

Prognozowanie i symulacje

Wykład 6

Własności estymatora KMNK

Dla modelu:

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \varepsilon_t$$

Przy spełnionych założeniach:

$$E(\varepsilon) = 0 \quad \text{var}(\varepsilon)^2 = \sigma_\varepsilon^2 \quad \text{cov}(\varepsilon_t, \varepsilon_s) = 0, \quad t \neq s$$

Najlepszym, nieobciążonym estymatorem liniowym nieznanych parametrów β jest estymator metody najmniejszych kwadratów

Autokorelacja

Autokorelacja pierwszego rzędu: $\varepsilon_t = \rho_1 \varepsilon_{t-1} + \xi_t$, $|\rho_1| < 1$

gdzie ξ jest procesem czysto losowym.

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \xi_t,$$

$$\varepsilon_{t-1} = \rho_1 \varepsilon_{t-2} + \xi_{t-1}$$

$$\varepsilon_{t-2} = \rho_1 \varepsilon_{t-3} + \xi_{t-2}$$

$$\varepsilon_{t-3} = \rho_1 \varepsilon_{t-4} + \xi_{t-3}$$

Autokorelacja składnika losowego

$$\begin{aligned}\varepsilon_t &= \rho_1(\rho_1(\rho_1(\rho_1\varepsilon_{t-4} + \xi_{t-3}) + \xi_{t-2}) + \xi_{t-1}) + \xi_t \\ &= \rho_1^4\varepsilon_{t-4} + \rho_1^3\xi_{t-3} + \rho_1^2\xi_{t-2} + \rho_1\xi_{t-1} + \xi_t\end{aligned}$$

$$\begin{aligned}\varepsilon_t &= \xi_t + \rho_1\xi_{t-1} + \rho_1^2\xi_{t-2} + \rho_1^3\xi_{t-3} + \rho_1^4\varepsilon_{t-4} + \dots \\ &= \sum_{p=1}^{\infty} \rho_1^p \xi_{t-p}\end{aligned}$$

$$\Omega = \begin{bmatrix} 1 & \rho_1 & \dots & \rho_1^{n-1} \\ \rho_1 & 1 & \dots & \rho_1^{n-2} \\ \dots & \dots & \dots & \dots \\ \rho_1^{n-1} & \rho_1^{n-2} & \dots & 1 \end{bmatrix}$$

Autokorelacja składnika losowego

- Macierz wariancji kowariancji składnika losowego bez autokorelacji

$$\text{var}(\varepsilon_t) = \sigma^2 I, \quad I = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

- Macierz wariancji kowariancji składnika losowego z autokorelacją

$$\text{var}(\varepsilon_t) = \sigma^2 \Omega, \quad \Omega = \begin{bmatrix} 1 & \rho_1 & \dots & \rho_1^{n-1} \\ \rho_1 & 1 & \dots & \rho_1^{n-2} \\ \dots & \dots & \dots & \dots \\ \rho_1^{n-1} & \rho_1^{n-2} & \dots & 1 \end{bmatrix}$$

Estymator KMNK i wariancja estymatora

$$\begin{aligned}\hat{\beta} &= (X^T X)^{-1} X^T Y = (X^T X)^{-1} X^T (X \beta + \varepsilon) \\ &= (X^T X)^{-1} X^T X \beta + (X^T X)^{-1} X^T \varepsilon \\ &= \beta + (X^T X)^{-1} X^T \varepsilon\end{aligned}$$

$$\begin{aligned}\text{var}(\hat{\beta}) &= E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)^T] \\ &= E[(X^T X)^{-1} X^T \varepsilon \varepsilon^T X (X^T X)^{-1}] \\ &= (X^T X)^{-1} X^T E(\varepsilon \varepsilon^T) X (X^T X)^{-1}\end{aligned}$$

wariancja – kowariancja składnika losowego

Estymator KMNK i wariancja estymatora

$$E(\varepsilon\varepsilon^T) = \text{var}(\varepsilon_t) = \sigma^2 I$$

$$\begin{aligned}\text{var}(\hat{\beta}) &= E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)^T] \\ &= E[(X^T X)^{-1} X^T \varepsilon\varepsilon^T X (X^T X)^{-1}] \\ &= (X^T X)^{-1} X^T E(\varepsilon\varepsilon^T) X (X^T X)^{-1} = \sigma^2 (X^T X)^{-1}\end{aligned}$$

