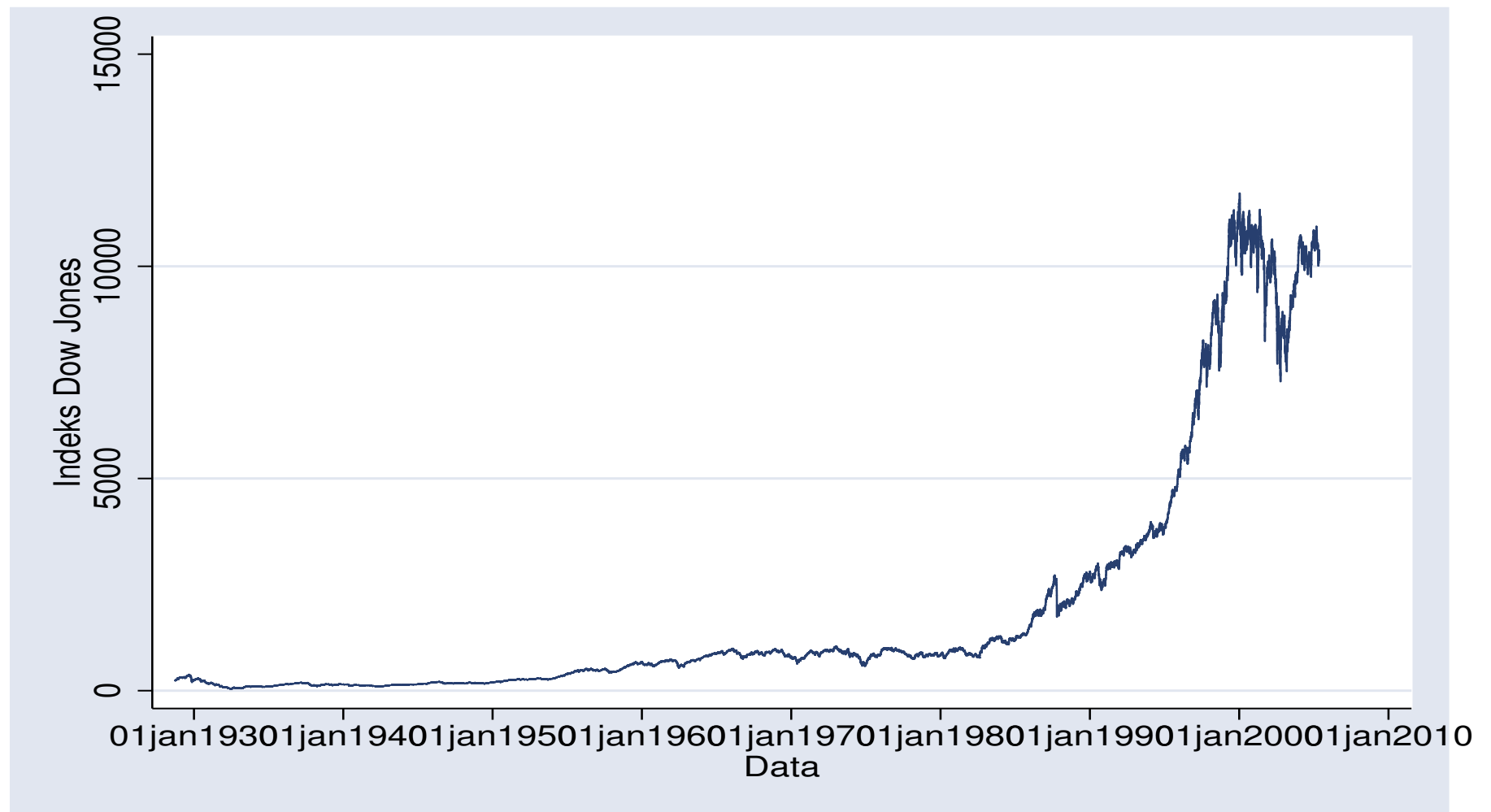


Heteroskedastyczność w szeregach czasowych

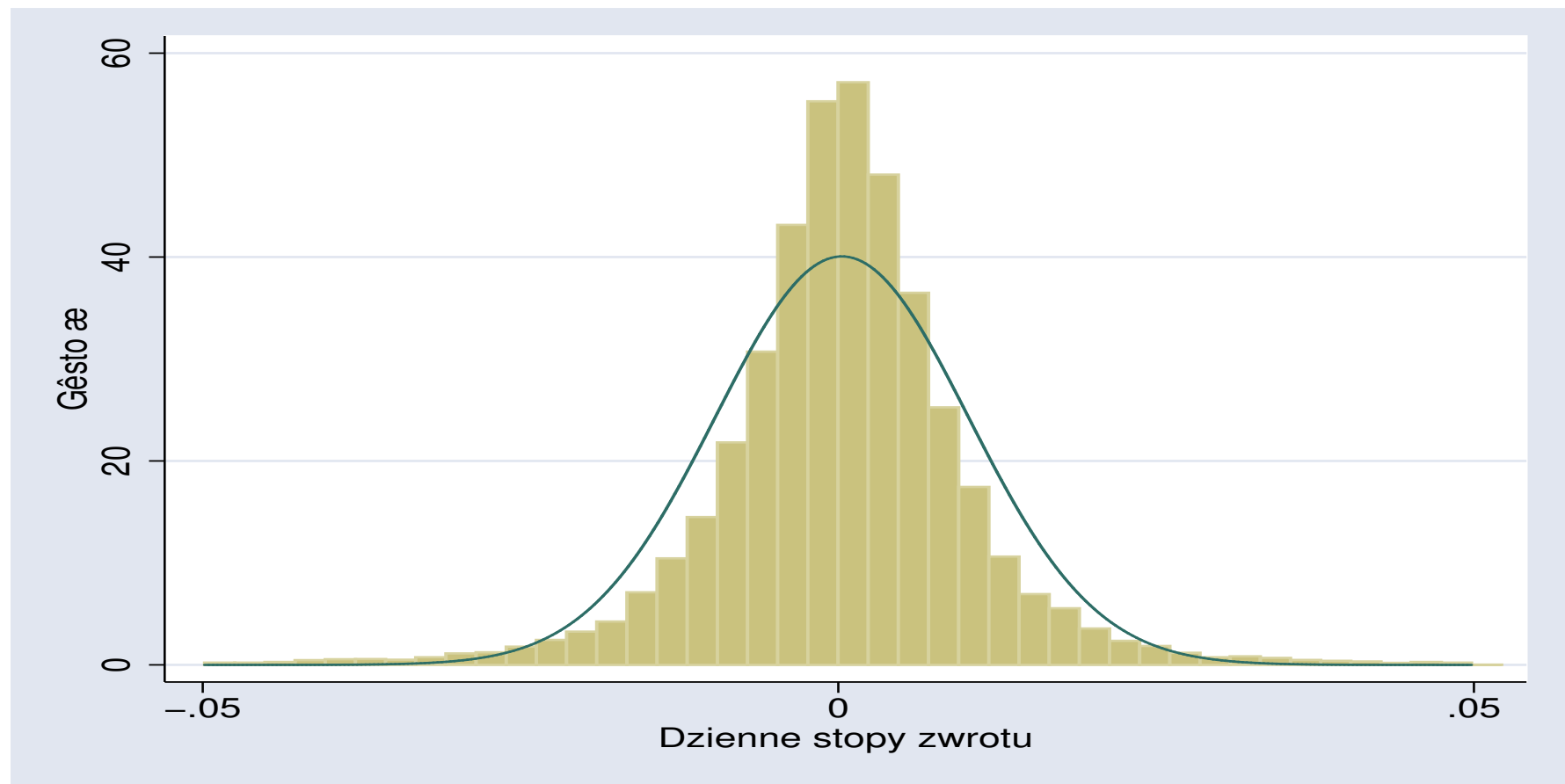
- Często zakłada się, że szeregi czasowe wykazują autokorelację ale są homoskedastyczne
- W rzeczywistości jednak często wariancja zmienia się w czasie
- Dobrym przykładem są tutaj dane z rynków finansowych
- Gdy zaczęto je badać zauważono, że szeregi te (np. ceny akcji, kursy walut etc.)
 1. zachowują się w przybliżeniu zgodnie z modelem błędzenia przypadkowego z dryfem
 2. w rozkładach pierwszych różnic obserwuje się silną leptokurtozę

