

Przykładowy model ekonometryczny

Sebastian Michalski

Spis treści

1	Postać modelu	3
1.1	Dane	4
1.2	Graficzna prezentacja danych	5
2	Dobór zmiennych objaśniających metodą Hellwig’a	6
3	Oszacowanie parametrów strukturalnych	8
3.1	Interpretacja oszacowanych parametrów strukturalnych	8
3.2	Interpretacja błędów szacunku parametrów	8
4	Badanie własności koincydencji	9
5	Normalność rozkładu składnika losowego	9
5.1	Test Jarque-Bera	9
6	Autokorelacja składnika losowego	10
6.1	Test Durbin’a-Watson’a	10
6.2	Test mnożnika Lagrange’a	11
7	Heteroskedastyczność składnika losowego	11
7.1	Test White’a	11
8	Liniowość modelu – test liczby serii	12
9	Współliniowość zmiennych objaśniających	13
10	Istotność zmiennych objaśniających	
	– test t-Studenta	13
11	Współczynnik determinacji	14
11.1	Skorygowany współczynnik determinacji	15
12	Prognozy	15
12.1	Prognoza zmiennej endogenicznej na podstawie modelu trendu linowego	15
12.2	Prognoza zmiennej endogenicznej na podstawie modelu trendu wielomianowego	16
12.2.1	Ocena <i>ex ante</i> prognozy	16
12.2.2	Prognoza przedziałowa	17
12.3	Prognoza zmiennej endogenicznej na podstawie modelu ekono- metrycznego (warunkowa)	18
12.3.1	Prognoza zmiennych egzogenicznych na podstawie modeli trendu liniowego	18
12.3.2	Prognoza zmiennych egzogenicznych na podstawie modeli trendu wielomianowego	21
12.3.3	Ocena <i>ex post</i> prognozy	22
12.4	Prognoza na podstawie modelu prostego wyrównania wykładni- czego Brown’a	24

1 Postać modelu

$$Y_t = \alpha_0 + \alpha_1 X_{1t} + \alpha_2 X_{2t} + \alpha_3 X_{3t} + \epsilon_t, \quad (1)$$

Y_t – wartość wyprodukowanych (przewiezionych) komputerów w USA w milionach USD w latach 1992–1998,
(ang. *Electronic Computer Manufacturing: Value of Shipments: Millions of Dollars: Seasonally Adjusted*)¹,

X_{1t} – wartość nowych zamówień na produkcję komputerów w USA w milionach USD w latach 1992–1998,
(ang. *Electronic Computer Manufacturing: Value of Shipments: Millions of Dollars: Seasonally Adjusted*)²,

X_{2t} – wartość wyprodukowanych (przewiezionych) nośników danych, pamięci masowych w USA w milionach USD w latach 1992–1998,
(ang. *Computer Storage Device Manufacturing: Value of Shipments: Millions of Dollars: Seasonally Adjusted*)³,

X_{3t} – wartość wyprodukowanych (przewiezionych) półprzewodników w USA w milionach USD w latach 1992–1998,
(ang. *Semiconductor and Related Device Manufacturing: Value of Shipments: Millions of Dollars: Seasonally Adjusted*)⁴,

α_j – parametry strukturalne, $j = 0, 1, 2, 3$,

ϵ_t – składnik losowy.

¹<http://www.economagic.com/em-cgi/data.exe/cenm3/a34avs>

²<http://www.economagic.com/em-cgi/data.exe/cenm3/a34ano>

³<http://www.economagic.com/em-cgi/data.exe/cenm3/a34bvs>

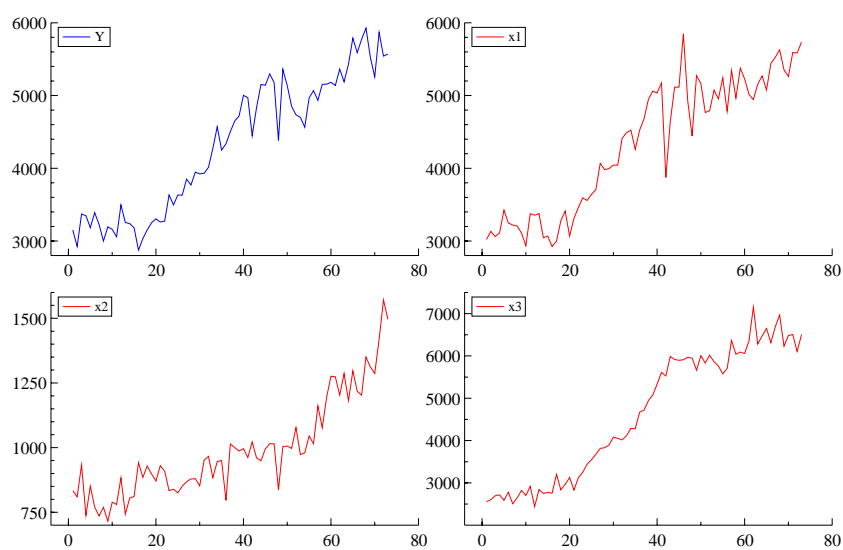
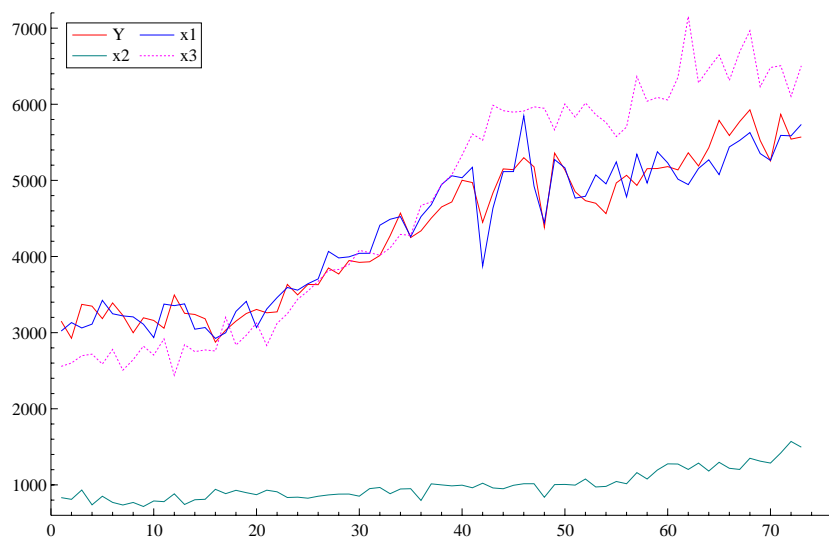
⁴<http://www.economagic.com/em-cgi/data.exe/cenm3/a34vgs>