$$E(\varepsilon\varepsilon^T) = \text{var}(\varepsilon_t) = \sigma^2 \Omega$$

$$\begin{aligned}\text{var}(\hat{\beta}) &= E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)^T] \\ &= E[(X^T X)^{-1} X^T \varepsilon\varepsilon^T X (X^T X)^{-1}] \\ &= (X^T X)^{-1} X^T E(\varepsilon\varepsilon^T) X (X^T X)^{-1} \\ &= \sigma^2 (X^T X)^{-1} X^T \Omega X (X^T X)^{-1}\end{aligned}$$

Skutki autokorelacji

- Estymator KMNK w dalszym ciągu jest estymatorem nieobciążonym i zgodnym
- Estymator KMNK przystaje być estymatorem najefektywniejszym (przestaje być estymatorem BLUE)
- Estymator wariancja składnika losowego jest obciążony – przy autokorelacji dodatniej jest to obciążenie w dół
- Średnie błędy szacunku parametrów są niedoszacowane
- Statystyki t-Studenta są zawyżone
- W teście t-Studenta częściej odrzucana jest hipoteza zerowa

Rozwiązanie problemu

- Zmiana postaci funkcyjnej modelu
- Uzupełnienie zbioru zmiennych objaśniających
- Uogólniona metoda najmniejszych kwadratów
- Estymator Cochrana-Orcutta
- Estymator Prais-Winstena
- Prognozowanie w warunkach autokorelacji

Estymator uogólnionej metody najmniejszych kwadratów

$$\hat{\beta} = (X^T \Omega^{-1} X)^{-1} X^T \Omega^{-1} Y$$

- Nieznane wartości w macierzy Ω
- Wykorzystujemy „stosowalną” metodę UMNK

Metoda Cochrane'a - Orcutta

Zastosowanie KMNK do zmodyfikowanych zmiennych

$$y_t^* = y_t - \rho_1 y_{t-1}$$

$$x_t^* = x_t - \rho_1 x_{t-1}$$

Wartość współczynnika korelacji pochodzi z estymacji modelu (AR(1)):

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \xi_t, \quad |\rho_1| < 1$$

Metoda Cochrane'a – Orcutta korzysta z n-1 obserwacji, prowadzi to do przybliżonych ocen parametrów metody UMNK (takie podejście daje mniej efektywne wyniki), jednak w przypadku dużych prób oszacowania są zbieżne.

Transformacja Prais - Winstena

Transformacja pierwszych wartości zmiennej objaśnianej i zmiennych objaśniających:

$$y_1^* = y_1 \sqrt{1 - \rho_1^2}$$

$$x_1^* = x_1 \sqrt{1 - \rho_1^2}$$

Dalsze wartości szeregów ustalane są zgodnie z przekształceniem Cochrane'a – Orcutta

Metoda Cochrane'a – Orcutta – krok po kroku

Model 12: Estymacja KMNK, wykorzystane obserwacje 1999:1-2015:4 (N = 68)
Zmienna zależna (Y): SI

	współczynnik	błąd standardowy	t-Studenta	wartość p	
const	27417,7	6093,12	4,500	2,83e-05	***
PKB	0,518236	0,0195025	26,57	5,76e-037	***
Średn. aryt. zm. zależnej	181903,4	Odch. stand. zm. zależnej	51063,10		
Suma kwadratów reszt	1,49e+10	Błąd standardowy reszt	15041,93		
Wsp. determ. R-kwadrat	0,914521	Skorygowany R-kwadrat	0,913225		
F(1, 66)	706,1151	Wartość p dla testu F	5,76e-37		
Logarytm wiarygodności	-749,5374	Kryt. inform. Akaike'a	1503,075		
Kryt. bayes. Schwarz	1507,514	Kryt. Hannana-Quinna	1504,834		
Autokorel. reszt - rho1	-0,329756	Stat. Durbina-Watsona	2,460178		