- * zbyt grube ogony rozkładu
- * zbyt wysoka koncentracja wokół zera

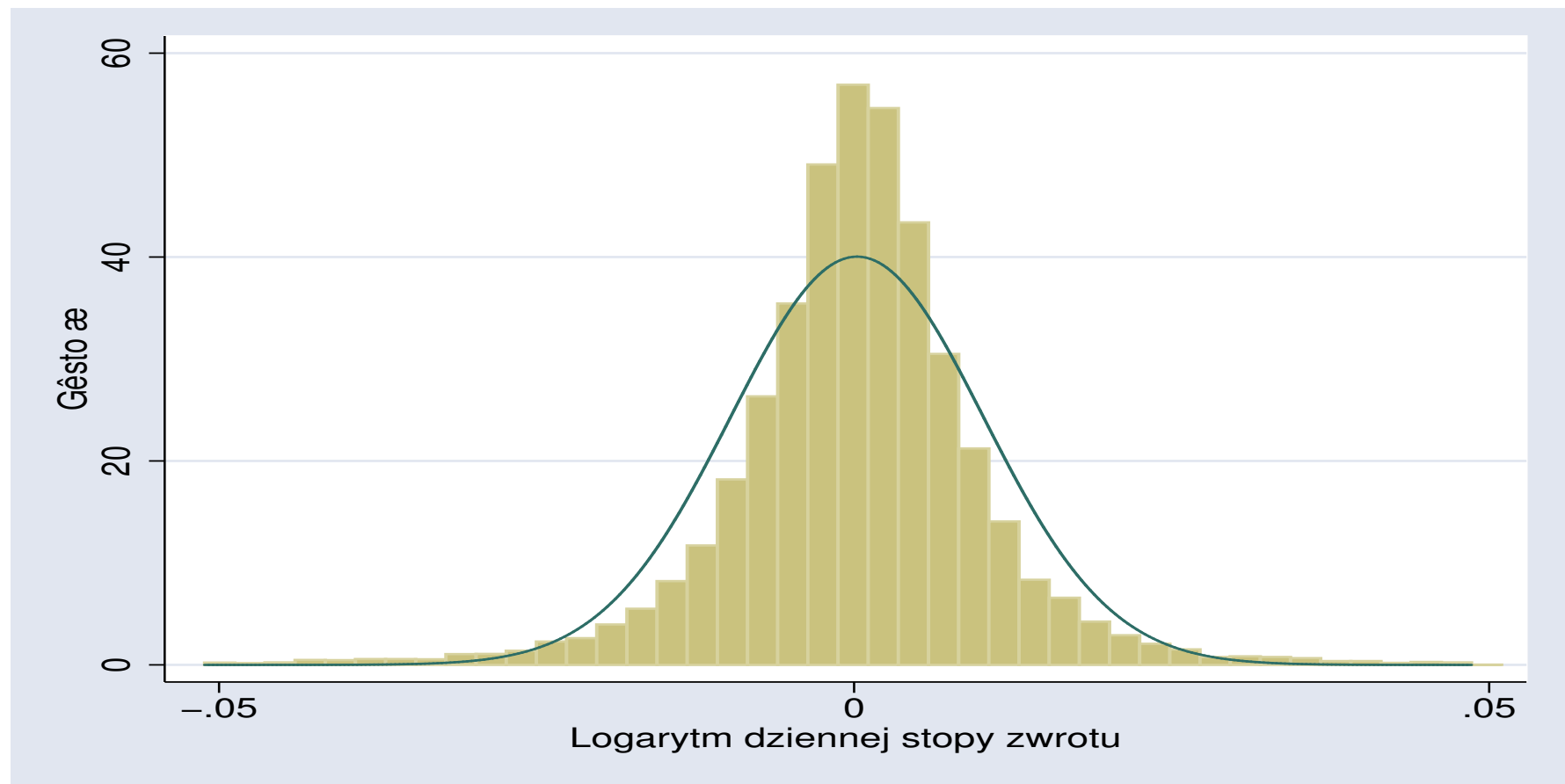
- Indeks Dow Jones 1928-2005



- Histogram zwrotów z Dow Jones 1928-2005



- Histogram logarytmów zwrotów z Dow Jones 1928-2005



- Wynik testu Jarque-Berra na normalność rozkładu logarytmów zwrotów z Dow-Jones

Skewness/Kurtosis tests for Normality				
Variable	Pr(Skewness)	Pr(Kurtosis)	----- joint ----- chi2(2)	Prob>chi2
lreturn	0.000	0.000	6413.84	0.0000