1.1 Dane

Dane do modelu obejmują 73 obserwacje od lutego 1992 do lutego 1998:

t	Y_t	X_{1t}	X_{2t}	X_{3t}		t	Y_t	X_{1t}	X_{2t}	X_{3t}
1992 02	3151	3021	833	2557		1995 03	4651	4946	1000	4944
1992 03	2926	3132	810	2601		1995 04	4717	5061	987	5074
1992 04	3371	3062	933	2697		1995 05	5002	5035	996	5332
1992 05	3348	3112	738	2717		1995 06	4971	5172	962	5610
1992 06	3184	3424	851	2587		1995 07	4449	3874	1022	5527
1992 07	3390	3248	770	2779		1995 08	4836	4636	960	5987
1992 08	3227	3220	736	2507		1995 09	5151	5115	949	5916
1992 09	2999	3207	769	2645		1995 10	5141	5116	995	5896
1992 10	3195	3110	716	2823		1995 11	5298	5848	1015	5911
1992 11	3160	2935	789	2705		1995 12	5178	4924	1015	5967
1992 12	3058	3375	780	2918		1996 01	4383	4442	837	5947
1993 01	3495	3355	883	2440		1996 02	5359	5273	1004	5663
1993 02	3254	3378	743	2842		1996 03	5137	5164	1006	6004
1993 03	3239	3045	805	2751		1996 04	4851	4767	997	5831
1993 04	3182	3067	811	2773		1996 05	4734	4791	1077	6017
1993 05	2875	2924	942	2759		1996 06	4700	5072	973	5864
1993 06	3033	2999	885	3202		1996 07	4564	4955	980	5762
1993 07	3151	3279	929	2837		1996 08	4967	5243	1045	5576
1993 08	3251	3411	898	2965		1996 09	5067	4784	1015	5701
1993 09	3305	3067	871	3127		1996 10	4934	5343	1161	6365
1993 10	3261	3311	930	2831		1996 11	5154	4965	1077	6040
1993 11	3273	3461	909	3122		1996 12	5156	5376	1195	6090
1993 12	3633	3593	834	3248		1997 01	5181	5230	1276	6057
1994 01	3496	3558	839	3439		1997 02	5138	5016	1274	6354
1994 02	3633	3642	825	3547		1997 03	5363	4945	1203	7154
1994 03	3630	3706	851	3675		1997 04	5189	5157	1286	6281
1994 04	3850	4067	868	3816		1997 05	5429	5270	1182	6471
1994 05	3770	3981	879	3830		1997 06	5788	5076	1297	6648
1994 06	3947	3996	880	3894		1997 07	5590	5440	1218	6314
1994 07	3923	4042	852	4081		1997 08	5772	5525	1203	6688
1994 08	3931	4043	951	4051		1997 09	5927	5628	1350	6966
1994 09	4013	4411	966	4016		1997 10	5524	5352	1312	6231
1994 10	4273	4489	884	4112		1997 11	5256	5263	1287	6483
1994 11	4570	4524	946	4290		1997 12	5868	5591	1419	6506
1994 12	4249	4260	950	4278		1998 01	5542	5585	1571	6105
1995 01	4337	4524	796	4672		1998 02	5570	5735	1496	6505
1995 02	4504	4680	1014	4716						

1.2 Graficzna prezentacja danych



2 Dobór zmiennych objaśniających metodą Helwig'a

Correlation matrix

	y	x1	x2	x3
y	1.0000			
x1	0.96849	1.0000		
x2	0.82428	0.78519	1.0000	
x3	0.96741	0.94488	0.79939	1.0000

$$\mathbf{R}_0 = \begin{bmatrix} 0,968 \\ 0,824 \\ 0,967 \end{bmatrix}, \mathbf{R} = \begin{bmatrix} 1 & 0,785 & 0,945 \\ 0,785 & 1 & 0,799 \\ 0,945 & 0,799 & 1 \end{bmatrix}, \quad (2)$$

gdzie:

$\mathbf{R}_0 = [r_j]$ – macierz współczynników korelacji liniowej pomiędzy j . zmienną objaśniającą a zmienną objaśnianą, $j = 1, 2, \dots, k$,

$\mathbf{R} = [r_{ij}]$ – macierz współczynników korelacji liniowej pomiędzy i . a j . zmienną objaśniającą, $i = 1, 2, \dots, k$, $j = 1, 2, \dots, k$,

k – liczba zmiennych objaśniających w modelu.

Ilość możliwych podzbiorów ze zbioru zmiennych objaśniających $\{X_1, X_2, X_3\}$ (bez zbioru pustego):

$$S = 2^k - 1 = 2^3 - 1 = 7. \quad (3)$$

Te możliwe podzbiory to:

numer podzbioru (s)	podzbiór	zbiór indeksów zmiennych tworzących dany podzbiór (C_s)
1	$\{X_1\}$	$\{1\}$
2	$\{X_2\}$	$\{2\}$
3	$\{X_3\}$	$\{3\}$
4	$\{X_1, X_2\}$	$\{1, 2\}$
5	$\{X_1, X_3\}$	$\{1, 3\}$
6	$\{X_2, X_3\}$	$\{2, 3\}$
7	$\{X_1, X_2, X_3\}$	$\{1, 2, 3\}$

Pojemność indywidualną j . zmiennej s . podzbioru określamy jako:

$$h_{sj} = \frac{r_j^2}{\sum_{i \in C_s} |r_{ij}|}. \quad (4)$$

Pojemność integralna s . podzbioru to suma jego pojemności indywidualnych:

$$H_s = \sum_{j \in C_s} h_{sj}. \quad (5)$$

Pojemność integralna i indywidualna dla zbioru jednoelementowego jest identyczna:

$$(1) \quad \{X_1\}: h_{11} = \frac{r_1^2}{\sum_{i \in \{1\}} |r_{i1}|} = \frac{(0,968)^2}{|1|} \approx 0,937 = H_1,$$

$$(2) \quad \{X_2\}: h_{22} = \frac{r_2^2}{\sum_{i \in \{2\}} |r_{i2}|} = \frac{(0,824)^2}{|1|} \approx 0,679 = H_2,$$

$$(3) \quad \{X_3\}: h_{33} = \frac{r_3^2}{\sum_{i \in \{3\}} |r_{i3}|} = \frac{(0,967)^2}{|1|} \approx 0,935 = H_3,$$

oraz:

$$(4) \quad \{X_1, X_2\}: h_{41} = \frac{r_1^2}{\sum_{i \in \{1,2\}} |r_{i1}|} = \frac{(0,968)^2}{|1|+|0,785|} \approx 0,525,$$

$$h_{42} = \frac{r_2^2}{\sum_{i \in \{1,2\}} |r_{i2}|} = \frac{(0,824)^2}{|0,785|+|1|} \approx 0,380,$$

$$H_4 = h_{41} + h_{42} = 0,905,$$

$$(5) \quad \{X_1, X_3\}: h_{51} = \frac{r_1^2}{\sum_{i \in \{1,3\}} |r_{i1}|} = \frac{(0,968)^2}{|1|+|0,945|} \approx 0,482,$$

$$h_{53} = \frac{r_3^2}{\sum_{i \in \{1,3\}} |r_{i3}|} = \frac{(0,967)^2}{|0,945|+|1|} \approx 0,481,$$

$$H_5 = h_{51} + h_{53} = 0,963,$$

$$(6) \quad \{X_2, X_3\}: h_{62} = \frac{r_2^2}{\sum_{i \in \{2,3\}} |r_{i2}|} = \frac{(0,824)^2}{|1|+|0,799|} \approx 0,377,$$

$$h_{63} = \frac{r_3^2}{\sum_{i \in \{2,3\}} |r_{i3}|} = \frac{(0,967)^2}{|0,799|+|1|} \approx 0,520,$$

$$H_6 = h_{62} + h_{63} = 0,897,$$

$$(7) \quad \{X_1, X_2, X_3\}: h_{71} = \frac{r_1^2}{\sum_{i \in \{1,2,3\}} |r_{i1}|} = \frac{(0,968)^2}{|1|+|0,785|+|0,945|} \approx 0,343,$$

$$h_{72} = \frac{r_2^2}{\sum_{i \in \{1,2,3\}} |r_{i2}|} = \frac{(0,824)^2}{|0,785|+|1|+|0,799|} \approx 0,319,$$

$$h_{73} = \frac{r_3^2}{\sum_{i \in \{1,2,3\}} |r_{i3}|} = \frac{(0,967)^2}{|0,945|+|0,799|+|1|} \approx 0,341,$$

$$H_7 = h_{71} + h_{72} + h_{73} \approx 1.$$

Zestawienie pojemności integralnych:

s	podzbiór	H_s
1	$\{X_1\}$	0,937
2	$\{X_2\}$	0,679
3	$\{X_3\}$	0,935
4	$\{X_1, X_2\}$	0,905
5	$\{X_1, X_3\}$	0,963
6	$\{X_2, X_3\}$	0,8974
7	$\{X_1, X_2, X_3\}$	1,000

Wybieramy ten podzbiór, dla którego pojemność integralna jest największa:

$$C_{opt} = \max\{H_s : s = 1, 2, \dots, S = 2^k - 1\} = C_7. \quad (6)$$

Zatem do modelu wchodzi zmienne: X_1, X_2, X_3 .

