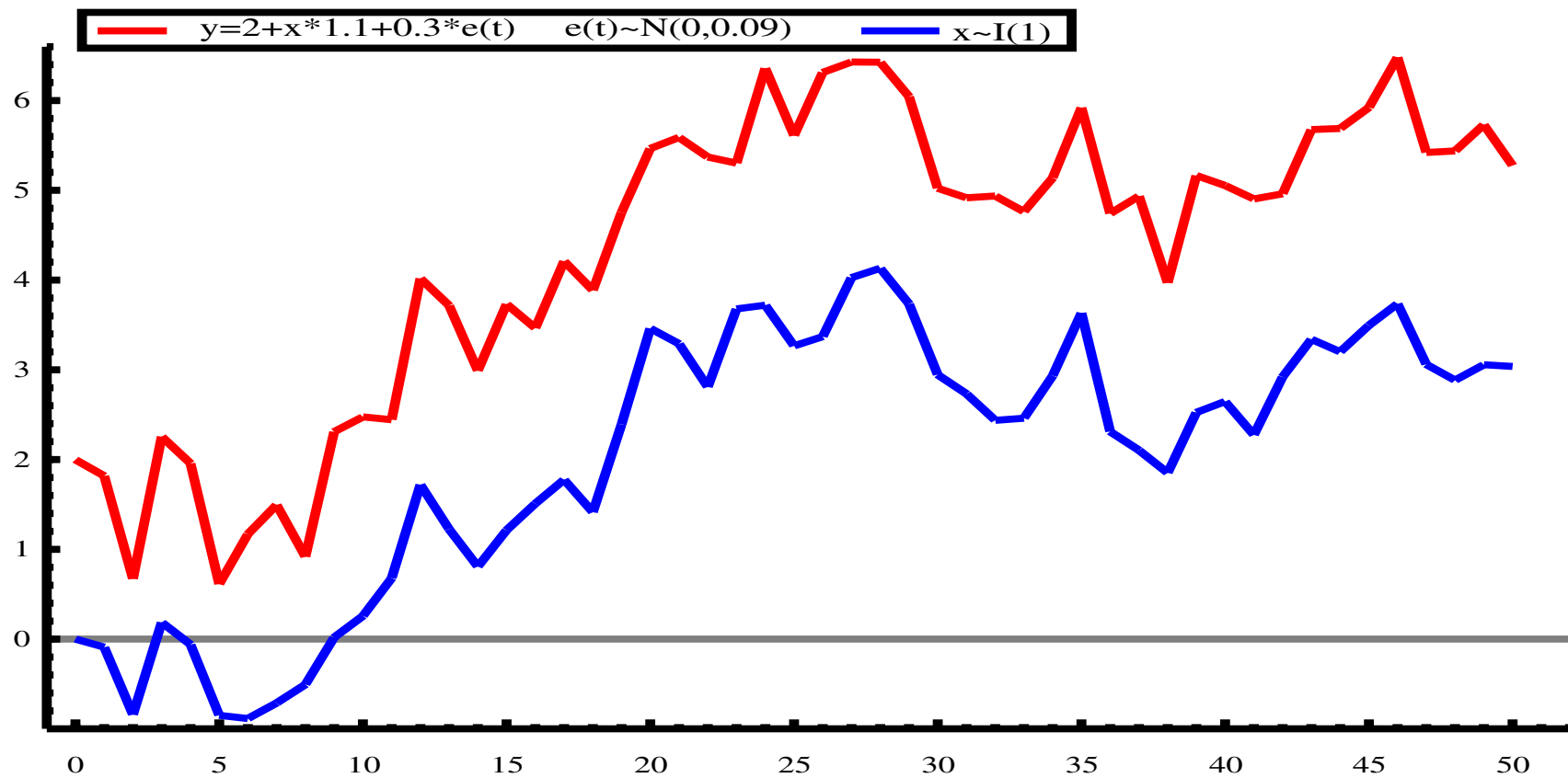
Lag order	Test statistic
0	.146
1	.107
2	.0979
3	.095
4	.0989
5	.104
6	.11
7	.115
8	.125
9	.137

- Pierwsze różnice wzrostu PKB są niewątpliwie stacjonarne

Procesy skointegrowane

- O procesach $x_{1t} \sim I(1)$ i $x_{2t} \sim I(1)$ mówimy, że są skointegrowane jeśli istnieje takie β , że $x_{1t} + \beta x_{2t}$ jest $I(0)$.
- W przypadku wektora x_t , jeśli każdy element wektora x_t jest $I(1)$, to mówimy, że elementy tego wektora są skointegrowane wektorem kointegrującym β , jeśli $\beta' x_t$ jest $I(0)$