- Wielkość współczynnika asymetrii i kurtozy:

variable	skewness	kurtosis
lreturn	-.666518	29.55787

- Hipoteza o normalności rozkładu logarytmów zwrotów z Dow Jonesa został bardzo silnie odrzucona
- Z wyliczonych z próby wartości współczynnika asymetrii i kurtozy

wnioskujemy, że główną przyczyną odrzucenia tej hipotezy była silna kurtoza

- Wyniki testu Dickey-Fullera dla logarytmu dziennej stopy zwrotu (regresja ze stałą)

Dickey-Fuller test for unit root Number of obs = 19236

		----- Z(t) has t-distribution -----		
	Test	1% Critical	5% Critical	10% Critical
	Statistic	Value	Value	Value

Z(t)	0.710	-2.327	-1.645	-1.282

p-value for Z(t) = 0.7611				

Dickey-Fuller test for unit root Number of obs = 19235

----- Interpolated Dickey-Fuller -----				
	Test Statistic	1% Critical Value	5% Critical Value	10% Critical Value
Z(t)	-134.572	-3.430	-2.860	-2.570
MacKinnon approximate p-value for Z(t) = 0.0000				

- H_0 o istnieniu pierwiastka jednostkowego w logarytmach indeksu Dow Jones nie został odrzucony
- H_0 o istnieniu pierwiastka jednostkowego w pierwszych różnicach logarytmu indeksu Dow Jones została odrzucona
- Wniosek: $\log \left(\frac{p_t}{p_{t-1}} \right) = \Delta \log (p_t) \sim I(0), p_t \sim I(1)$

- Modelowanie $\log\left(\frac{p_t}{p_{t-1}}\right)$ w latach 1928-2005

Source	SS	df	MS	Number of obs = 19226		
Model	.008681166	10	.000868117	F(10, 19215) = 6.62		
Residual	2.52068163	19215	.000131183	Prob > F = 0.0000		
Total	2.52936279	19225	.000131566	R-squared = 0.0034		
				Adj R-squared = 0.0029		
				Root MSE = .01145		

lreturn	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lreturn						
L1	.0301748	.0072139	4.18	0.000	.0160349	.0443148
L2	-.0187431	.007215	-2.60	0.009	-.032885	-.0046012
L3	.0070471	.0072152	0.98	0.329	-.0070953	.0211896
L4	.0139131	.0072149	1.93	0.054	-.0002288	.0280549
L5	.0086085	.0072127	1.19	0.233	-.0055289	.0227459
L6	-.0282714	.0072126	-3.92	0.000	-.0424088	-.014134
L7	.0060472	.0072148	0.84	0.402	-.0080945	.0201889
L8	.0166248	.0072149	2.30	0.021	.002483	.0307666
L9	.0253229	.0072146	3.51	0.000	.0111816	.0394641

	L10		.0064811	.0072136	0.90	0.369	- .0076582	.0206203
_cons			.0001796	.0000827	2.17	0.030	.0000175	.0003418

- Współczynniki wychodzą istotne: czy tak zbudowany model można wykorzystać do zarabiania na NYSE?

- Wyniki regresji dla lat 1980-2005

Source	SS	df	MS	Number of obs = 6404		
Model	.001840057	10	.000184006	F(10, 6393) = 1.60		
Residual	.734122305	6393	.000114832	Prob > F = 0.0992		
Total	.735962362	6403	.00011494	R-squared = 0.0025		
				Adj R-squared = 0.0009		
				Root MSE = .01072		

lreturn	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lreturn						
L1	.0114255	.012507	0.91	0.361	-.0130924	.0359435
L2	-.0379114	.0125089	-3.03	0.002	-.0624331	-.0133898
L3	-.0159845	.0125178	-1.28	0.202	-.0405236	.0085546
L4	-.0185163	.0125189	-1.48	0.139	-.0430575	.0060248
L5	.0084503	.0125211	0.67	0.500	-.0160952	.0329957
L6	-.0023308	.0125208	-0.19	0.852	-.0268757	.0222141
L7	-.0156891	.0125187	-1.25	0.210	-.04023	.0088518
L8	-.0032441	.0125199	-0.26	0.796	-.0277872	.0212991
L9	.0002451	.0125104	0.02	0.984	-.0242794	.0247696
L10	.0023473	.0125098	0.19	0.851	-.0221761	.0268707

