

Ekonometria

Dobór postaci analitycznej, transformacja liniowa i estymacja modelu KMNK

Paweł Cibis
pawel@cibis.pl

9 marca 2007

- 1 Miary dopasowania modelu do danych empirycznych
 - Współczynnik determinacji
 - Współczynnik zbieżności
 - Skorygowany R^2
- 2 Dobór postaci analitycznej
 - Metoda aprioryczna
 - Metoda heurystyczna
 - Metoda oceny wzrokowej rozrzutu punktów
- 3 Transformacja liniowa
 - Etapy transformacji liniowej
 - Przydatne funkcje
 - Przykłady transformacji liniowej
- 4 Estymacja modelu KMNK
 - Estymacja „macierzowa”
 - Estymacja – funkcja REGLINP
 - Estymacja – Analiza danych
 - Estymacja – Interpretacja parametrów struktury stochastycznej
- 5 Literatura

- 1 Miary dopasowania modelu do danych empirycznych
 - Współczynnik determinacji
 - Współczynnik zbieżności
 - Skorygowany R^2
- 2 Dobór postaci analitycznej
 - Metoda aprioryczna
 - Metoda heurystyczna
 - Metoda oceny wzrokowej rozrzutu punktów
- 3 Transformacja liniowa
 - Etapy transformacji liniowej
 - Przydatne funkcje
 - Przykłady transformacji liniowej
- 4 Estymacja modelu KMNK
 - Estymacja „macierzowa”
 - Estymacja – funkcja REGLINP
 - Estymacja – Analiza danych
 - Estymacja – Interpretacja parametrów struktury stochastycznej
- 5 Literatura

Współczynnik determinacji

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

W Excelu:

- Wykres/Dodaj linię trendu.../Opcje/Wyświetl wartości R-kwadrat na wykresie (tylko dla regresji jednej zmiennej)
- Analiza danych/Regresja
- REGLINP(Y;X;stała;1) – trzeci wiersz, pierwsza kolumna
- R.KWADRAT(Y;X) – tylko dla regresji jednej zmiennej
- Kwadrat współczynnika korelacji wielorakiej

Współczynnik zbieżności

$$\phi^2 = \frac{\sum_{t=1}^n e_t^2}{\sum_{t=1}^n (y_t - \bar{y})^2}$$

$$e_t = y_t - \hat{y}_t, \quad t = 1, 2, \dots, n$$

$$R^2 = 1 - \phi^2$$

Skorygowany (zmodyfikowany) R^2

$$\bar{R}^2 = 1 - \phi^2 \left(\frac{n-1}{n-k-1} \right)$$

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

n - liczba obserwacji

k - liczba zmiennych objaśniających w modelu

- 1 Miary dopasowania modelu do danych empirycznych
 - Współczynnik determinacji
 - Współczynnik zbieżności
 - Skorygowany R^2
- 2 Dobór postaci analitycznej
 - Metoda aprioryczna
 - Metoda heurystyczna
 - Metoda oceny wzrokowej rozrzutu punktów
- 3 Transformacja liniowa
 - Etapy transformacji liniowej
 - Przydatne funkcje
 - Przykłady transformacji liniowej
- 4 Estymacja modelu KMNK
 - Estymacja „macierzowa”
 - Estymacja – funkcja REGLINP
 - Estymacja – Analiza danych
 - Estymacja – Interpretacja parametrów struktury stochastycznej
- 5 Literatura

Metoda aprioryczna

Polega na doborze postaci analitycznej modelu na podstawie informacji pozastatystycznych, dotyczących związku łączącego zmienną objaśnianą ze zmiennymi objaśniającymi. Źródłami tych informacji mogą być:

- teoria ekonomii i ekonomik branżowych,
- opinie ekspertów,
- tradycje i doświadczenia badawcze.

Metoda aprioryczna

Wnioski wynikające z apriorycznej wiedzy o zależnościach między zmiennymi:

- stałe absolutne przyrosty zmiennej objaśnianej – funkcja liniowa,
- stałe względne przyrosty zmiennej objaśnianej – funkcja wykładnicza,
- stałe współczynniki elastyczności – funkcja potęgowa,
- coraz większe przyrosty zmiennej objaśnianej przy wzroście zmiennej objaśniającej o każdą kolejną jednostkę – funkcja potęgowa, wykładnicza lub kwadratowa,
- coraz mniejsze przyrosty zmiennej objaśnianej przy wzroście zmiennej objaśniającej o każdą kolejną jednostkę – funkcja potęgowa lub logarytmiczna.

