

EDWARD NOWAK

## MODELOWANIE PROCESÓW EKONOMICZNYCH W WARUNKACH DEZAKTUALIZACJI INFORMACJI

Jednym z głównych wyrazów złożoności zjawisk społeczno-ekonomicznych jest to, że na ich kształtowanie wpływa wiele różnorodnych czynników (w tym także o charakterze losowym), które z kolei pozostają między sobą we wzajemnych współzależnościach. Powoduje to, że obiektywne wykrycie siły i kierunku zależności występujących między rozpatrywanymi zjawiskami jest niezmiernie trudne. W takich sytuacjach badacz zmuszony jest korzystać z narzędzi statystyczno-ekonometrycznych. Chodzi tutaj zwłaszcza o modele regresyjne, które w sposób przybliżony opisują związki między zjawiskami ekonomicznymi.

Wiadomo, że ekonometryczne modelowanie zjawisk społeczno-ekonomicznych oparte jest na analizie i przetwarzaniu odpowiednich danych statystycznych będących w posiadaniu badacza. Dysponowanie dostatecznie licznym zbiorem danych statystycznych jest podstawowym warunkiem, który musi być spełniony przy stosowaniu metod statystyczno-ekonometrycznych. Dane te są liczbowymi realizacjami rozpatrywanych zjawisk (zmiennych). W zależności od rodzaju badania informacje statystyczne są zaobserwowanymi wartościami wyróżnionych zmiennych w kolejnych okresach lub momentach czasu w pewnym przedziale czasowym i dotyczą jednego obiektu badania (tzw. dane dynamiczne), bądź realizacjami zmiennych w zbiorze obiektów badania w ustalonej jednostce czasu (tzw. dane przekrojowe).

W przypadku, gdy dane statystyczne przybierają postać szeregów czasowych istotnym problemem staje się ocena wartości informacyjnych obserwacji pochodzących z różnych okresów. Ponieważ często zjawiska ekonomiczne charakteryzują się wysoką dynamiką, dlatego wartość informacyjna danych statystycznych pochodzących z bardziej odległych okresów może być niewielka. Właśnie niniejszy artykuł poświęcony jest niektórym problemom modelowania ekonometrycznego, zwłaszcza doborowi zmiennych objaśniających i wyznaczaniu wartości parametrów modelu, w przypadku gdy istnieje przypuszczenie, że dane statystyczne będące szeregami czasowymi podlegają procesowi „starzenia”. Rozważania te będą dotyczyły modelu ekonometrycznego liniowego.

Oznaczmy przez  $Y$  zjawisko ekonomiczne objaśniane, natomiast przez

$$X_1, X_2, \dots, X_K \quad (1)$$

zjawiska ekonomiczne wpływające na zmienną  $Y$  powołane do roli zmiennych objaśniających. Dane statystyczne dotyczące wyróżnionych zmiennych zaobserwowane w przedziale czasu  $[1, T]$  można przedstawić w postaci wektora obserwacji zmiennej objaśnianej.

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix}, \quad (2)$$

gdzie  $y_t$  ( $t=1, 2, \dots, T$ ) oznacza wartość zmiennej  $Y$  w jednostce czasu  $t$ , oraz w postaci macierzy obserwacji zmiennych objaśniających

$$X = \begin{bmatrix} 1 & x_{11} & x_{21} & \dots & x_{K1} \\ 1 & x_{12} & x_{22} & \dots & x_{K2} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{1T} & x_{2T} & \dots & x_{KT} \end{bmatrix}, \quad (3)$$

gdzie  $x_{kt}$  ( $k=1, 2, \dots, K, t=1, 2, \dots, T$ ) jest obserwacją zmiennej  $X_k$  w momencie lub okresie  $t$ .

Klasyczny model ekonometryczny opisujący zależność zmiennej  $Y$  od zmiennych  $X_1, X_2, \dots, X_K$

$$\hat{Y} = a_0 + a_1 X_1 + a_2 X_2 + \dots + a_K X_K, \quad (4)$$

gdzie  $\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_K$  oznaczają oceny parametrów strukturalnych, wyznaczany jest na podstawie danych statystycznych zawartych w wektorze obserwacji  $y$  i w macierzy obserwacji  $X$ .

Oparcie modelowania ekonometrycznego bezpośrednio na informacjach zawartych w wektorze obserwacji  $y$  i w macierzy obserwacji  $X$  oznacza jednocześnie przyjęcie założenia o jednakowym znaczeniu nadawanym danym pochodzącym z różnych okresów lub momentów czasu. Takie podejście jest uzasadnione tylko wówczas, gdy wraz z upływem czasu informacje o wyróżnionych zjawiskach ekonomicznych nie podlegają procesowi „starzenia”, a więc nie dezaktualizują się. Wtedy rzeczywiście informacje starsze można traktować na równi z informacjami nowszymi. Będzie to miało miejsce wówczas, gdy warunki prowadzenia działalności gospodarczej przez dany obiekt ekonomiczny nie ulegają zmianom w czasie, lub gdy zmiany zachodzące w tych warunkach są niewielkie.