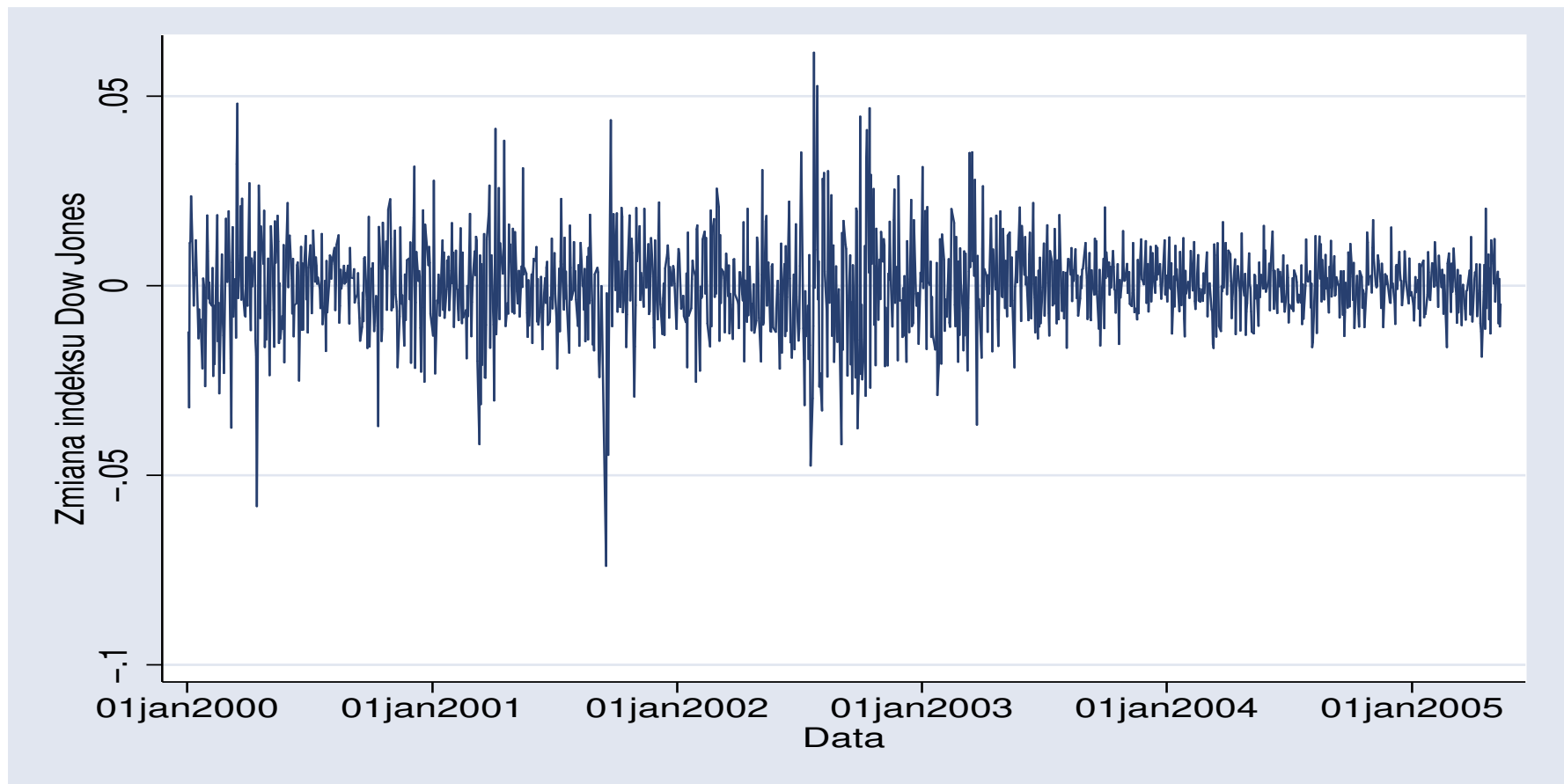
_cons		.0004171	.0001349	3.09	0.002	.0001527	.0006815

- Dla ostatnich 25 lat nie da się niestety zaobserwować żadnej istotnej korelacji!
- Czy więc zwroty z Dow Jones są białym szumem: ciągiem nieskorelowanych ε_t ?
- Dlaczego w zwrotach występuje leptokurtoza?