- Istnieje taka kombinacja liniowa elementów wektora x_t , która jest $I(0)$, mimo że elementy tego wektora są $I(1)$



- zmienne skointegrowane

Regresja pozorna

- Problem regresji pozornej może się pojawić jeśli zmienne w modelu nie są $I(0)$
- Może to dotyczyć zarówno zmiennych objaśnianych jak i objaśniających
- W regresji ze zmiennymi $I(1)$ uzyskujemy w dużych próbach istotne statystyki t w MNK nawet jeśli zmienna endogeniczna i zmienne egzogeniczne

są od siebie niezależne. Estymator MNK w takiej regresji jest dalej zgodny.

- Przeprowadźmy regresję y_t na x_t , gdzie y_t i x_t są *niezależnymi* procesami $I(1)$.

$$y_t = \beta_0 + \beta_1 x_t + e_t$$

- Wartości p z rozkładu t -studenta i uzyskane z sy-

mulacji dla hipotezy $H_0 : \beta_1 = 0$.

		p-stwo, że $ t > t_\alpha$	
liczba obserwacji	t_α	t-student	symulacja
10	2.31	0.05	0.28
100	1.98	0.05	0.76
1000	1.96	0.05	0.93

- Dla długich obserwacji jest niemal pewne, że $|t| > t_\alpha$ i że w konsekwencji odrzucimy prawdziwą hipotezę zerową o braku związku między zmiennymi!

- Problem regresji pozornej jest potencjalnie bardzo poważny, ponieważ może doprowadzić do zbudowania modelu, w którym zmienne są całkowicie niepowiązane ze sobą mimo pozornie istotnych statystyk t
- UWAGA: mimo, że statystyki t mają całkowicie nietypowe rozkłady w przypadku, kiedy zmienna zależna jest $I(1)$ i zmienne niezależne są $I(1)$, to jednak estymatory współczynników regresji są zgodne.

- Proste rozwiązanie problemu regresji pozornej: w przypadku zmiennych $I(1)$ możemy oszacować model na pierwszych różnicach - te będą $I(0)$ więc statystyki t będą miały standardowe rozkłady.
 - Problem: za pomocą regresji na pierwszych różnicach nie da się oszacować równowagi długookresowej:

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

i przy założeniu, że $y^* = E(y_t) = E(y_{t-1}) = \dots$
i $x^* = E(x_t) = E(x_{t-1}) = \dots$ otrzymujemy małą
mówiącą tożsamość $0 = 0\beta$.

Mechanizm korekty błędem

- Praktyczne znaczenie pojęcia kointegracji związane jest z twierdzeniem Grangera,

Twierdzenie Jeśli (y_t, x_t) są skointegrowane, oraz y_t i x_t są $I(1)$, to y_t można przedstawić w postaci Mechanizmu Korekty Błędem (ECM – **E**rror **C**orrection

Mechanism) gdzie $y_{t-1} - \mathbf{x}_{t-1} \boldsymbol{\beta} \sim I(0)$

$$\Delta y_t = \boldsymbol{\alpha} (y_{t-1} - \mathbf{x}_{t-1} \boldsymbol{\beta}) + \sum_{i=1}^{k-1} \boldsymbol{\theta}_i \Delta y_{t-i} + \sum_{i=0}^{k-1} \Delta \mathbf{x}_{t-i} \boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}_t$$

- Znaczenie twierdzenia Grangera: umożliwia ono nam interpretację wektora kointegrującego jako równowagi długookresowej:

$$\begin{aligned}
y^* &= E(y_t) = \dots = E(y_{t-k}) \\
\implies E(\Delta y_t) &= \dots = E(\Delta y_{k-1}) = 0 \\
x^* &= E(x_t) = \dots = E(x_{t-k}) \\
\implies E(\Delta x_t) &= \dots = E(\Delta x_{k-1}) = 0
\end{aligned}$$

a więc z *ECM*:

$$0 = \alpha (y^* - x^* \beta)$$

- Równowaga długookresowa dana wzorem $y^* =$

$x^*\beta$ jest interpretowana jako taka relacja między zmiennymi, do której zmienne dążą przy braku zakłóceń związanych z zaburzeniami losowymi

- $y_t - x_t\beta$ jest odchyleniem od tej równowagi długookresowej.
- Odchylenie to jest korygowane za pośrednictwem $\alpha (y_{t-1} - x_{t-1}\beta)$.
- Współczynnik α związany jest więc szybkością dostosowania y_t do poziomu równowagi.