Metoda heurystyczna

Polega na zastosowaniu modeli o różnych postaciach analitycznych i wyborze jednego z nich na podstawie wyróżnionego kryterium dobroci dopasowania modelu do rzeczywistości. Istnieją 2 warianty tej metody:

- metoda kolejnych przybliżeń,
- metoda zadowalającego wyboru.

Metoda heurystyczna – metoda kolejnych przybliżeń

Analizie poddawanych jest wiele funkcji matematycznych, na podstawie których budowane są modele ekonometryczne. W etapie weryfikacji oblicza się dla każdego modelu współczynnik determinacji R^2 (przyjmujący wartości z przedziału $[0,1]$). Do modelu dobierana jest funkcja o maksymalnej wartości współczynnika determinacji R^2 .

Metoda heurystyczna – metoda zadowalającego wyboru

Należy ustalić wartość krytyczną współczynnika determinacji (R_*^2).
Analizę rozpoczyna się od najprostszych funkcji, porównując wartość R^2 badanego modelu z wartością krytyczną R_*^2 i jeżeli ta pierwsza jest jej co najmniej równa, należy przyjąć daną postać analityczną modelu.

Metoda heurystyczna – warunki stosowalności

- brak informacji pozastatystycznych o związku łączącym zmienne modelu,
- występowanie wielu zmiennych objaśniających,
- jeśli metoda aprioryczna lub metoda oceny wzrokowej nie dają jednoznacznego rozstrzygnięcia, co do postaci analitycznej modelu – metoda heurystyczna stosowana jest wtedy pomocniczo.

Wady: subiektywizm i pracochłonność

Metoda oceny wzrokowej rozrzutu punktów

Polega na przedstawieniu na wykresie korelacyjnym rozrzutu punktów empirycznych i przypisaniu badanej zależności funkcji, której przebieg zmienności jest najbardziej zbliżony do uzyskanej smugi punktów.

Warunkiem stosowania tej metody jest występowanie tylko jednej zmiennej objaśniającej.

Metoda aproksymacji segmentowej

Jest to szczególny przypadek metody oceny wzrokowej. Ma zastosowanie, gdy rozrzut punktów jest nieciągły lub funkcja jest powikłana.

Wykres rozrzutu należy podzielić na części zwane segmentami. Każdemu segmentowi przypisuje się oddzielną funkcję opisującą związek między zmiennymi, zwaną aproksymantą segmentową. Wartość zmiennej objaśniającej, dla której dokonuje się cięcia wykresu nazywamy modulatorem.

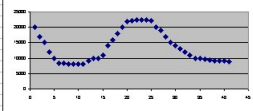
Metoda aproksymacji segmentowej

Funkcja obliczająca wartość oceny zmiennej objaśnianej w danym punkcie wynikającą z linii regresji oszacowanej na podstawie zadanego ciągu obserwacji:

$$= REGLINW(Znane_Y; Znane_X; Nowe_X; Wyraz_wolny)$$

Metoda aproksymacji segmentowej

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	t	Y	t	t ²						yar	y*	(y*-yar)*2	(y-yr)*2	(y-y*)*2
2	1	20000	1	1		REGLINP I segment				13682,46	19245,32	30945393,5	39911268,5	569539,2
3	2	17000	2	4		152,88464	-2979,24	22071,68		R*2	16724,73	9255409,31	11006049	75771,46
4	3	15000	3	9		5,6948247	117,236	509,1696		0,982962	14509,92	684676,359	1735902,65	240183,2
5	4	12000	4	16		0,9785147	663,2943	#N/D!		Phi*2	12600,87	1169653,57	2830683,14	361039,6
6	5	10000	5	25		320595872	7039349	#N/D!		0,017038	10997,59	7208568,98	13560536,8	995177
7	6	8417	6	36						suma	9700,075	15894919,7	27725105	1646281
8	7	8267	7	49						1	8708,333	24741972,1	29327244	194774,9
9	8	8117	8	64							8022,361	32036761,5	30974383	8956,576
10	9	8100	9	81							7642,158	36485292,2	31163897,8	208619,5
11	10	8110	10	100							7567,724	37390037,9	31052348,5	294063,2
12	11	8120	11	121							7799,06	34614441,3	30940999,2	103002,8
13	12	9000	12	144							8336,164	28582914	21925463,6	440677,8
14	13	9800	13	169							9179,038	20280837	15073522,2	385593,3
15	14	10000	14	196							10327,68	11254560,1	13560536,8	107375,3
16	15	11000	15	225							11782,09	3611402,57	7195609,97	611671,6
17	16	14000	16	256							13542,28	19652,4413	100629,483	209511
18	17	16000	17	289							15608,23	3708567,01	5370975,82	153485,8
19	18	18000	18	324							17979,95	18468372,6	18641122,2	402,0879
20	19	20000	19	361							20657,44	48650264,8	39911268,5	432224,2
21	20	22000	20	400		REGLINP II segment					22532,96	76331368,2	69181414,8	284051,1
22	21	22200	21	441		5,8676583	-518,705	14022,98	-97385,7		22686,14	81102201,2	72548429,5	238280,1
23	22	22500	22	484	10648	0,8374662	76,76777	2294,236	22314,23		22545,23	78548608,2	77748951,4	2045,634
24	23	22400	23	529	12167	347,71875	18	#N/D!	#N/D!		22139,42	71520419,5	57995444,1	67892,54
25	24	22400	24	576	13824	633039091	10923295	#N/D!	#N/D!		21505,97	61207309	57995444,1	799283,5
26	25	22200	25	625	15625						20680,04	48966090,1	72548429,5	2310276
27	26	20000	26	676	17576						19696,85	36172801,1	39911268,5	91902,19
28	27	19000	27	729	19683						18591,6	24099578,7	28276195,3	166794,2
29	28	17000	28	784	21952						17399,49	13816322,2	11006049	159596,1
30	29	15000	29	841	24389						16155,75	6117145,54	1735902,65	1335758
31	30	14000	30	900	27000						14895,57	1471619,3	100629,483	802039,5
32	31	13000	31	961	29791						13654,15	801,587772	465756,312	427913,6
33	32	12000	32	1024	32768						12466,71	1478058,16	2830683,14	217817,5
34	33	11000	33	1089	35937						11368,45	5354671,71	7195609,97	135753,2
35	34	10000	34	1156	39304						10394,57	10810240,4	13560536,8	155685,8
36	35	10000	35	1225	42875						9580,285	16827865,3	13560536,8	176160,4
37	36	9600	36	1296	46656						8960,798	22294127,6	16666507,5	408579,6
38	37	9500	37	1369	50653						8571,313	26123854,3	17493000,2	862458,8
39	38	9200	38	1444	54872						8447,039	27409673,8	20092478,3	566951
40	39	9200	39	1521	59319						8623,179	25596358,8	20092478,3	332722,5
41	40	9000	40	1600	64000						9134,941	20679962,7	21925463,6	18209
42	41	8850	41	1681	68921						10017,53	13431739,5	23352702,7	1363125
43												1036329214	1054291858	17962644



- 1 Miary dopasowania modelu do danych empirycznych
 - Współczynnik determinacji
 - Współczynnik zbieżności
 - Skorygowany R^2
- 2 Dobór postaci analitycznej
 - Metoda aprioryczna
 - Metoda heurystyczna
 - Metoda oceny wzrokowej rozrzutu punktów
- 3 **Transformacja liniowa**
 - Etapy transformacji liniowej
 - Przydatne funkcje
 - Przykłady transformacji liniowej
- 4 Estymacja modelu KMNK
 - Estymacja „macierzowa”
 - Estymacja – funkcja REGLINP
 - Estymacja – Analiza danych
 - Estymacja – Interpretacja parametrów struktury stochastycznej
- 5 Literatura

Etapy transformacji liniowej

- 1 Dobór postaci analitycznej modelu (np. analiza wykresu korelacyjnego)
- 2 Transformacja liniowa funkcji i dokonanie podstawień zmiennych i/lub parametrów
- 3 Szacowanie parametrów modelu liniowego
- 4 Obliczenie parametrów modelu nieliniowego