Można uniknąć liczenia pojemności integralnych dla wszystkich podzbiorów: z możliwych podzbiorów należy wykluczyć te, które nie zawierają zmiennej najbardziej skorelowanej ze zmienną objaśnianą (w naszym przykładzie odrzucamy kombinacje bez zmiennej X_1 : $s = 2, 3, 6$).

3 Oszacowanie parametrów strukturalnych

Pakiet PC-Give podaje wyniki:

Variable	Coefficient	Std.Error	t-value	t-prob	PartR ²
Constant	493.30	152.46	3.236	0.0019	0.1317
x1	0.48675	0.068517	7.104	0.0000	0.4224
x2	0.49478	0.17920	2.761	0.0074	0.0995
x3	0.26531	0.042543	6.236	0.0000	0.3605

R² = 0.967122 F(3,69) = 676.56 [0.0000] \sigma = 171.707 DW = 1.96
 RSS = 2034352.204 for 4 variables and 73 observations

Oszacowany metodą najmniejszych kwadratów model przyjmuje zatem postać:

$$\hat{Y}_t = \hat{\alpha}_0 + \hat{\alpha}_1 X_{1t} + \hat{\alpha}_2 X_{2t} + \hat{\alpha}_3 X_{3t}, \quad (7)$$

$$\hat{Y}_t = 493,30 + 0,487X_{1t} + 0,495X_{2t} + 0,265X_{3t}, \quad (8)$$

$$(152,46) \quad (0,069) \quad (0,179) \quad (0,043). \quad (9)$$

W nawiasach podano średnie błędy szacunku.

3.1 Interpretacja oszacowanych parametrów strukturalnych

$\hat{\alpha}_1 = 0,487$ – wzrost wartości zamówień na produkcję komputerów o 1 mln USD (*ceteris paribus* - przy pozostałych warunkach niezmiennych) wywoła wzrost produkcji komputerów średnio o 487 tys. USD,

$\hat{\alpha}_2 = 0,495$ – wzrost wartości produkcji nośników danych o 1 mln USD (*ceteris paribus*) wywoła wzrost produkcji komputerów średnio o 495 tys. USD,

$\hat{\alpha}_3 = 0,265$ – wzrost wartości wyprodukowanych półprzewodników o 1 mln USD (*ceteris paribus*) wywoła wzrost produkcji komputerów średnio o 265 tys. USD.

3.2 Interpretacja błędów szacunku parametrów

$S_{\hat{\alpha}_0} = 152,46$ – szacując α_0 na poziomie 493,30 mylimy się średnio o $\pm 152,46$,

$S_{\hat{\alpha}_1} = 0,069$ – szacując α_1 na poziomie 0,487 mylimy się średnio o $\pm 0,069$,

$S_{\hat{\alpha}_2} = 0,197$ – szacując α_2 na poziomie 0,495 mylimy się średnio o $\pm 0,197$,

$S_{\hat{\alpha}_3} = 0,043$ – szacując α_3 na poziomie 0,265 mylimy się średnio o $\pm 0,043$.

4 Badanie własności koïncydencji

Model jest koïncydentny, gdy spełniony jest warunek:

$$\bigwedge \text{sgn } r_j = \text{sgn } \hat{\alpha}_j, \quad j = 1, 2, \dots, k.^5 \quad (10)$$

$$\begin{aligned} \text{sgn } (r_1 = 0,968) &= " + " \text{ oraz } \text{sgn } (\hat{\alpha}_1 = 0,487) = " + " \\ \text{sgn } (r_2 = 0,824) &= " + " \text{ oraz } \text{sgn } (\hat{\alpha}_2 = 0,495) = " + " \\ \text{sgn } (r_3 = 0,967) &= " + " \text{ oraz } \text{sgn } (\hat{\alpha}_3 = 0,265) = " + " \end{aligned}$$

Ponieważ dla wszystkich par znaki są zgodne, zatem model spełnia postulat koïncydencji.

5 Normalność rozkładu składnika loswego

5.1 Test Jarque-Bera

Po oszacowaniu modelu:

$$Y_t = \alpha_0 + \alpha_1 X_{1t} + \alpha_2 X_{2t} + \alpha_3 X_{3t} + \epsilon_t, \quad (11)$$

obliczamy statystykę JB :

$$JB = n \left[\frac{1}{6} \left(\frac{\frac{1}{n} \sum_{t=1}^n e_t^3}{\left(\sqrt{\frac{1}{n} \sum_{t=1}^n e_t^2} \right)^3} \right)^2 + \frac{1}{24} \left(\frac{\frac{1}{n} \sum_{t=1}^n e_t^4}{\left(\sqrt{\frac{1}{n} \sum_{t=1}^n e_t^2} \right)^4} - 3 \right)^2 \right], \quad (12)$$

która ma rozkład $\chi_{df=2}^2$.

Jeżeli $JB < \chi_{\alpha=0,05; df=2}^2$ to:

$H_0 : \epsilon_t \sim N(0, \sigma^2)$ – składnik losowy ma rozkład normalny,

jeżeli $JB > \chi_{\alpha=0,05; df=2}^2$ to:

$H_1 : \epsilon_t \not\sim N(0, \sigma^2)$ – składnik losowy nie ma rozkładu normalnego.

```
-----
Normality test for Residual
Sample size 73: 1992 (2) to 1998 (2)
-----
Normality Chi^2(2)=    0.025478 [0.9873]
-----
```

$\chi_{df=2}^2 = 0,0255$ oraz $\alpha^* = 0,9873$.

Zatem dopiero przy prawie 99% poziomie istotności moglibyśmy odrzucić H_0 , co znacznie przekracza 5% poziom błędu. Uznajemy więc, że w naszym modelu składnik losowy ma rozkład normalny.

⁵Por. równania (2) oraz (7), (8)

6 Autokorelacja składnika losowego

6.1 Test Durbin'a-Watson'a

Test DW możemy stosować, gdy:

- (1) w modelu występuje wyraz wolny ($\alpha_0 \neq 0$),
- (2) składnik losowy ma rozkład normalny ($\epsilon_t \sim N(0, \sigma^2)$),
- (3) model nie ma postaci autoregresyjnej – nie występuje opóźniona zmienna objaśniana jako zmienna objaśniająca.

Składnik losowy ϵ_t możemy przedstawić w postaci:

$$\epsilon_t = \rho\epsilon_{t-1} + \xi_t, \quad |\rho| < 1, \quad (13)$$

gdzie:

- ρ – współczynnik autokorelacji,
- ξ_t – składnik losowy.

Nieobciążonym estymatorem współczynnika ρ jest:

$$\hat{\rho} = \frac{\sum_{t=2}^n e_t e_{t-1}}{\sqrt{\sum_{t=2}^n e_t^2 \sum_{t=2}^n e_{t-1}^2}}. \quad (14)$$

Statystykę DW obliczamy jako:

$$DW = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}. \quad (15)$$

Pomiędzy DW a $\hat{\rho}$ zachodzi związek:

$$DW \approx 2(1 - \hat{\rho}). \quad (16)$$

	$DW < 2$	$DW > 2$
	$H_0 : \rho = 0$ brak autokorelacji $H_1 : \rho > 0$ dodatnia autokorelacja	$H_0 : \rho = 0$ brak autokorelacji $H_1 : \rho < 0$ ujemna autokorelacja
Proces weryfikacji:	H_1 ? H_0 $\frac{d_L}{d_U}$	H_0 ? H_1 $\frac{4 - d_U}{4 - d_L}$

 $R^2 = 0.967122$ $F(3,69) = 676.56$ $[0.0000]$ $\sigma = 171.707$ $DW = 1.96$
 $RSS = 2034352.204$ for 4 variables and 73 observations

Zatem: $H_1 : \rho > 0$,

odczytane wartości krytyczne z tablic dla $\alpha = 0,05$, $n = 73$, $k = 3$ to:

$d_L = 1,543$,

$d_U = 1,709$.