Test LM na autokorelację rzędu 1 -

Hipoteza zerowa: brak autokorelacji składnika losowego

Statystyka testu: LMF = 7,87251

z wartością p = P(F(1, 65) > 7,87251) = 0,00661486

Z tego modelu zapisujemy reszty i szacujemy model pomocniczy

$$\hat{\varepsilon}_t = \rho_1 \hat{\varepsilon}_{t-1} + \xi_t$$

Metoda Cochrane'a – Orcutta – krok po kroku

Model 13: Estymacja KMNK, wykorzystane obserwacje 1999:2-2015:4 (N = 67)
Zmienna zależna (Y): uhat12

	współczynnik	błąd standardowy	t-Studenta	wartość p	
-----	-----	-----	-----	-----	-----
uhat12_1	-0,329756	0,124729	-2,644	0,0102	**
Średn.aryt.zm.zależnej	41,63880	Odch.stand.zm.zależnej	15037,95		
Suma kwadratów reszt	1,35e+10	Błąd standardowy reszt	14299,86		
Niecentrowany R-kwadr.	0,095762	Centrowany R-kwadrat	0,095755		
F(1, 66)	6,989605	Wartość p dla testu F	0,010232		
Logarytm wiarygodności	-735,6215	Kryt. inform. Akaike'a	1473,243		
Kryt. bayes. Schwarz	1475,448	Kryt. Hannana-Quinna	1474,115		
Autokorel.reszt - rho1	-0,033929	Stat. Durbina-Watsona	1,939682		

Ustalamy skorygowane wartości zmiennych – y i x

$$SI_t^* = y_t - (-0,329756) SI_{t-1}$$

$$PKB_t^* = x_t - (-0,329756) PKB_{t-1}$$

Szacujemy model KMNK z wykorzystaniem skorygowanych zmiennych

Metoda Cochrane'a – Orcutta – krok po kroku

Model 14: Estymacja KMNK, wykorzystane obserwacje 1999:2-2015:4 (N = 67)
Zmienna zależna (Y): **SI1**

	współczynnik	błąd standardowy	t-Studenta	wartość p	
const	29232,7	5965,93	4,900	6,69e-06	***
PKB1	0,537038	0,0143535	37,42	1,08e-045	***
Średn.aryt.zm.zależnej	242774,3	Odch.stand.zm.zależnej	66996,68		
Suma kwadratów reszt	1,31e+10	Błąd standardowy reszt	14220,74		
Wsp. determ. R-kwadrat	0,955628	Skorygowany R-kwadrat	0,954946		
F(1, 65)	1399,895	Wartość p dla testu F	1,08e-45		
Logarytm wiarygodności	-734,7382	Kryt. inform. Akaike'a	1473,476		
Kryt. bayes. Schwarza	1477,886	Kryt. Hannana-Quinna	1475,221		
Autokorel.reszt - rho1	-0,141743	Stat. Durbina-Watsona	2,087910		

Test LM na autokorelację rzędu 1 -

Hipoteza zerowa: brak autokorelacji składnika losowego

Statystyka testu: LMF = 1,18828

z wartością p = $P(F(1, 64) > 1,18828) = 0,279763$

ITER	RHO	ESS
1	-0,32976	1,31449e+010
2	-0,39841	1,30788e+010
3	-0,40367	1,30784e+010
4	-0,40401	1,30784e+010
5	-0,40403	1,30784e+010
6	-0,40403	1,30784e+010

Model 15: Estymacja Cochrane-Orcutt, wykorzystane obserwacje 1999:2-2015:4 (N = 67)

Zmienna zależna (Y): SI

rho = -0,404033

	współczynnik	błąd standardowy	t-Studenta	wartość p	
const	21331,7	4242,38	5,028	4,14e-06	***
PKB	0,539311	0,0135835	39,70	2,68e-047	***

Podstawowe statystyki dla danych quasi-różnicowanych (rho):