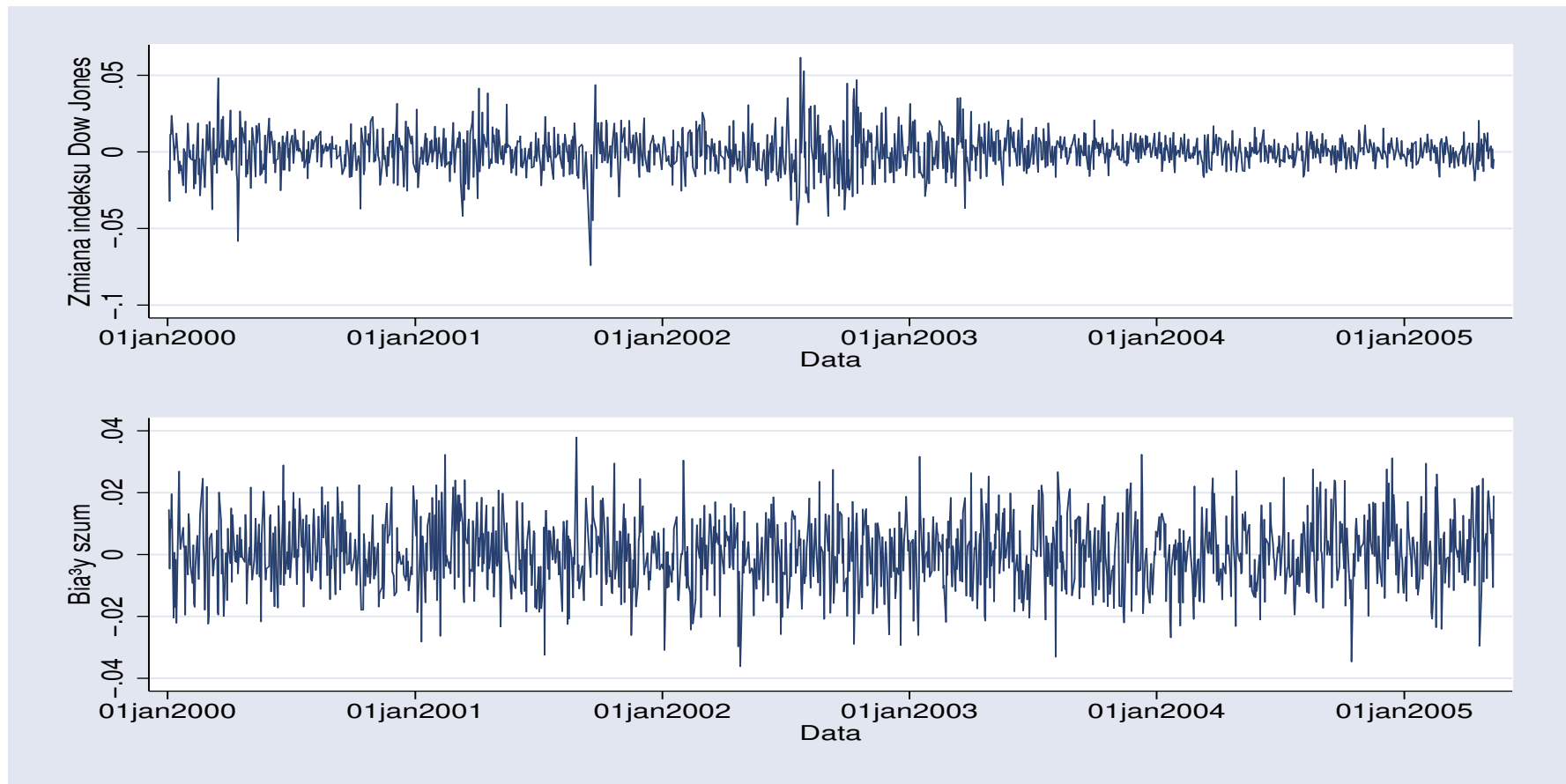
Portmanteau test for white noise

Portmanteau (Q) statistic =	81.8469
Prob > chi2(40) =	0.0001

- Zwroty z Dow Jones



- Zwroty z Dow Jones i biały szum



- Najprawdopodobniej przyczyną występowania leptokurtozy jest zjawisko grupowania obserwacji o dużej wariancji
- Oczekiwanej wielkości zwrotów nie da się przewidzieć (z pewnym zastrzeżeniem do którego jeszcze do tego wrócimy)
- Do pewnego stopnia przewidywalne jest jednak wariancja
- Modelowanie zmian wariancji jest w praktyce bardzo ważne ponieważ jest to parametr, od którego zależą ceny instrumentów pochodnych (np. opcji)