Ponieważ $1,96 > 1,709$ ($DW > d_U$), zatem na postawie testu Durbin'a-Watson'a nie możemy odrzucić H_0 o braku autokorelacji składnika losowego.

6.2 Test mnożnika Lagrange'a

Po oszacowaniu modelu:

$$Y_t = \alpha_0 + \alpha_1 X_{1t} + \alpha_2 X_{2t} + \alpha_3 X_{3t} + \epsilon_t, \quad (17)$$

szacujemy model:

$$e_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 e_{t-1} + \xi_t \quad (18)$$

i obliczamy R^2 .

Obliczamy statystykę $(n-1)R^2$ i odczytujemy z tablic $\chi^2_{\alpha=0,05; df=1}$.

Jeżeli $(n-1)R^2 < \chi^2_{\alpha=0,05; df=1}$ to:

$H_0 : \rho = 0$ – brak autokorelacji składnika losowego,

jeżeli $(n-1)R^2 \geq \chi^2_{\alpha=0,05; df=1}$ to:

$H_1 : \rho \neq 0$ – autokorelacja składnika losowego.

```
-----  
Testing for Error Autocorrelation from lags 1 to 1  
Chi^2(1) = 0.01015 [0.9198] and F-form(1,68) = 0.0094563 [0.9228]  
-----
```

$\chi^2_{df=1} = 0,0102$ oraz $\alpha^* = 0,9198$.

Zatem dopiero przy prawie 92% poziomie istotności moglibyśmy odrzucić H_0 , co znacznie przekracza 5% poziom błędu. Uznajemy więc, że w naszym modelu nie występuje zjawisko autokorelacji składnika losowego.

Możemy również odczytać wartości oszacowanych parametrów równania (13):

```
-----  
Autoregression for Residual: lags from 1 to 1  
The present sample is: 1992 (3) to 1998 (2)  
-----
```

	Constant	Lag 1
Coeff.	-1.373	0.012
Std.Err	20.04	0.1202

```
-----
```

$\hat{\epsilon}_t = -1,373 + 0,012\epsilon_{t-1}$, co potwierdza niską wartość współczynnika autokorelacji.

7 Heteroskedastyczność składnika losowego

7.1 Test White'a

Po oszacowaniu modelu:

$$Y_t = \alpha_0 + \alpha_1 X_{1t} + \alpha_2 X_{2t} + \alpha_3 X_{3t} + \epsilon_t, \quad (19)$$

szacujemy model:

$$e_t^2 = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{1t}^2 + \beta_5 X_{2t}^2 + \beta_6 X_{3t}^2 + \xi_t, \quad (20)$$

lub:

$$e_t^2 = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{1t}^2 + \beta_5 X_{2t}^2 + \beta_6 X_{3t}^2 + \beta_7 X_{1t} X_{2t} + \beta_8 X_{1t} X_{3t} + \beta_9 X_{2t} X_{3t} + \xi_t \quad (21)$$

i obliczamy R^2 .

Obliczamy statystkę nR^2 i odczytujemy z tablic:

$\chi_{\alpha=0,05; df=6}^2$ dla modelu (20),

$\chi_{\alpha=0,05; df=9}^2$ dla modelu (21).

Jeżeli $nR^2 < \chi_{\alpha; df}^2$ to:

$H_0 : \sigma_i^2 = \sigma^2$ – składnik losowy jest homoskedastyczny,

jeżeli $nR^2 > \chi_{\alpha; df}^2$ to:

$H_1 : \sigma_i^2 \neq \sigma^2$ – składnik losowy jest heteroskedastyczny.

```
-----
Testing for Heteroscedastic errors (squares)
Chi^2(6) = 7.121 [0.3098]    and  F-form(6,62) = 1.117 [0.3629]
-----
Testing for Heteroscedastic errors (squares and cross-products)
Chi^2(9) = 8.6952 [0.4659]    and  F-form(9,59) = 0.88643 [0.5427]
-----
```

Czyli:

$\chi_{df=6}^2 = 7,121$ oraz $\alpha^* = 0,3098$, dla modelu (20),

$\chi_{df=9}^2 = 8,6952$ oraz $\alpha^* = 0,4659$, dla modelu (21).

W obu przypadkach krytyczny poziom istotności znacznie przekracza 5%. Uznajemy więc, że w naszym modelu składnik losowy jest homoskedastyczny.

8 Liniowość modelu – test liczby serii

H_0 : oszacowany model jest liniowy,

H_1 : oszacowany model nie jest liniowy.

poziom istotności: $\alpha = 0,05$,

liczba serii: $r=37$

liczba reszt dodatnich: $n_1=35$,

liczba reszt ujemnych: $n_2=38$,

poziom krytyczny odczytany z tablic: $r^*=30$.

Przy 5% poziomie istotności $r > r^*$, więc nie ma podstaw do odrzucenia H_0 o liniowości modelu.

9 Współliniowość zmiennych objaśniających

Dla k modeli:

$$X_{1t} = \alpha_{1,0} + \alpha_{1,2}X_{2t} + \alpha_{1,3}X_{3t} + \dots + \alpha_{1,k}X_{kt} + \epsilon_{1t},$$

$$X_{2t} = \alpha_{2,0} + \alpha_{2,1}X_{1t} + \alpha_{2,3}X_{3t} + \dots + \alpha_{2,k}X_{kt} + \epsilon_{2t},$$

\vdots

$$X_{kt} = \alpha_{k,0} + \alpha_{k,1}X_{1t} + \alpha_{k,2}X_{2t} + \dots + \alpha_{k,k-1}X_{(k-1)t} + \epsilon_{kt},$$

obliczamy współczynnik determinacji R_j^2 oraz czynnik inflacji wariancji estymatora α_j :

$$CIW_j = \frac{1}{1 - R_j^2}. \quad (22)$$

Jeżeli:

$R_j^2 = 0$ oraz $CIW_j = 1$ – brak współliniowości zmiennych,

$R_j^2 > 0$ oraz $CIW_j > 1$ – przybliżona współliniowość zmiennych,

$CIW_j > 10$ – współliniowość zmiennych trwale zakłócająca jakość modelu.

Obliczenia:

$$\begin{aligned} \hat{X}_{1t} &= 1484,8 + 0,397X_{2t} + 0,529X_{3t}, \quad R_1^2 = 0,895 \implies CIW = 9,52, \\ \hat{X}_{2t} &= 426,75 + 0,058X_{1t} + 0,067X_{3t}, \quad R_2^2 = 0,647 \implies CIW = 2,832, \\ \hat{X}_{3t} &= -2494,7 + 1,373X_{1t} + 1,194X_{2t}, \quad R_3^2 = 0,901 \implies CIW = 10,1. \end{aligned}$$

Zatem w modelu zmienna X_3 powinna zostać usunięta. W celach czysto rachunkowych pozostawimy tę zmienną w modelu, przedstawiając jej wpływ na dalszą analizę.