Jeśli zmiany zachodzące w czasie w gospodarce charakteryzują się wysoką dynamiką, wówczas zarejestrowane dane statystyczne dotyczące przebiegu rozpatrywanych zjawisk społeczno-ekonomicznych w poszczególnych jednostkach czasu posiadają różne znaczenie z punktu wi-

dzenia ich wartości informacyjnych dla wnioskowania ekonometrycznego. Wtedy często wraz z upływem czasu informacje te dezaktualizują się, a przez to dane z jednostek czasu bardziej odległych nie zawsze mogą być traktowane na równi z danymi dotyczącymi okresów lub momentów czasu bliższych teraźniejszości. W takich przypadkach nowsze dane statystyczne posiadają znacznie większe wartości informacyjne niż starsze. W tych warunkach należy zastosować taką procedurę modelowania ekonometrycznego, która uwzględniałaby dezaktualizację posiadanych informacji, a tym samym realizowała tzw. zasadę „postarzania informacji”. Dotychczas zasada ta jest stosowana głównie w prognozowaniu ekonometrycznym<sup>1</sup>. Chodzi tutaj zwłaszcza o metodę wag harmonicznych autorstwa Z. Hellwiga<sup>2</sup>.

W literaturze ekonometrycznej istnieją pewne propozycje dotyczące modelowania w przypadku, gdy zmiany w przebiegu procesów gospodarczych mają charakter „rewolucyjny”, gdy kształtowanie się tych zjawisk w badanym przedziale czasowym nie wykazuje wyraźnej prawidłowości. Jedno z zalecanych podejść, to budowa tzw. modeli segmentowych dla odpowiednio wydzielonych okresów, w których można zaobserwować gładki przebieg zjawisk. Takie rozwiązanie charakteryzuje się jednakże tym, że nie może być uważane jako całościowy opis badanych procesów gospodarczych w przedziale czasu  $[1, T]$ , lecz jedynie jako opis ich przebiegu w wydzielonych segmentach. Inne znane podejście polega na wprowadzeniu do modelu ekonometrycznego tzw. zmiennych objaśniających zero-jedynkowych. Modele ze zmiennymi zero-jedynkowymi, podobnie jak i modele segmentowe, nie najlepiej jednakże ujmują proces starzenia się informacji, który z reguły nie jest procesem skokowym, lecz zachodzącym stopniowo.

Niekiedy postępuje się w ten sposób, że badanie ekonometryczne opiera się jedynie na danych pochodzących z niewielu ostatnich jednostek czasu, co do których istnieje przypuszczenie, że są to informacje aktualne. Takie rozwiązanie posiada jednakże poważne wady. Kłóci się ono z wymogiem oparcia badania ekonometrycznego na możliwie najbardziej obszernym materiale statystycznym. Zbyt mała liczba obserwacji w stosunku do liczby zmiennych może w ogóle uniemożliwić oszacowanie modelu<sup>3</sup>. Nawet jeśli formalny warunek, by liczba obserwacji  $T$

<sup>1</sup> O zasadzie „postarzania informacji” w prognozowaniu ekonometrycznym pisze S. Bartosiewicz w pracy *Ekonometria. Technologia ekonometrycznego przetwarzania informacji*, Warszawa 1976.

<sup>2</sup> Porównaj Z. Hellwig, *Schemat budowy prognozy statystycznej metodą wag harmonicznych*, Przegląd Statystyczny 1967, nr 2.

<sup>3</sup> Należy w tym miejscu wskazać na specjalne metody szacowania parametrów modeli ekonometrycznych w sytuacji, gdy liczba obserwacji jest mniejsza od liczby zmiennych, które zaproponował ostatnio Z. Hellwig w pracy *Wyznaczanie parametrów regresji w warunkach skąpej inormacji*, Zeszyty Naukowe Politechniki Szczecińskiej 1983, nr 236.

była większa od liczby zmiennych  $K+1$ , będzie spełniony, to zbyt mała liczba danych w istotny sposób rzutuje na jakość oszacowania modelu, zwłaszcza parametrów strukturalnych i wariancji odchyłeń losowych, a tym samym i na dokładność wnioskowania. Ponadto najczęściej trudno jest rozstrzygnąć w sposób jednoznaczny, od którego momentu można uznać informacje za aktualne. W praktyce jest bowiem tak, że chociaż w miarę upływu czasu dane starzeją się, to w dalszym ciągu posiadają one pewne wartości informacyjne. W takich przypadkach najbardziej zasadne jest takie postępowanie, które dopuszcza możliwość wykorzystania informacji z szerokiego przedziału czasu, przy czym danym tym nadaje się zróżnicowane znaczenie.

Dla zróżnicowania wpływu obserwacji z różnych okresów na wyniki modelowania proponuje się wyznaczanie modelu na podstawie informacji ważonych w ten sposób, by informacjom starszym nadawać mniejsze wagi niż informacjom bardziej aktualnym.