Warunkowa autoregresyjna heteroskedastyczność *ARCH* (*q*)

- Model *ARCH* (1) (**A**utoregressive **C**onditional **H**eteroskedasticity)

$$\begin{aligned}y_t &= \mathbf{x}_t \boldsymbol{\beta} + \varepsilon_t \\ \varepsilon_t &= u_t \sqrt{\boldsymbol{\theta}_0 + \boldsymbol{\theta}_1 \varepsilon_{t-1}^2}\end{aligned}$$

gdzie $u_t \sim N(0, 1)$, u_t jest niezależne od ε_{t-1} i $E(\varepsilon_t | \mathbf{x}_t, \varepsilon_{t-1}) = 0$

- Bazwarunkowa wariancja ε_t jest stała

$$\begin{aligned}\text{Var}(\varepsilon_t) &= E(\varepsilon_t^2) = E(u_t^2) E(\boldsymbol{\theta}_0 + \boldsymbol{\theta}_1 \varepsilon_{t-1}^2) \\ &= \boldsymbol{\theta}_0 + \boldsymbol{\theta}_1 \text{Var}(\varepsilon_{t-1}^2)\end{aligned}$$

- Jeśli proces generujący wariancje jest stacjonarny to $\text{Var}(\epsilon_t^2) = \text{Var}(\epsilon_{t-1}^2) = \dots$
- Podstawiając i rozwiązując dla $\text{Var}(\epsilon_t)$ otrzymujemy:

$$\text{Var}(\epsilon_t^2) = \frac{\theta_0}{1 - \theta_1}$$

- Założenia *KMRL* są spełnione - estymator *MNK* jest najlepszym liniowym i niobciążonym estymatorem β
- Z racji na występowanie warunkowej heteroskedastyczności istnieją jednak lepsze *nieliniowe* estymatory β
- Wariancja nie jest jednak stała warunkowo:

$$\text{Var}(\epsilon_t | \epsilon_{t-1}) = E(\epsilon_t^2 | \epsilon_{t-1}) = \theta_0 + \theta_1 \epsilon_{t-1}^2$$

- Ogólna postać modelu $ARCH(q)$

$$y_t = x_t \beta + \varepsilon_t$$

$$\varepsilon_t = u_t \sigma_t$$

$$\sigma_t^2 = \theta_0 + \theta_1 \varepsilon_{t-1}^2 + \dots + \theta_q \varepsilon_{t-q}^2$$

- Zauważmy, że model opisujący wariancję jest w gruncie rzeczy procesem $MA(q)$
- Test na występowanie warunkowej heteroskedastyczności rzędu q :
 1. przeprowadzamy regresję y_t na x_t
 2. uzyskane reszty tej regresji podniesimy do kwadratu
 3. przeprowadzamy regresję e_t^2 na $e_{t-1}^2, \dots, e_{t-q}^2$
 4. liczymy statystykę $LM = TR^2 \sim \chi_q^2$

- Wynik testu LM na heteroskedastyczność warunkową dla logarytmów zwrotów z Dow Jones'a.

LM test for autoregressive conditional heteroskedasticity (ARCH)

lags(p)	chi2	df	Prob > chi2
1	52.980	1	0.0000
2	174.122	2	0.0000
3	189.930	3	0.0000

H0: no ARCH effects vs. H1: ARCH(p) disturbance

- Wynik testu: bardzo silnie odrzucamy hipotezę o braku warunkowej heteroskedystyczności

- Model *ARCH* (3)

ARCH family regression

Sample: -6404 to -1	Number of obs	=	6404
	Wald chi2(.)	=	.
Log likelihood = 20547.74	Prob > chi2	=	.