- Współczynniki θ_i i γ_i związane są z krótkookresową dynamiką procesu.
- Do modelu powinniśmy wstawić taką liczbę opóźnionych różnic y_t i x_t , by wyeliminować autokorelację reszt. Przy ustalaniu k można posłużyć się metodą od ogólnego do szczegółowego bądź kryteriami informacyjnymi.

Dwustopniowa metoda Engela-Grangera

Uwaga: Testowanie kointegracji ma sens jedynie wtedy, gdy y_t i *wszystkie* zmienne zawarte w x_t są $I(1)$.

- Pierwszym etapem analizy kointegracji musi więc być przetestowanie, czy wszystkie analizowane zmienne są $I(1)$

1. Estymujemy równanie na poziomach

$$y_t = x_t \beta + u_t$$

otrzymujemy relację długookresową $1, -\hat{\beta}'$. Ponieważ w regresji zmiennej $I(1)$ na $I(1)$ oszacowanie β jest zgodne więc reszty $\hat{u}_t$ stanowią "dobre" oszacowanie błędów losowych u_t .

2. Przeprowadzamy test DF dla $\hat{u}_t$: jeśli $\hat{u}_t$ stacjonarne to występuje kointegracja

3. Rozkład statystyki testu DF w przypadku testowania kointegracji zależy od liczby zmiennych objaśniających zawartych w wektorze $\mathbf{x}_t$.

4. Szacujemy ECM z wykorzystując otrzymane $\hat{\beta}$

$$\Delta y_t = \alpha \left(y_{t-1} - \mathbf{x}_{t-1} \hat{\beta} \right) + \sum_{i=1}^{k-1} \theta_i \Delta y_{t-i} + \sum_{i=0}^{k-1} \Delta \mathbf{x}_{t-i} \gamma_i + \epsilon_t$$

- Tak uzyskane estymatory parametrów α i β są zgodne ale nieefektywne

- Istnieją także inne metody szacowania współczynników w *ECM*

Przykład Badamy istnienie kointegracji między realnym wzrostem spożycia indywidualnego gospodarstw domowych i realnym wzrostem *PKB*. Używamy danych kwartalnych 1996.1–2004.3, wzrost liczony kwartał do analogicznego kwartału. Wyliczony wcześniej test *ADF* wskazał, że *PKB* jest niestacjonarne.

- Wynik testu ADF dla pierwszych różnic PKB :

Dickey-Fuller test for unit root

Number of obs = 33

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

Z(t)	-3.872	-3.696	-2.978	-2.620

* MacKinnon approximate p-value for Z(t) = 0.0023

- Wniosek: PKB niestacjonarne, ΔPKB stacjonarne $\implies PKB \sim I(1)$
- Podobnie dla spożycia indywidualnego dla poziomów otrzymujemy

Dickey-Fuller test for unit root Number of obs = 34

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

Z(t)	-2.265	-3.689	-2.975	-2.619

* MacKinnon approximate p-value for Z(t) = 0.1834

• A dla pierwszych różnic:

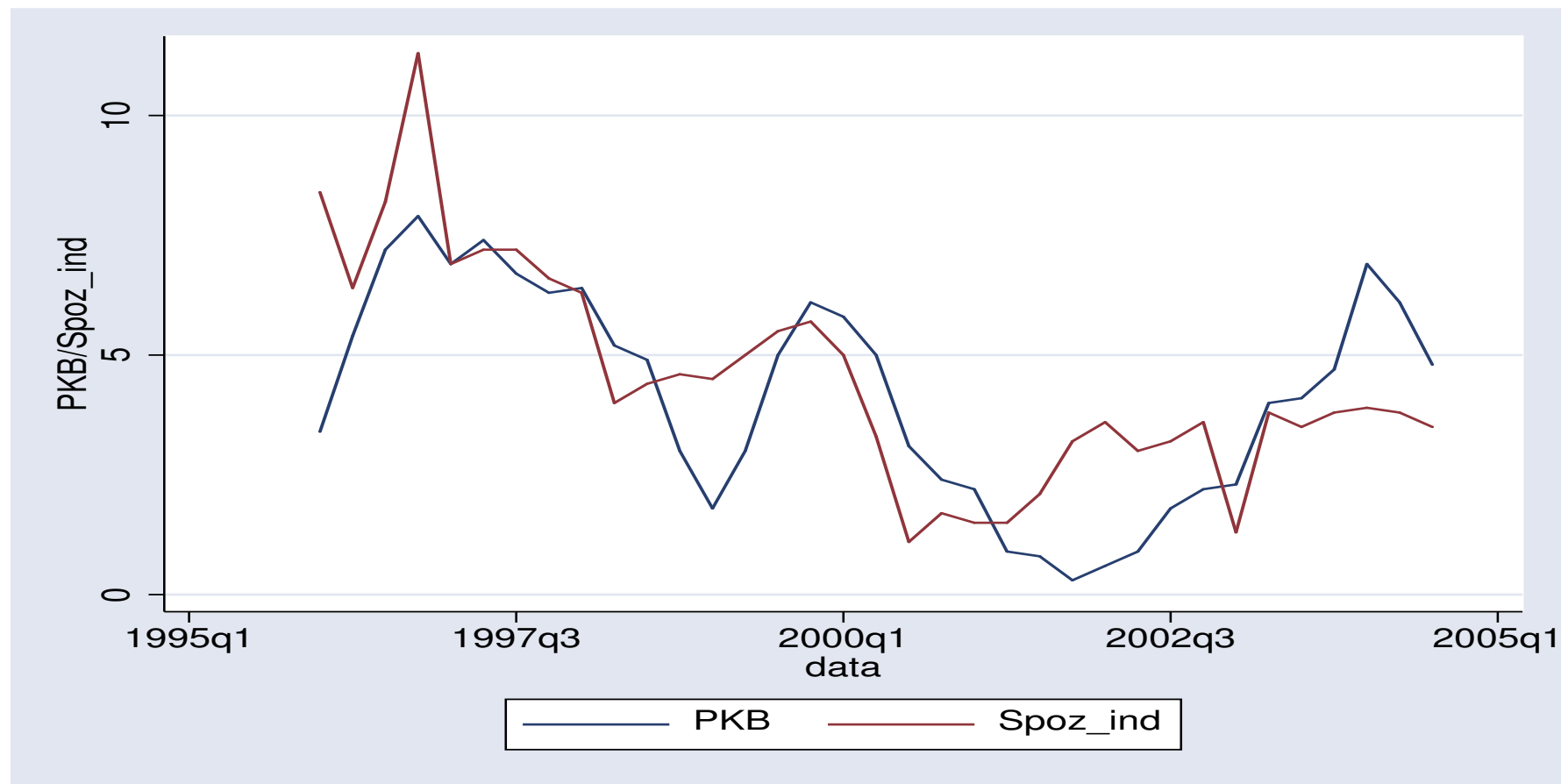
Dickey-Fuller test for unit root Number of obs = 33

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

Z(t)	-7.426	-3.696	-2.978	-2.620

* MacKinnon approximate p-value for Z(t) = 0.0000

- Wynika z tego, że spożycie indywidualne jest też $I(1)$
- Kointegracja między zmiennymi jest możliwa!



- zmienna PKB i spożycie indywidualne

- Pierwszy stopień procedury Engla-Grangera - regresja *PKB* na stałej i spożyciu indywidualnym

Source	SS	df	MS	Number of obs = 35		
Model	85.7028718	1	85.7028718	F(1, 33)	=	31.38
Residual	90.1325542	33	2.73128952	Prob > F	=	0.0000
Total	175.835426	34	5.17163018	R-squared	=	0.4874
				Adj R-squared	=	0.4719
				Root MSE	=	1.6527

spoz_ind	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
pkb	.7095951	.1266767	5.60	0.000	.4518694	.9673208
_cons	1.58154	.5961194	2.65	0.012	.3687263	2.794354

- Tworzymy reszty i przeprowadzamy dla nich test Dickey-Fullera:

Dickey-Fuller test for unit root

Number of obs = 34

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

Z(t)	-3.505	-3.689	-2.975	-2.619

* MacKinnon approximate p-value for Z(t) = 0.0079

- Uzyskane automatycznie wartości krytyczne testu nie są prawidłowe (dotyczą przypadku testowania rzędu integracji)
- Znajdujemy prawidłową wartość $p = .0319 < 0.05$
- Odrzucamy hipotezę zerową o niestacjonarności

reszt i przyjmujemy hipotezę alternatywną o ich stacjonarności - między zmiennymi występuje kointegracja!