W programie STATISTICA możemy odczytać R_j^2 wciskając po oszacowaniu przycisk *Redundancy*:

Redundancy of Independent Variables; DV: Y (compfull2sta.sta)				
Continue...	R-square column contains R-square of respective variable with all other independent variables			
variable	Toleran.	R-square	Partial Cor.	Semipart Cor.
X1	,104736	,895264	,649952	,155072
X2	,352660	,647340	,315414	,060268
X3	,098592	,901408	,600382	,136127

10 Istotność zmiennych objaśniających – test t-Studenta

$H_0 : \alpha_j = 0$ – j . zmienna jest nieistotna w modelu, $j = 0, 1, 2, \dots, k$,

$H_1 : \alpha_j \neq 0$ – j . zmienna jest istotna w modelu, $j = 0, 1, 2, \dots, k$.

Zmienna losowa:

$$t_{\hat{\alpha}_j} = \left| \frac{\hat{\alpha}_j}{S_{\hat{\alpha}_j}} \right|, \quad (23)$$

ma rozkład t-Studenta z $n - (k + 1)$ stopniami swobody.

Statystyki t-Studenta oraz krytyczne (nominalne) poziomy istotności dla poszczególnych zmiennych wynoszą odpowiednio:

Variable	t-value	t-prob
Constant	3.236	0.0019
x1	7.104	0.0000
x2	2.761	0.0074
x3	6.236	0.0000

Z powyższych krytycznych poziomów istotności wynika, że wszystkie zmienne (+wyraz wolny) są statystycznie istotne dla poziomów istotności (odpowiednio): α_0 od 0,19%, X_1 od 0,00%, X_2 od 0,74%, X_3 od 0,00%.

2. sposób:

Statystyka t-Studenta odczytana z tablic dla 5% poziomu istotności i $df=73-(3+1)=69$ stopni swobody wynosi:

$$t_{\alpha=0,05;df=69} = 2 \quad (24)$$

i jest mniejsza od wartości statystyk t-Studenta dla poszczególnych zmiennych, co prowadzi do wniosku o statystycznej istotności zmiennych.

11 Współczynnik determinacji

Współczynnik determinacji możemy obliczyć korzystając ze wzorów:

$$\mathbf{R}^2 = \frac{\sum_{t=1}^n (\hat{y}_t - \bar{y})^2}{\sum_{t=1}^n (y_t - \bar{y})^2} = 1 - \frac{\sum_{t=1}^n (y_t - \hat{y})^2}{\sum_{t=1}^n (y_t - \bar{y})^2} = \quad (25)$$

$$= 1 - \frac{\mathbf{e}^T \mathbf{e}}{\mathbf{y}^T \mathbf{y} - n\bar{y}^2} = 1 - \frac{\mathbf{y}^T \mathbf{y} - \hat{\alpha}^T \mathbf{X}^T \mathbf{y}}{\mathbf{y}^T \mathbf{y} - n\bar{y}^2} = \quad (26)$$

$$= \mathbf{R}_0^T \mathbf{R}^{-1} \mathbf{R}_0 \quad (27)$$

$\mathbf{R}^2 \approx 0,967$: około 97% zmienności produkcji komputerów jest wyjaśniane przez zmienność zmiennych:

- (a) zamówienia na produkcję komputerów,
- (b) produkcję nośników danych,
- (c) produkcję półprzewodników.

11.1 Skorygowany współczynnik determinacji

$$\bar{R}^2 = R^2 - \frac{k}{n - (k + 1)}(1 - R^2), \quad (28)$$

$$\bar{R}^2 \approx 0,966.$$

12 Prognozy

Wprowadźmy oznaczenia dla okresu prognozowanego τ :

$$\tau = n + 1, n + 2, \dots, n + s = T.$$

Zatem długość okresu prognozy wynosi: $T - n = s$.

12.1 Prognoza zmiennej endogenicznej na podstawie modelu trendu linowego

Szacujemy model postaci:

$$Y_t = \alpha_0 + \alpha_1 t + \epsilon_t, \quad (29)$$

otrzymując:

$$\hat{Y}_t = 2787,3 + 41,522t, \quad (30)$$

na podstawie wyników:

Variable	Coefficient	Std.Error	t-value	t-prob	PartR ²
Constant	2787.3	68.721	40.559	0.0000	0.9586
Trend	41.522	1.6140	25.727	0.0000	0.9031

$R^2 = 0.903123$ $F(1,71) = 661.89$ $[0.0000]$ $\sigma = 290.565$ $DW = 0.845$
 $RSS = 5994401.909$ for 2 variables and 73 observations

Prognoza Y na okres $\tau=74$:

$$Y_{74}^P = 2787,3 + 41,522 * 74 = 5859,93. \quad (31)$$

Ponieważ w modelu występuje zjawisko autokorelacji składnika losowego (niskie DW), zatem aby dokonać prognozy punktowej zmiennej Y_t należy skorygować powyższy wynik o $\hat{\rho}e_{73}$, gdzie:

$$e_{73} - 73. \text{ reszta modelu trendu obliczona jako: } Y_{73} - \hat{Y}_{73},$$

$$\hat{\rho} - \text{współczynnik autokorelacji składnika losowego obliczony ze wzoru (16),}$$

$$e_{73} = 5570 - (2787,3 + 41,522 * 73) = -248,406,$$

$$\hat{\rho} = 1 - \frac{0,845}{2} = 0,57.$$

Zatem:

$$\hat{\rho}e_{73} = -143,45,$$

$$Y_{74}^P = 5859,93 - 143,45 = 5716,5.$$

Ze względu na dość kłopotliwy sposób obliczania wariancji predykcji (średniego błędu prognozy) w warunkach autokorelacji składnika losowego⁶, można poszukiwać innej postaci trendu, która będzie pozbawiona tej niedogodności.

12.2 Prognoza zmiennej endogenicznej na podstawie modelu trendu wielomianowego

Postać trendu wielomianowego dla zmiennej Y_t :

$$Y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \dots + \alpha_q t^q + \epsilon_t, \quad (32)$$

Rząd q ustalamy na podstawie otrzymanego optymalnego modelu:

Variable	Coefficient	Std.Error	t-value	t-prob	PartR ²
Constant	2785.8	196.43	14.182	0.0000	0.7529
t	238.52	70.960	3.361	0.0013	0.1462
t2	-35.794	8.1443	-4.395	0.0000	0.2264
t3	2.0889	0.40802	5.120	0.0000	0.2842
t4	-0.053483	0.0099665	-5.366	0.0000	0.3038
t5	0.00062735	0.00011668	5.377	0.0000	0.3046
t6	-2.7638e-006	5.2381e-007	-5.276	0.0000	0.2967

R² = 0.956664 F(6,66) = 242.83 [0.0000] \sigma = 201.564 DW = 1.86
 RSS = 2681453.905 for 7 variables and 73 observations

Zatem oszacowany model przyjmuje postać:

$$\hat{Y}_t = 2785,8 + 238,52t - 35,79t^2 + 2,09t^3 - 0,0534t^4 + 0,0006t^5 - 0,0000028t^6. \quad (33)$$

Podstawiając za $t = 74$ otrzymujemy⁷:

$$Y_{74}^P = 5392,764608.$$

Ponieważ DW wskazuje na brak autokorelacji składnika losowego (potwierdza to również poniższy test mnożnika Lagrange'a), możemy przejść do analizy błędów *ex ante* prognozy.

Testing for Error Autocorrelation from lags 1 to 1
 Chi²(1) = 0.28669 [0.5923] and F-form(1,65) = 0.25628 [0.6144]

12.2.1 Ocena *ex ante* prognozy

Średni błąd prognozy *ex ante* obliczamy jako:

$$S_\tau^P = \sqrt{S^2 + \mathbf{x}_\tau^T \hat{D}^2(\hat{\alpha}) \mathbf{x}_\tau} = S \sqrt{1 + \mathbf{x}_\tau^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_\tau}, \quad (34)$$

gdzie⁸:

$$S^2 = \frac{\mathbf{e}^T \mathbf{e}}{n - (k + 1)}, \quad (35)$$

⁶Zeliaś A., *Teoria Prognozy*, PWE, Warszawa 1997, s. 217

⁷Precyzyjnych obliczeń dokonano w Excel.