Średn.aryt.zm.zależnej	183121,0	Odch.stand.zm.zależnej	50444,15
Suma kwadratów reszt	1,31e+10	Błąd standardowy reszt	14184,70
Wsp. determ. R-kwadrat	0,922405	Skorygowany R-kwadrat	0,921211
F(1, 65)	1576,354	Wartość p dla testu F	2,68e-47
Autokorel.reszt - rho1	-0,092692	Stat. Durбина-Watsona	2,002183

Wartość współczynnika autkorelacji ustalana jest aby zminimalizować resztową sumę kwadratów (ESS)

$$SI_t^* = y_t - (-0,40403)SI_{t-1} \quad PKB_t^* = x_t - (-0,40403)PKB_{t-1}$$

	współczynnik	błąd standardowy	t-Studenta	wartość p	
const	29950,5	5956,44	5,028	4,14e-06	***
PKB2	0,539311	0,0135835	39,70	2,68e-047	***
Średn.aryt.zm.zależnej	256211,1	Odch.stand.zm.zależnej	70737,41		KMNK
Suma kwadratów reszt	1,31e+10	Błąd standardowy reszt	14184,70		
Wsp. determ. R-kwadrat	0,960399	Skorygowany R-kwadrat	0,959789		
F(1, 65)	1576,353	Wartość p dla testu F	2,68e-47		
Logarytm wiarygodności	-734,5682	Kryt. inform. Akaike'a	1473,136		
Kryt. bayes. Schwarz	1477,546	Kryt. Hannana-Quinna	1474,881		
Autokorel.reszt - rho1	-0,092692	Stat. Durbina-Watsona	2,002183		

	współczynnik	błąd standardowy	t-Studenta	wartość p	
const	21331,7	4242,38	5,028	4,14e-06	***
PKB	0,539311	0,0135835	39,70	2,68e-047	***
Podstawowe statystyki dla danych quasi-różnicowanych (rho):					Cochrane - Orcutt
Średn.aryt.zm.zależnej	183121,0	Odch.stand.zm.zależnej	50444,15		
Suma kwadratów reszt	1,31e+10	Błąd standardowy reszt	14184,70		
Wsp. determ. R-kwadrat	0,922405	Skorygowany R-kwadrat	0,921211		
F(1, 65)	1576,354	Wartość p dla testu F	2,68e-47		
Autokorel.reszt - rho1	-0,092692	Stat. Durbina-Watsona	2,002183		

Prognozowanie na podstawie modelu z autokorelacją

Model 4: Estymacja KMNK, wykorzystane obserwacje 1999:1-2015:4 (N = 68)

Zmienna zależna (Y): PKB

	współczynnik	błąd standardowy	t-Studenta	wartość p	
const	139216	3004,32	46,34	2,17e-050	***
time	4605,29	75,7197	60,82	1,14e-057	***
v1	-16222,6	2572,52	-6,306	3,18e-08	***
v2	-6300,20	2570,29	-2,451	0,0170	**
v3	-5784,94	2570,29	-2,251	0,0279	**
Średn.aryt.zm.zależnej	298099,0	Odch.stand.zm.zależnej	94227,13		
Suma kwadratów reszt	9,43e+09	Błąd standardowy reszt	12235,70		
Wsp. determ. R-kwadrat	0,984145	Skorygowany R-kwadrat	0,983138		
F(4, 63)	977,6150	Wartość p dla testu F	6,46e-56		
Logarytm wiarygodności	-733,9149	Kryt. inform. Akaike'a	1477,830		
Kryt. bayes. Schwarz	1488,927	Kryt. Hannana-Quinna	1482,227		
Autokorel.reszt - rho1	0,667916	Stat. Durbina-Watsona	0,657088		

Test LM na autokorelację rzędu 1 -

Hipoteza zerowa: brak autokorelacji składnika losowego

Statystyka testu: LMF = 43,3884

z wartością p = $P(F(1, 62) > 43,3884) = 1,10929e-008$

Test Breuscha-Godfrey na autokorelację rzędu pierwszego
 Estymacja KMNK, wykorzystane obserwacje 1999:1-2015:4 (N = 68)
 Zmienna zależna (Y): uhat

	współczynnik	błąd standardowy	t-Studenta	wartość p	
const	-633,863	2324,83	-0,2726	0,7860	
time	26,5956	58,6831	0,4532	0,6520	
v1	890,954	1993,58	0,4469	0,6565	
v2	-270,389	1987,69	-0,1360	0,8922	
v3	-296,985	1987,78	-0,1494	0,8817	
uhat_1	0,674858	0,102453	6,587	1,11e-08	***