Przydatne funkcje

$$\ln(x)=\text{LN}(x)$$

$$\log_a x=\text{LOG}(x; [a])$$

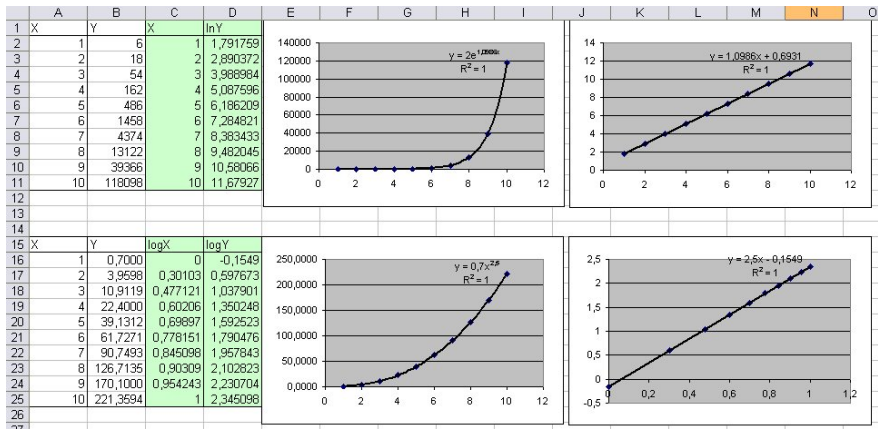
$$\log x=\text{LOG}(x)$$

$$e^x=\text{EXP}(x)$$

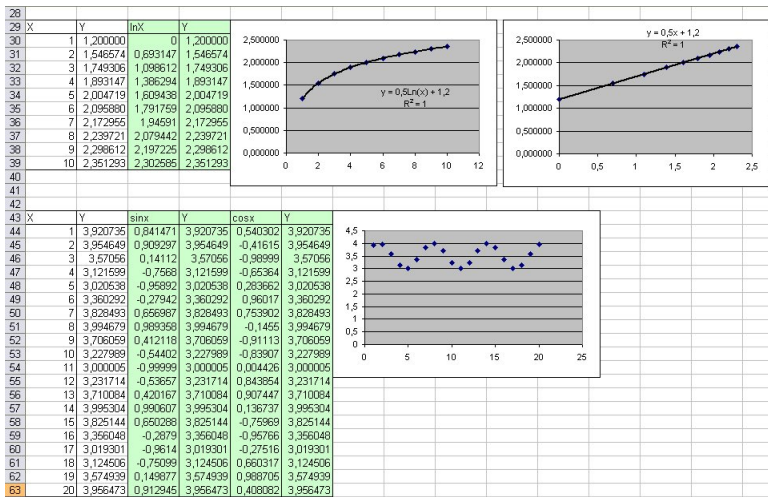
$$\sin(x)=\text{SIN}(x)$$

$$\cos(x)=\text{COS}(x)$$

Przykłady transformacji liniowej 1



Przykłady transformacji liniowej 2



- 1 Miary dopasowania modelu do danych empirycznych
 - Współczynnik determinacji
 - Współczynnik zbieżności
 - Skorygowany R^2
- 2 Dobór postaci analitycznej
 - Metoda aprioryczna
 - Metoda heurystyczna
 - Metoda oceny wzrokowej rozrzutu punktów
- 3 Transformacja liniowa
 - Etapy transformacji liniowej
 - Przydatne funkcje
 - Przykłady transformacji liniowej
- 4 Estymacja modelu KMNK
 - Estymacja „macierzowa”
 - Estymacja – funkcja REGLINP
 - Estymacja – Analiza danych
 - Estymacja – Interpretacja parametrów struktury stochastycznej
- 5 Literatura

Estymacja „macierzowa” – przydatne funkcje

$$AB = \text{MACIERZ.ILOCZYN}(\text{macierz_A}; \text{macierz_B})$$

$$A^{-1} = \text{MACIERZ.ODW}(\text{macierz})$$

$$\sum_{i=1}^n e_i^2 = \text{SUMA.KWADRATÓW}(\text{zakres_e})$$

oraz procedura transponowania macierzy (opisana na poprzedniej prezentacji)

Estymacja „macierzowa” – szacowanie parametrów modelu

- 1 Tworzymy wektor Y i macierze X oraz X^T .
- 2 Obliczamy $X^T X$.
- 3 Odwracamy otrzymaną macierz, obliczając $(X^T X)^{-1}$.
- 4 Obliczamy $X^T Y$.
- 5 Obliczamy wektor ocen parametrów: $b = (X^T X)^{-1} X^T Y$.