⁸W pakietach komputerowych S podawane jest najczęściej jako σ .

\mathbf{x}_τ – wektor prognozowanych wartości zmiennych objaśniających.

$$S_{74}^P = 281,4462.^9$$

Średni względny błąd prognozy *ex ante*:

$$V_\tau = \left| \frac{S_\tau^P}{Y_\tau^P} \right|, \quad (36)$$

$$V_{74} = 0,0522.$$

Zatem średni względny błąd *ex ante* prognozy na 74 okres wynosi 5,2%.

12.2.2 Prognoza przedziałowa

Jeśli składnik losowy ma rozkład normalny to zmienna losowa:

$$u = \frac{|y_\tau - y_\tau^P|}{S_\tau^P}, \quad (37)$$

ma rozkład t-Studenta z $n - (k + 1)$ stopniami swobody.

Przedział ufności prognozowanej zmiennej wyznaczamy jako:

$$P(y_\tau^P - t_{\alpha, df} S_\tau^P < y_\tau < y_\tau^P + t_{\alpha, df} S_\tau^P) = 1 - \alpha. \quad (38)$$

Zatem dla $\alpha = 0,05$ oraz $df = 73 - (6 + 1) = 66$:

$$\begin{aligned} P(5392,76 - 2 * 281,45 < y_{74} < 5392,76 + 2 * 281,45) &= 0,95, \\ P(4829,87 < y_{74} < 5955,66) &= 0,95. \end{aligned}$$

Przedział o końcówkach (4829,87; 5955,66) pokrywa prognozowaną wartość Y_{74} (na 1 okres do przodu) z prawdopodobieństwem 95%.

⁹Obliczeń dokonano w Excel.

12.3 Prognoza zmiennej endogenicznej na podstawie modelu ekonometrycznego (warunkowa)

Oszacowany model z równania (1) na postać:¹⁰

$$\hat{Y}_t = 493,30 + 0,487X_{1t} + 0,495X_{2t} + 0,265X_{3t}. \quad (39)$$

12.3.1 Prognoza zmiennych egzogenicznych na podstawie modeli trendu liniowego

Dokonyjemy prognoz zmiennych egzogenicznych na $\tau = 74$ okres na podstawie modeli trendu liniowego:

$$X_{jt} = \alpha_0 + \alpha_1 t + \epsilon_t, \quad (40)$$

otrzymując:

```
-----
dla X1:
Variable      Coefficient      Std.Error    t-value    t-prob    PartR^2
Constant      2849.9         78.514      36.298    0.0000    0.9489
t              40.108         1.8440     21.751    0.0000    0.8695
R^2 = 0.86951  F(1,71) = 473.1 [0.0000]  \sigma = 331.974  DW = 0.999
RSS = 7824670.687 for 2 variables and 73 observations
-----

dla X2:
Variable      Coefficient      Std.Error    t-value    t-prob    PartR^2
Constant      700.02         21.838      32.055    0.0000    0.9354
t              7.8513         0.51288     15.308    0.0000    0.7675
R^2 = 0.767478  F(1,71) = 234.35 [0.0000]  \sigma = 92.3348  DW = 0.816
RSS = 605325.7327 for 2 variables and 73 observations
-----

dla X3:
Variable      Coefficient      Std.Error    t-value    t-prob    PartR^2
Constant      2089.1         94.342      22.144    0.0000    0.8735
t              68.914         2.2157      31.103    0.0000    0.9316
R^2 = 0.931624  F(1,71) = 967.38 [0.0000]  \sigma = 398.895  DW = 0.485
RSS = 11297300.23 for 2 variables and 73 observations
-----
```

Dla każdego z trzech powyższych modeli DW jest zbyt niskie by pomyślnie przejść test autokorelacji składnika losowego. Potwierdza to również test mnożnika Lagrange’a:

```
-----
Testing for Error Autocorrelation from lags 1 to 1
-----

dla X1:
Chi^2(1) = 18.211 [0.0000] ** and F-form(1,70) = 23.266 [0.0000] **
-----

dla X2:
Chi^2(1) = 23.135 [0.0000] ** and F-form(1,70) = 32.477 [0.0000] **
-----

dla X3:
Chi^2(1) = 40.745 [0.0000] ** and F-form(1,70) = 88.424 [0.0000] **
-----
```

Podsumowując:

Modele trendu liniowego dla zmiennych egzogenicznych mają postać:

$$\hat{X}_{1t} = 2849,9 + 40,12t,$$

¹⁰Por. równanie (8)

$$\hat{X}_{2t} = 700,02 + 7,85t,$$

$$\hat{X}_{3t} = 2089,1 + 68,91t.$$

Każdy z nich obarczony jest autokorelacją składnika losowego. Prognozy punktowe na ich podstawie muszą być skorygowane o $\hat{\rho}e_{73}$ (jak w przypadku modelu trendu liniowego dla Y_t):

dla X_{1t} :

$$e_{73} = 5735 - (2849,9 + 40,12 * 73) = -42,77,$$

$$\hat{\rho} = 1 - \frac{0,999}{2} = 0,5,$$

$$\hat{\rho}e_{73} = -21,39,$$

zatem:

$$Y_{74}^P = 2849,9 + 40,12 * 74 - 21,39 = 5796,49,$$

dla X_{2t} :

$$e_{73} = 1496 - (700,02 + 7,85 * 73) = 222,83,$$

$$\hat{\rho} = 1 - \frac{0,816}{2} = 0,59,$$

$$\hat{\rho}e_{73} = 131,47,$$

zatem:

$$Y_{74}^P = 700,02 + 7,85 * 74 + 131,47 = 1412,49,$$

dla X_{3t} :

$$e_{73} = 6505 - (2089,1 + 68,91 * 73) = -614,75,$$

$$\hat{\rho} = 1 - \frac{0,485}{2} = 0,76,$$

$$\hat{\rho}e_{73} = -467,21,$$

zatem:

$$Y_{74}^P = 2089,1 + 68,91 * 74 - 467,21 = 6721,45.$$

Wektor prognozowanych zmiennych egzogenicznych wynosi:

$$\mathbf{x}_{74} = \begin{bmatrix} 1 \\ 5796,49 \\ 1412,49 \\ 6721,45 \end{bmatrix}. \quad (41)$$

Podstawiając go do równania (39) otrzymujemy wartość punktowej prognozy warunkowej Y_{74} :

$$Y_{74}^P = 493,30 + 0,487 * 5796,49 + 0,495 * 1412,49 + 0,265 * 6721,45 = 5796,84.$$

Ocena *ex ante* prognozy

Korzystając z równania (34) obliczmy średni błąd prognozy *ex ante*:

$$S = 171,707,^{11}$$

$$S_{19}^P = 179,61,^{12}$$

Średni względny błąd prognozy *ex ante*:

¹¹PC-Give: po oszacowaniu regresji odczytujemy *sigma*.

¹²Na podstawie obliczeń w Excel.

$$V_{74} = \frac{179,61}{5796,84} = 0,031.$$

Zatem średni względny błąd *ex ante* prognozy na 74 okres wynosi 3,1%.

Prognoza przedziałowa

Korzystając ze wzoru (38) obliczamy końcówki prognozy przedziałowej:

$$P(5796,84 - 2 * 179,61 < y_{74} < 5796,84 + 2 * 179,61) = 0,95,$$

$$P(5437,62 < y_{74} < 6156,05) = 0,95.$$

Przedział o końcówkach (5437,62; 6156,05) pokrywa prognozowaną wartość Y_{74} (na 1 okres do przodu) z prawdopodobieństwem 95%¹³.