Wsp. determ. R-kwadrat = 0,411700

Statystyka testu: LMF = 43,388417,
 z wartością p = $P(F(1,62) > 43,3884) = 1,11e-008$

Statystyka testu: $TR^2 = 27,995604$,
 z wartością p = $P(\text{Chi-kwadrat}(1) > 27,9956) = 1,22e-007$

Ljung-Box $Q' = 26,4404$,
 z wartością p = $P(\text{Chi-kwadrat}(1) > 26,4404) = 2,72e-007$

Iteracyjne szacowanie parametru rho...

ITER	RHO	ESS
1	0,66792	5,18234e+009
2	0,67326	5,18204e+009
3	0,67380	5,18204e+009
4	0,67386	5,18204e+009
5	0,67386	5,18204e+009

Model 5: Estymacja Cochrane-Orcutt, wykorzystane obserwacje 1999:2-2015:4 (N = 67)

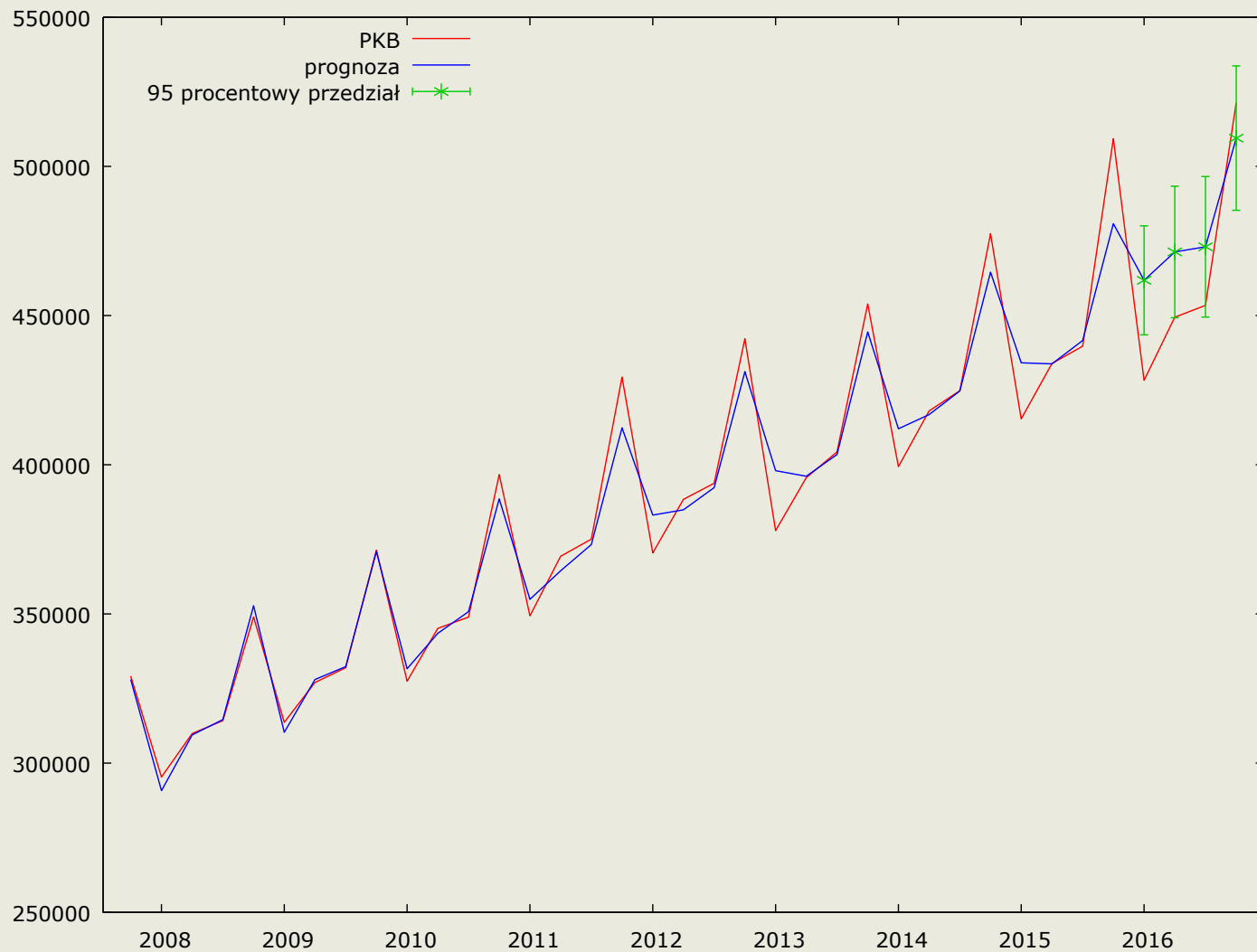
Zmienna zależna (Y): PKB

rho = 0,673863

	współczynnik	błąd standardowy	t-Studenta	wartość p	
const	133428	7408,05	18,01	2,55e-026	***
time	4762,49	177,205	26,88	7,19e-036	***
v1	-16222,0	1484,84	-10,93	4,37e-016	***
v2	-6260,87	1460,46	-4,287	6,44e-05	***