Estymacja „macierzowa” – standardowy błąd oceny

- 1 Obliczamy wektor wartości teoretycznych zmiennej objaśnianej: $\hat{Y} = Xb$.
- 2 Obliczamy wektor reszt: $e = Y - Xb = Y - \hat{Y}$.
- 3 Obliczamy sumę kwadratów reszt ($\sum_{i=1}^n e_i^2$).
- 4 Obliczamy wariancję resztową:

$$S^2 = \frac{\sum_{i=1}^n e_i^2}{n - (k + 1)}$$

- 5 Obliczamy standardowy błąd oceny: $S_e = \sqrt{S^2}$

Estymacja „macierzowa” – błędy średnie ocen

- 1 Obliczamy macierz wariancji i kowariancji estymatorów parametrów strukturalnych:

$$D^2(b) = S^2(X^T X)^{-1}$$

- 2 Obliczamy błędy średnie ocen parametrów strukturalnych modelu:

$$S(b_i) = \sqrt{V(b_i)}$$

(czyli pierwiastki kwadratowe z elementów głównej przekątnej otrzymanej macierzy).

Estymacja „macierzowa” – szacowanie parametrów struktury stochastycznej

7					
3		20,4144			-0,8
3		24,62911			-0,9
3		26,24089			-2,0
1		26,03754			0,5
2		27,20495			1,1
3		22,66637			2,1
4	Y^	25,89443		e	0,9
5		22,94504			1,1
5		27,92799			-0,4
7		23,59277			-1,7
3		25,59619			-2,1
3		23,58523			1,4
3		26,51956			-0,2
1		27,76229			-0,3
2		28,58324			1,4
3					
4	Suma kwadratów reszt:				
5		25,15853			
3	Wariancja resztowa:				
7		S ² 2,096544			
3	Standardowy błąd oceny:				
3		S _e 1,447945			
3					
1		0,000455	-0,00138	-0,05774	
2	D ² (b)	-0,00138	0,586557	-0,17427	
3		-0,05774	-0,17427	7,68285	
4					
5	S(b ₀)	S(b ₁)	S(b ₂)		
5	0,021323	0,76587	2,771795		
7					

Estymacja – funkcja REGLINP

REGLINP(zmienna_Y;zmiennne_X;stała;statystyka)

- I wiersz – oceny parametrów (**UWAGA**: wyraz wolny jest zawsze ostatni, a pozostałe parametry są w odwrotnej kolejności niż odpowiadające im zmienne w macierzy obserwacji!)
- II wiersz – błędy średnie ocen parametrów strukturalnych ($S(b_i)$)
- III wiersz – R^2 i standardowy błąd oceny zmiennej objaśnianej (S_e)
- V wiersz, I kolumna – wyjaśniona suma kwadratów odchyleń
- V wiersz, II kolumna – suma kwadratów reszt

Estymacja – funkcja REGLINP

	b) reglinp		
0	3,393761	0,075317	13,5681
1	0,76587	0,021323	2,771795
2	0,744711	1,447945	#N/D!
3	17,5028	12	#N/D!
4	73,3908	25,15853	#N/D!

Estymacja – Analiza danych

Postępujemy podobnie jak przy regresji liniowej z jedną zmienną objaśniającą, ale zaznaczamy całą macierz zmiennych objaśniających jako zakres wejściowy X oraz pole "Składniki resztowe".

Objaśnienia:

- Błąd standardowy (statystyki regresji) – standardowy błąd oceny zmiennej objaśnianej (S_e)
- Przecięcie – wyraz wolny
- Współczynniki – oceny parametrów strukturalnych modelu
- Błąd standardowy (tabela ze współczynnikami) – błędy średnie ocen parametrów strukturalnych ($S(b_i)$)
- Przewidywane Y – wartości teoretyczne zmiennej objaśnianej
- Składniki resztowe – reszty modelu