• Szacujemy *ECM*

Source	SS	df	MS	Number of obs = 34			
Model	5.54977348	1	5.54977348	F(1, 32)	=	3.00	
Residual	59.2740495	32	1.85231405	Prob > F	=	0.0931	
				R-squared	=	0.0856	
				Adj R-squared	=	0.0570	
Total	64.823823	33	1.96435827	Root MSE	=	1.361	

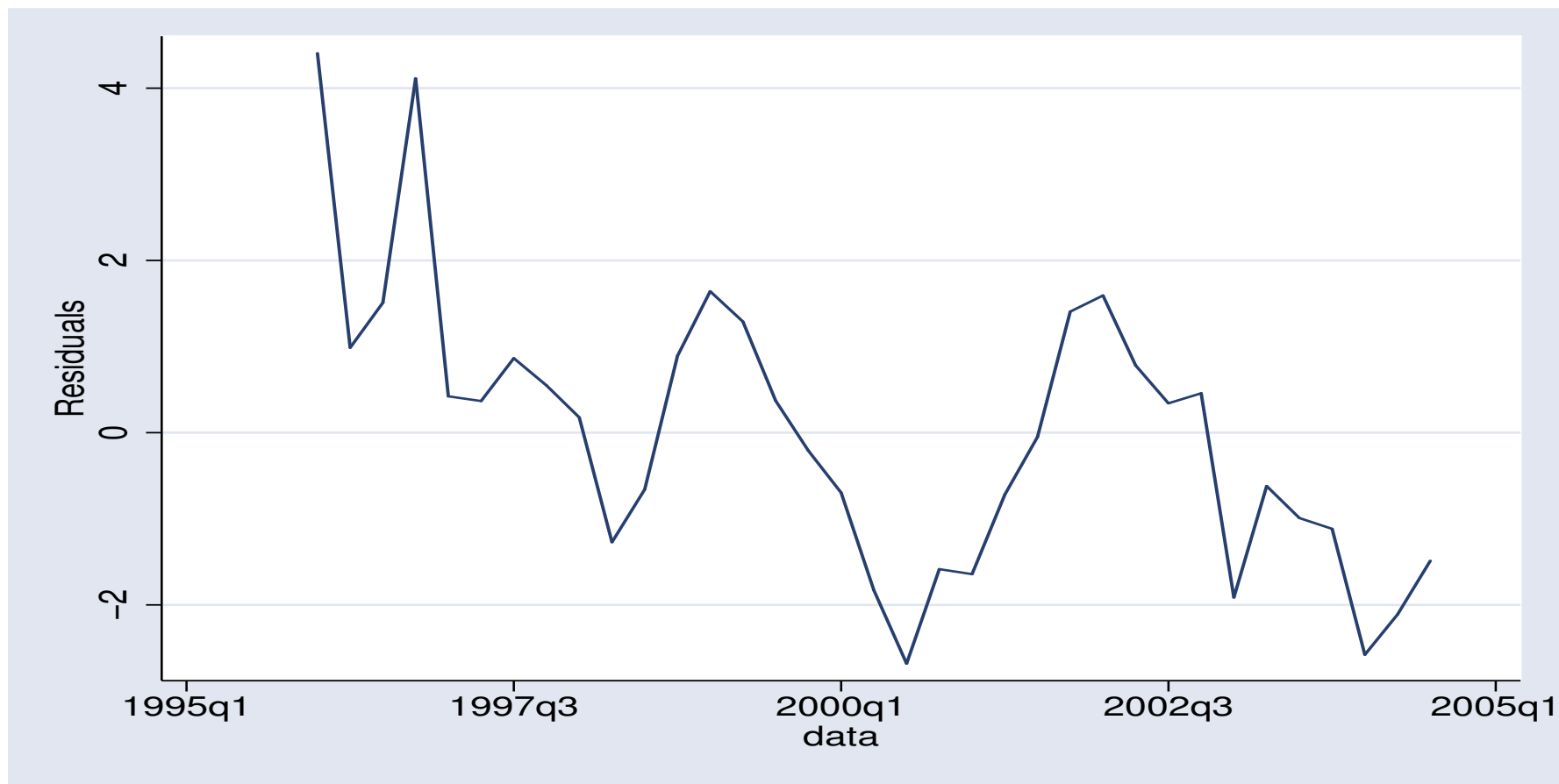
D.spoz_ind	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
ecm						
L1	-.2513365	.1452029	-1.73	0.093	-.5471051	.0444321
_cons	-.1331209	.2334956	-0.57	0.573	-.6087359	.342494

Breusch-Godfrey LM test for autocorrelation

lags(p)	chi2	df	Prob > chi2
1	0.318	1	0.5727

H0: no serial correlation

- Wynik testu Breusch-Godfrey wskazuje na to że nie ma potrzeby wstawiania do ECM dodatkowych elementów związanych z opóźnionymi po pierwszymi różnicami x_t i y_t .
- Współczynnik przy ecm wskazuje na to, 25% różnicy między zależnością długokresową a aktualnymi wartościami y_t i x_t jest redukowana w ciągu jednego kwartału.



- odchylenia od stanu równowagi - zmienna ecm