		Coef.	OPG Std. Err.	z	P> z	[95% Conf. Interval]	
lreturn							
_cons		.0006031	.0001107	5.45	0.000	.0003861	.0008201
ARCH							
arch							
	L1	.1217862	.0089875	13.55	0.000	.104171	.1394015
	L2	.1316005	.010769	12.22	0.000	.1104936	.1527074
	L3	.1520855	.0119788	12.70	0.000	.1286074	.1755635
_cons		.0000633	1.00e-06	63.25	0.000	.0000613	.0000652

- Test LR na łączną nieistotność parametrów

likelihood-ratio test
(Assumption: . nested in model)

LR chi2(3) = 1176.93
Prob > chi2 = 0.0000

Model uogólnionej warunkowej heteroskedastyczności

$GARCH(p, q)$

- Model $GARCH(p, q)$ (**G**eneralized **A**utoregressive **C**onditional **H**eteroskedasticity) ma dla wariancji postać podobną do modelu $ARMA(p, q)$

$$y_t = x_t\beta + \varepsilon_t$$

$$\varepsilon_t = u_t\sigma_t$$

$$\sigma_t^2 = \alpha_1\sigma_{t-1}^2 + \alpha_2\sigma_{t-2}^2 + \dots + \alpha_p\sigma_{t-p}^2 + \theta_0 + \theta_1\varepsilon_{t-1}^2 + \dots + \theta_q\varepsilon_{t-q}^2$$

- Wyniki estymacji modelu $GARCH(3, 3)$

ARCH family regression

Sample:	-6404 to -1	Number of obs	=	6404
		Wald chi2(.)	=	.
Log likelihood =	20793.3	Prob > chi2	=	.

			OPG				
lreturn		Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
-----+-----							
lreturn							
_cons		.000574	.0001083	5.30	0.000	.0003617	.0007863
-----+-----							
ARCH							
arch							
	L1	.0748263	.0041613	17.98	0.000	.0666704	.0829823
	L2	.0807364	.0045208	17.86	0.000	.0718758	.089597
	L3	.0642312	.0032802	19.58	0.000	.057802	.0706604
garch							
	L1	- .008597	.0131908	-0.65	0.515	- .0344505	.0172565
	L2	- .103563	.0164509	-6.30	0.000	- .1358061	- .0713198
	L3	.8596985	.014648	58.69	0.000	.8309889	.888408
_cons		4.54e-06	5.37e-07	8.47	0.000	3.49e-06	5.60e-06

Model ARCH w średnich

- Model *ARCHM* (*ARCH* in **M**eans) ma następującą postać:

$$y_t = x_t\beta + \delta\sigma_t^2 + \varepsilon_t$$

$$\varepsilon_t = u_t\sigma_t$$

$$\sigma_t^2 = \theta_0 + \theta_1\varepsilon_{t-1}^2 + \dots + \theta_q\varepsilon_{t-q}^2$$

- W przypadku tego modelu zmiany wariancji wpływają także na średnią
- Model ten ma zastosowania w finansach, gdzie zakłada się, że na ceny akcji wpływa także ryzyko

- Jeśli ryzyko zmienia się w czasie to także powinno to wpływać na ceny akcji
- Parametr δ jest więc związany z awersją do ryzyka
- Ponieważ wyższe ryzyko powinno się wiązać z wyższymi zwrotami, więc parametr ten powinien być dodatni

- Wyniki estymacji modelu *ARCHM* (3)

Log likelihood = 20551.6 Prob > chi2 = 0.0000

		OPG					
		Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
		-----+-----					
lreturn							
_cons		- .0000286	.000202	-0.14	0.887	- .0004246	.0003674
		-----+-----					
ARCHM							
sigma2		7.041259	1.671361	4.21	0.000	3.765453	10.31707
		-----+-----					
ARCH							
arch							
	L1	.119549	.008333	14.35	0.000	.1032165	.1358814
	L2	.1356328	.011151	12.16	0.000	.1137773	.1574884
	L3	.14759	.0122571	12.04	0.000	.1235666	.1716135
_cons		.0000633	1.00e-06	63.32	0.000	.0000614	.0000653

- Istotnie $\delta = 7.041259 > 0$
- Z modelu tego mogłoby wynikać, że przychody są jednak częściowo przewidywalne - jednak jedynie w tym sensie, że w okresie wysokiego ryzyka są nieco wyższe!