v3	-5770,80	1463,43	-3,943	0,0002	***

Podstawowe statystyki dla danych quasi-różnicowanych (rho):

Średn.aryt.zm.zależnej	300368,1	Odch.stand.zm.zależnej	93047,61
Suma kwadratów reszt	5,18e+09	Błąd standardowy reszt	9142,277
Wsp. determ. R-kwadrat	0,990931	Skorygowany R-kwadrat	0,990346
F(4, 62)	298,3721	Wartość p dla testu F	9,66e-40
Autokorel.reszt - rho1	-0,286579	Stat. Durbina-Watsona	2,320392



2016Q1	428292.3	461813.8	9142.28	1.98%	-33521.5	-7.83%
2016Q2	449436.9	471320.4	11024.28	2.34%	-21883.5	-4.87%
2016Q3	453415.1	473057.4	11780.03	2.49%	-19642.3	-4.33%
2016Q4	521388.8	509475.4	12107.65	2.38%	11913.4	2.28%

Iteracyjne szacowanie parametru rho...

ITER	RHO	ESS
1	0,66792	5,41763e+009
2	0,67255	5,41548e+009
3	0,67273	5,41541e+009
4	0,67274	5,41540e+009
5	0,67274	5,41540e+009

Model 6: Estymacja Prais-Winsten, wykorzystane obserwacje 1999:1-2015:4 (N = 68)