Prognozę przedziałową można obliczyć przy pomocy programu STATISTICA, korzystając z opcji *Compute prediction limits*, podając wektor prognozowanych zmiennych egzogenicznych. Należy pamiętać, że STATISTICA nie bierze pod uwagę zjawiska autokorelacji składnika losowego - nie obliczy więc poprawnie przedziałów prognozy gdy autokorelacja występuje. Nieznaczne różnice w

Predicting Values for (compfull2sta.sta)			
variable: Y			
variable	B-Weight	Value	B-Weight * Value
X1	,486746	5796,490	2821,421
X2	,494776	1412,490	698,866
X3	,265307	6721,450	1783,245
Intercpt			493,304
Predictd			5796,836
-95,0%PL			5438,530
+95,0%PL			6155,142

końcówkach wynikają z dokładniejszej wartości t – *Studenta* w STATISTICE ($t_{\alpha=0,05; df=69} = 1,99494479602436$).

¹³Uwaga: Powinniśmy obliczyć końcówki prognozy *ex ante* dla każdej z prognoz punktowych dla zmiennych egzogenicznych, następnie lewe i prawe (oddzielnie) końcówki potraktować jako prognozowany wektor zmiennych egzogenicznych, obliczyć prognozy warunkowe, dla każdej z nich obliczyć końcówki prognozy i dopiero ich skrajne wartości potraktować jako prognozę przedziałową.

12.3.2 Prognoza zmiennych egzogenicznych na podstawie modeli trendu wielomianowego

Postać trendu wielomianowego dla zmiennych X_{jt} :

$$X_{jt} = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \dots + \alpha_q t^q + \epsilon_t. \quad (42)$$

Rząd q ustalamy na podstawie otrzymanego optymalnego modelu:

dla X1

Variable	Coefficient	Std.Error	t-value	t-prob	PartR ²
Constant	2851.9	239.80	11.892	0.0000	0.6818
t1	171.71	86.630	1.982	0.0516	0.0562
t2	-25.836	9.9428	-2.598	0.0115	0.0928
t3	1.5302	0.49812	3.072	0.0031	0.1251
t4	-0.038621	0.012167	-3.174	0.0023	0.1324
t5	0.00043984	0.00014244	3.088	0.0029	0.1262
t6	-1.8661e-006	6.3948e-007	-2.918	0.0048	0.1143

R² = 0.933351 F(6,66) = 154.04 [0.0000] \sigma = 246.075 DW = 1.94
 RSS = 3996506.212 for 7 variables and 73 observations

dla X2

Variable	Coefficient	Std.Error	t-value	t-prob	PartR ²
Constant	995.45	87.673	11.354	0.0000	0.6682
t1	-119.85	51.230	-2.339	0.0224	0.0788
t2	22.425	9.8433	2.278	0.0261	0.0750
t3	-1.9307	0.87550	-2.205	0.0310	0.0706
t4	0.091006	0.041736	2.180	0.0329	0.0692
t5	-0.0024702	0.0011327	-2.181	0.0329	0.0692
t6	3.8397e-005	1.7519e-005	2.192	0.0320	0.0698
t7	-3.1684e-007	1.4367e-007	-2.205	0.0310	0.0706
t8	1.0747e-009	4.8445e-010	2.218	0.0301	0.0714

R² = 0.906386 F(8,64) = 77.458 [0.0000] \sigma = 61.7081 DW = 2.04
 RSS = 243704.7817 for 9 variables and 73 observations

dla X3

Variable	Coefficient	Std.Error	t-value	t-prob	PartR ²
Constant	2024.0	354.75	5.705	0.0000	0.3407
t1	560.62	251.81	2.226	0.0296	0.0729
t2	-153.37	59.530	-2.576	0.0123	0.0953
t3	19.066	6.6355	2.873	0.0055	0.1159
t4	-1.2621	0.40645	-3.105	0.0028	0.1327
t5	0.048533	0.014714	3.299	0.0016	0.1473
t6	-0.0011118	0.00032268	-3.446	0.0010	0.1586
t7	1.4914e-005	4.2076e-006	3.545	0.0007	0.1663
t8	-1.0797e-007	2.9992e-008	-3.600	0.0006	0.1706
t9	3.2545e-010	8.9934e-011	3.619	0.0006	0.1721

R² = 0.983537 F(9,63) = 418.19 [0.0000] \sigma = 207.79 DW = 1.93
 RSS = 2720131.434 for 10 variables and 73 observations

Otrzymujemy zatem trzy modele, w których składnik losowy pozbawiony jest autokorelacji:

Testing for Error Autocorrelation from lags 1 to 1

dla X1: Chi²(1) = 0.062101 [0.8032] and F-form(1,65) = 0.055342 [0.8148]
dla X2: Chi²(1) = 0.089631 [0.7646] and F-form(1,63) = 0.077448 [0.7817]
dla X3: Chi²(1) = 0.072393 [0.7879] and F-form(1,62) = 0.061545 [0.8049]

Ostatecznie otrzymujemy:

	2851,8548981299	2851,8549
74	171,7147703116	12706,8930
5476	-25,8363537988	-141479,8734
405224	1,5301687837	620061,1152
29986576	-0,0386205902	-1158099,2630
2219006624	0,0004398382	976003,7888
164206490176	-0,0000018661	-306427,1760

prognoza punktowa X1		5617,3395
=====		
	995,4463420096	995,4463
74	-119,8488444520	-8868,8145
5476	22,4250085310	122799,3467
405224	-1,9307208689	-782374,4334
29986576	0,0910066508	2728977,8518
2219006624	-0,0024701630	-5481307,9792
164206490176	0,0000383974	6305097,4293
12151280273024	-0,0000003168	-3850022,1410
899194740203776	0,0000000011	966371,5556

prognoza punktowa X2		1668,2617
=====		
	2023,9155485267	2023,9155
74	560,6594004983	41488,7956
5476	-153,3749261354	-839881,0955
405224	19,0667711390	7726313,2680
29986576	-1,2621443714	-37847388,1166
2219006624	0,0485354937	107700581,9258
164206490176	-0,0011118794	-182577817,9782
12151280273024	0,0000149146	181231660,8512
899194740203776	-0,0000001080	-97086796,8241
66540410775079400	0,0000000003	21656386,0580

prognoza punktowa X3		6570,7998
=====		

Wektor prognozowanych zmiennych egzogenicznych wynosi:

$$\mathbf{x}_{74} = \begin{bmatrix} 1 \\ 5617,3395 \\ 1668,2617 \\ 6570,7998 \end{bmatrix}. \quad (43)$$

Podstawiając go do równania (39) otrzymujemy wartość punktowej prognozy warunkowej Y_{74} :

$$Y_{74}^P = 493,30 + 0,487 * 5617,3395 + 0,495 * 1668,2617 + 0,265 * 6570,7998 = 5796,216.$$

12.3.3 Ocena *ex post* prognozy

Szacujemy model postaci:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + \eta_t, \quad (44)$$

skracać liczbę obserwacji o 3. Te 3 obserwacje potraktujemy jako realizacje prognozy (*ex post*).