Estymacja – Analiza danych

PODSUMOWANIE - WYJŚCIE								
<i>Statystyki regresji</i>								
Wielokrotność R	0,862966575							
R kwadrat	0,74471131							
Dopasowany R kwadrat	0,702163194							
Błąd standardowy	1,447944815							
Obserwacje	15							
ANALIZA WARIANCJI								
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Istotność F</i>			
Regresja	2	73,39080308	36,69540154	17,50280378	0,000276815			
Resztkowy	12	25,15853025	2,096544188					
Razem	14	98,54933333						
	<i>Współczynniki</i>	<i>Błąd standardowy</i>	<i>t Stat</i>	<i>Wartość-p</i>	<i>Dołne 95%</i>	<i>Górne 95%</i>	<i>Dołne 95,0%</i>	<i>Górne 95,0%</i>
Przecięcie	13,56809872	2,771795475	4,895057677	0,000369011	7,528875183	19,60732225	7,528875183	19,60732225
X2i	0,075316808	0,02132307	3,532174624	0,004129855	0,028857829	0,121775787	0,028857829	0,121775787
X4i	3,393761238	0,765870019	4,431249632	0,000819295	1,725073818	5,062448658	1,725073818	5,062448658
SKŁADNIKI RESZTOWE - WYJŚCIE								
<i>Obserwacja</i>	<i>Przewidywane Yt</i>	<i>Składniki resztowe</i>						
1	20,41439656	-0,814396565						
2	24,62911101	-0,92911101						
3	26,2408907	-2,040890701						
4	26,03753532	0,462464681						
5	27,20494584	1,095054157						
6	22,66636912	2,133630876						
7	25,89443338	0,905566616						
8	22,94504131	1,054958687						
9	27,9279872	-0,4279872						
10	23,59276586	-1,692765862						
11	25,59619295	-2,096192955						
12	23,58523418	1,414765819						
13	26,51956289	-0,219562891						
14	27,76229022	-0,262290222						
15	28,58324343	1,41675657						

Estymacja – Standardowy błąd oceny

$$S_e = \sqrt{\frac{\sum_{i=1}^n e_i^2}{n - m - 1}}$$

- jest to pierwiastek kwadratowy z wariancji składnika resztowego (S_e^2), która jest estymatorem wariancji składnika losowego;
- informuje o przeciętnych odchyleniach wartości empirycznych (rzeczywistych) zmiennej objaśnianej od jej wartości teoretycznych, wyliczonych z modelu;
- inaczej – o ile, średnio rzecz biorąc, model myli się przy szacowaniu wartości zmiennej objaśnianej (wielkość tej „pomyłki” wyrażona jest w jednostkach tej zmiennej).

Estymacja – Błędy średnie ocen parametrów strukturalnych

$$S(b_i) = \sqrt{V(b_i)}$$

- $V(b_i)$ to wariancja estymatora parametru strukturalnego (b_i) – element z i -tego wiersza i i -tej kolumny macierzy wariancji i kowariancji estymatorów parametrów strukturalnych $D^2(b)$;
- informuje, o ile mogłaby się wahać ocena parametru strukturalnego (b_i), gdybyśmy mogli pobrać inną próbę o tej samej liczebności;
- inaczej – o ile jednostek wartość oceny b_i różni się od rzeczywistej wartości parametru β_i .

- 1 Miary dopasowania modelu do danych empirycznych
 - Współczynnik determinacji
 - Współczynnik zbieżności
 - Skorygowany R^2
- 2 Dobór postaci analitycznej
 - Metoda aprioryczna
 - Metoda heurystyczna
 - Metoda oceny wzrokowej rozrzutu punktów
- 3 Transformacja liniowa
 - Etapy transformacji liniowej
 - Przydatne funkcje
 - Przykłady transformacji liniowej
- 4 Estymacja modelu KMNK
 - Estymacja „macierzowa”
 - Estymacja – funkcja REGLINP
 - Estymacja – Analiza danych
 - Estymacja – Interpretacja parametrów struktury stochastycznej
- 5 Literatura

Literatura



Strahl D., Sobczak E., Markowska M., Bal-Domańska B. *Modelowanie ekonometryczne z Excelem*. Wrocław: AE 2002.



Ekonometria. Metody, przykłady, zadania. Red. J. Dziechciarz. Wrocław: AE 2002.



Ekonometria. Metody i analiza problemów ekonomicznych. Red. K. Jajuga. Wrocław: AE 1999.