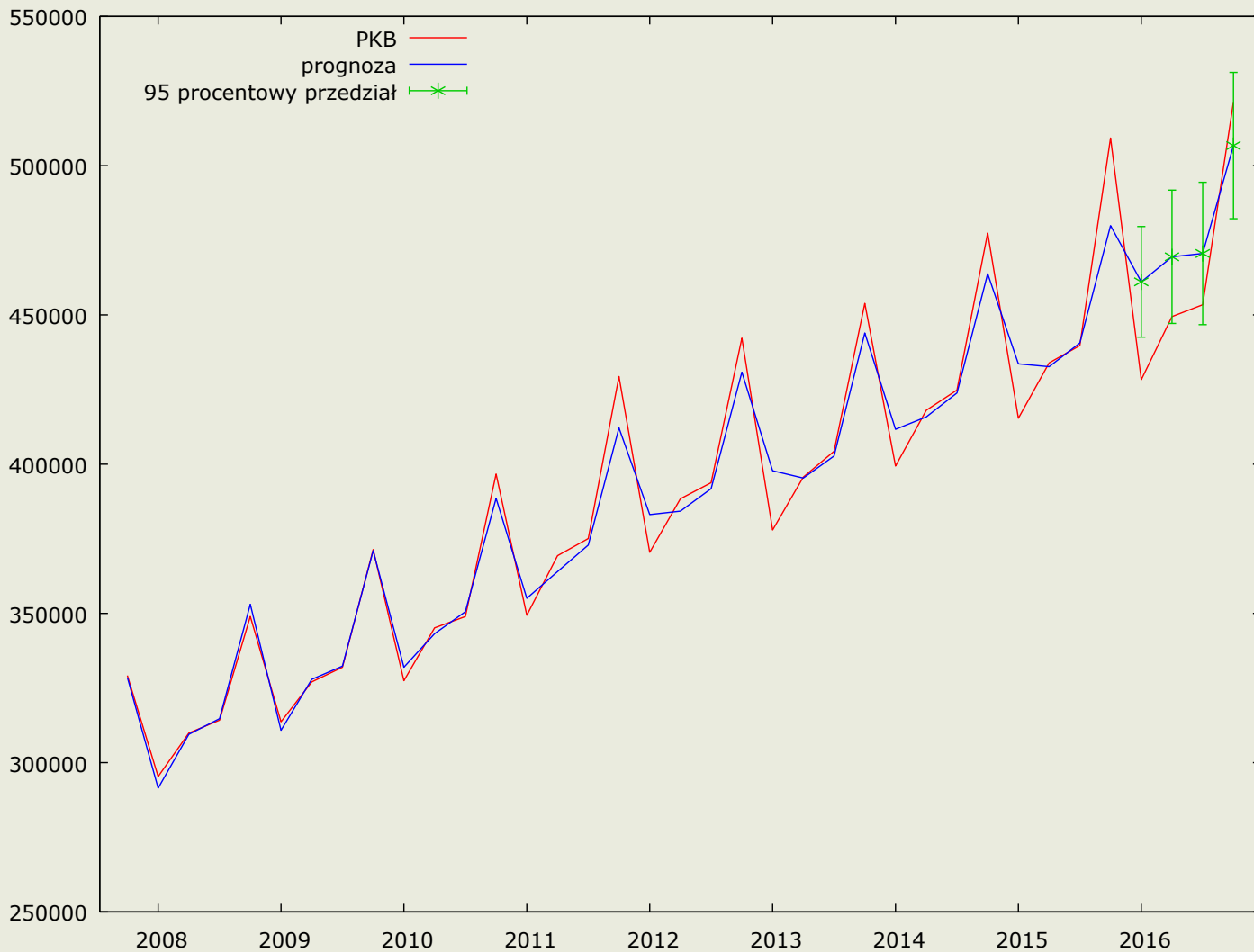
Zmienna zależna (Y): PKB

rho = 0,67274

	współczynnik	błąd standardowy	t-Studenta	wartość p	
const	139669	6464,72	21,60	7,90e-031	***
time	4630,89	160,496	28,85	5,07e-038	***
v1	-15933,8	1496,41	-10,65	1,02e-015	***
v2	-6337,80	1481,17	-4,279	6,51e-05	***
v3	-5944,33	1481,17	-4,013	0,0002	***

Podstawowe statystyki dla danych quasi-różnicowanych (rho):

Średn. aryt. zm. zależnej	298099,0	Odch. stand. zm. zależnej	94227,13
Suma kwadratów reszt	5,42e+09	Błąd standardowy reszt	9271,396
Wsp. determ. R-kwadrat	0,990924	Skorygowany R-kwadrat	0,990348
F(4, 63)	290,6016	Wartość p dla testu F	7,73e-40
Autokorel. reszt - rho1	-0,250111	Stat. Durbina-Watsona	2,229900



2016Q1	428292.3	461083.1	9271.4	2.01%	-32790.8	-7.66%
2016Q2	449436.9	469479.4	11174.17	2.38%	-20042.5	-4.46%
2016Q3	453415.1	470581.3	11936.03	2.54%	-17166.2	-3.79%
2016Q4	521388.8	506733.6	12265.29	2.42%	14655.2	2.81%

Dziękuję za uwagę!
Aleksandra.Kordalska@pg.edu.pl