Prognozy i realizacje podaje tabela:

Analysis of 1-step forecasts					
Date	Actual	Forecast	Y-Yhat	Forecast SE	t-value
1997 12	5868.00	5661.31	206.685	186.790	1.10651
1998 1	5542.00	5639.77	-97.7730	206.570	-0.473316
1998 2	5570.00	5774.83	-204.827	194.607	-1.05252
Tests of parameter constancy over: 1997 (12) to 1998 (2)					
Forecast	Chi ² (3)=	3.1988	[0.3620]		
Chow	F(3, 66) =	1.0194	[0.3898]		

(1) Średni błąd

(ang. *average error*, *AE* lub *mean error*, *ME*):

$$ME = \frac{1}{s} \sum_{\tau=n+1}^T (y_{\tau} - y_{\tau}^P) = -31,97. \quad (45)$$

(2) Średni błąd procentowy

(ang. *average percentage error*, *APE*
lub *mean percentage error*, *MPE*):

$$MPE = \frac{1}{s} \sum_{\tau=n+1}^T \frac{y_{\tau} - y_{\tau}^P}{y_{\tau}} 100\% = \frac{\sum_{\tau=n+1}^T (y_{\tau} - y_{\tau}^P)}{\sum_{\tau=n+1}^T y_{\tau}} 100\% = -0,64\%. \quad (46)$$

(3) Średni absolutny błąd

(ang. *average absolute error*, *AAE*
lub *mean absolute error*, *MAE*):

$$MAE = \frac{1}{s} \sum_{\tau=n+1}^T |y_{\tau} - y_{\tau}^P| = 169,76. \quad (47)$$

(4) Średni absolutny błąd procentowy

(ang. *average absolute percentage error*, *AAPE*
lub *mean absolute percentage error*, *MAPE*):

$$MAPE = \frac{1}{s} \sum_{\tau=n+1}^T \left| \frac{y_{\tau} - y_{\tau}^P}{y_{\tau}} \right| 100\% = \frac{\sum_{\tau=n+1}^T |y_{\tau} - y_{\tau}^P|}{\sum_{\tau=n+1}^T y_{\tau}} 100\% = 2,999\%. \quad (48)$$

(5) Błąd średniokwadratowy

(ang. *mean square error*, *MSE*):

$$MSE = \frac{1}{s} \sum_{\tau=n+1}^T (y_{\tau} - y_{\tau}^P)^2 = 31410,78. \quad (49)$$

- (6) Procentowy błąd średniokwadratowy
(ang. *mean square percentage error*, $MSPE$):

$$MSPE = \frac{1}{s} \sum_{\tau=n+1}^T \left(\frac{y_{\tau} - y_{\tau}^P}{y_{\tau}} \right)^2 100\% = \frac{\sum_{\tau=n+1}^T (y_{\tau} - y_{\tau}^P)^2}{\sum_{\tau=n+1}^T y_{\tau}^2} 100\% = 0,098\%. \quad (50)$$

- (7) Pierwiastek błędu średniokwadratowego
(ang. *root mean square error*, $RMSE$):

$$RMSE = \sqrt{\frac{1}{s} \sum_{\tau=n+1}^T (y_{\tau} - y_{\tau}^P)^2} = 177,23. \quad (51)$$

- (8) Pierwiastek procentowego błędu średniokwadratowego
(ang. *root mean square percentage error*, $RMSPE$):

$$RMSPE = \sqrt{\frac{1}{s} \sum_{\tau=n+1}^T \left(\frac{y_{\tau} - y_{\tau}^P}{y_{\tau}} \right)^2} 100\% = \sqrt{\frac{\sum_{\tau=n+1}^T (y_{\tau} - y_{\tau}^P)^2}{\sum_{\tau=n+1}^T y_{\tau}^2}} 100\% = 3,13\%. \quad (52)$$

- (9) Współczynnik rozbieżności 1
(ang. *inequality coefficient 1*, $IC1$):

$$IC1 = \frac{\sqrt{\frac{1}{s} \sum_{\tau=n+1}^T (y_{\tau} - y_{\tau}^P)^2}}{\sqrt{\frac{1}{s} \sum_{\tau=n+1}^T y_{\tau}^2} + \sqrt{\frac{1}{s} \sum_{\tau=n+1}^T y_{\tau}^2}} = 0,0156. \quad (53)$$

- (10) Współczynnik rozbieżności 2
(ang. *inequality coefficient 2*, $IC2$):

$$IC2 = \frac{\sqrt{\frac{1}{s} \sum_{\tau=n+1}^T (y_{\tau} - y_{\tau}^P)^2}}{\sqrt{\frac{1}{s} \sum_{\tau=n+1}^T y_{\tau}^2}} = 0,031. \quad (54)$$

12.4 Prognoza na podstawie modelu prostego wyrównania wykładniczego Brown'a

Dokonyjemy wygładzenia szeregu czasowego zmiennej Y_t na podstawie wzoru:

$$\hat{y}_{t+1} = \alpha y_t + (1 - \alpha) \hat{y}_t, \quad \alpha \in [0, 1], \quad (55)$$

gdzie:

\hat{y}_{t+1} – wartość wygładzona w czasie $t + 1$,

\hat{y}_t – wartość wygładzona w czasie t ,

y_t – wartość empiryczna Y w czasie t ,

α – stała wygładzania,

$\hat{y}_1 = y_1$.

Wartość prognozy obliczamy jako:

$$\hat{y}_{t+1} = \hat{y}_t + \alpha e_t, \quad (56)$$

gdzie:

$e_t = y_t - \hat{y}_t$ – błąd *ex post* prognozy.

Korzystając z programu STATISTICA wybieramy moduł *Time series, Forecasting*, wybieramy zmienną Y , następnie *exponential smoothing and forecasting*, następnie *Automatic search for best parameters (2)*, otrzymując wartość prognozy równą:

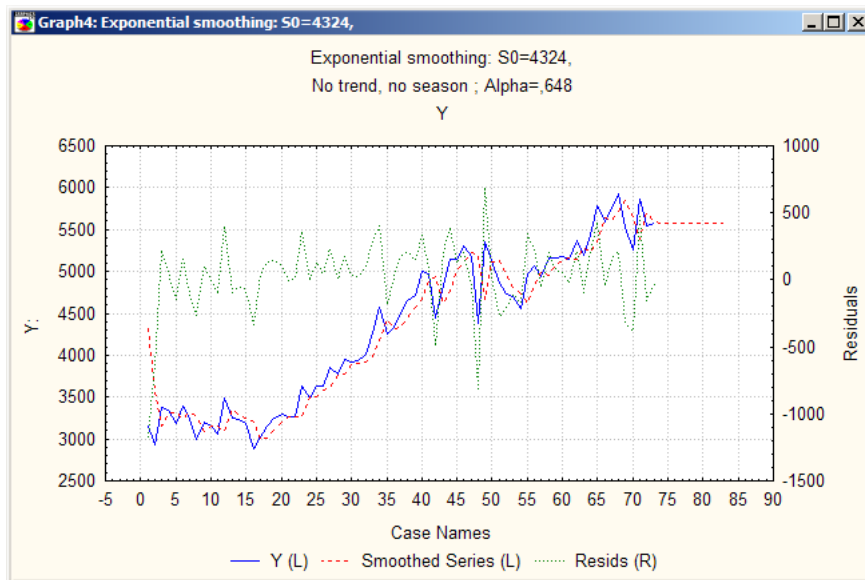
$\hat{y}_{74} = 5579,731$,

przy $\alpha = 0,648$ ¹⁴.

Dobór α następuje na podstawie minimalnych błędów ocen prognozy *ex post*:

Exponential smoothing: S0=4324, (compfull2sta.sta)	
TIME	No trend, no season ; Alpha=,648
SERIES	Y
Summary of error	
Error	
Mean error	26,55409140
Mean absolute error	202,93828551
Sums of squares	5950139,44688939
Mean square	81508,75954643
Mean percentage error	,20478428
Mean abs. perc. error	4,86947220

Graficzną prezentację wyrównania wykładniczego przedstawia rysunek:



¹⁴... $\alpha > 0,5$ może świadczyć o obecności w danych trendu czasowego, efektów sezonowych czy cyklicznych. Warto wtedy rozważyć zastosowanie innej metody. Por.: Gajda J. B., *Prognozowanie i symulacja a decyzje gospodarcze*, C.H. BECK, Warszawa 2001, s. 187