W przypadku doboru zmiennych objaśniających proponuje się procedurę opartą na odpowiednio zwielokrotnionych obserwacjach pierwotnych<sup>4</sup>, która jest procedurą numeryczną odpowiadającą ważeniu. Proponuje się, by obserwacje zmiennej objaśnianej  $Y$  oraz potencjalnych zmiennych objaśniających  $X_1, X_2, \dots, X_M$  z okresu lub momentu  $t$  zwielokrotnić  $t$  razy. Liczba danych wykorzystanych do doboru zmiennych objaśniających wynosić będzie teraz

$$T^* = T(T+1)/2. \quad (5)$$

Tak zwielokrotnione obserwacje można przedstawić za pomocą wektora  $y^*$  i macierzy  $X^*$ , w postaci:

$$y^* = \begin{bmatrix} y_1 \\ y_2 \\ y_2 \\ y_3 \\ y_3 \\ y_3 \\ \vdots \\ y_T \\ y_T \\ \vdots \\ y_T \end{bmatrix}, \quad X^* = \begin{bmatrix} x_{11} & x_{21} & \dots & x_{M1} \\ x_{12} & x_{22} & \dots & x_{M2} \\ x_{12} & x_{22} & \dots & x_{M2} \\ x_{13} & x_{23} & \dots & x_{M3} \\ x_{13} & x_{23} & \dots & x_{M3} \\ x_{13} & x_{23} & \dots & x_{M3} \\ \vdots & \vdots & \vdots & \vdots \\ x_{1T} & x_{2T} & \dots & x_{MT} \\ x_{1T} & x_{2T} & \dots & x_{MT} \\ \vdots & \vdots & \vdots & \vdots \\ x_{1T} & x_{2T} & \dots & x_{MT} \end{bmatrix} \quad (6)$$

<sup>4</sup> Zwielokrotnianie obserwacji zaproponowała po raz pierwszy S. Bartosiewicz dla wyznaczania regresji segmentowych w sytuacji niejednorodności zbioru przekrojowych danych statystycznych. Porównaj w tym zakresie pracę S. Bartosiewicz, J. Dziechciarz, E. Nowak, W. Pluta, *Problem separowalności zbiorów obiektów i zbiorów cech. Autonomiczne funkcje quasi-regresji, w: Zastosowanie technik wielowymiarowej analizy porównawczej w dynamicznych i przekrojowych badaniach ekonometrycznych*, red. S. Bartosiewicz, Wrocław 1982 (maszynopis Akademii Ekonomicznej).

Przekształcone w ten sposób zmienne oznaczymy, jako  $Y^*$  oraz  $X_1^*$ ,  $X_2^*$ , ...,  $X_M^*$

Na podstawie danych zawartych w wektorze  $y^*$  oraz w macierzy  $X^*$  przeprowadzana jest redukcja zbioru potencjalnych zmiennych objaśniających. Ponieważ budowany model jest modelem liniowym, dlatego podstawą do jej przeprowadzenia będą współczynniki korelacji:

— przekształconej zmiennej objaśnianej z przekształconymi potencjalnymi zmiennymi objaśniającymi:

$$r_i^* = r(Y^*, X_i^*) \quad (i=1, 2, \dots, M), \quad (7)$$

— między przekształconymi potencjalnymi zmiennymi objaśniającymi:

$$r_{ij}^* = r(X_i^*, X_j^*) \quad (i, j=1, 2, \dots, M, i \neq j).$$

Dla zrealizowania tego zadania można wykorzystać dowolną procedurę statystyczną opartą na współczynnikach korelacji liniowej.

Postulat ważenia informacji przy szacowaniu modelu zmiennej  $Y$  względem wybranych zmiennych objaśniających  $X_1, X_2, \dots, X_K$  ( $K \leq M$ ):

$$\hat{Y} = a_0 + a_1 X_1 + a_2 X_2 + \dots + a_K X_K \quad (9)$$

można zrealizować za «pomocą tzw. ważonej metody najmniejszych kwadratów»<sup>5</sup>.

W tym przypadku macierz wagowa będzie miała następującą postać:

$$W = \begin{bmatrix} w_1 & 0 & \dots & 0 \\ 0 & w_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & w_T \end{bmatrix}. \quad (10)$$

Elementy macierzy wagowej  $W$  zdefiniowane są natomiast następująco:

$$w_t = \sqrt{t} \quad (t=1, 2, \dots, T). \quad (11)$$

W przypadku ważonej metody najmniejszych kwadratów podstawą do wyznaczenia parametrów modelu jest ważony wektor obserwacji zmiennej objaśnianej.

$$y_w = Wy, \quad (12)$$

oraz ważona macierz obserwacji zmiennych objaśniających:

$$X_w = WX. \quad (13)$$

Stosując metodę najmniejszych kwadratów dla danych zawartych w wektorze  $y_w$  oraz w macierzy  $X_w$  otrzymujemy oceny parametrów strukturalnych modelu (9).

Szczególnym przypadkiem ważonej metody najmniejszych kwadra-

<sup>5</sup> O metodzie tej pisze np. H. Theil w pracy *Zasady ekonometrii*, Warszawa 1978.

Tabela 1

Wartości zmiennych w latach 1965 - 1982

Lata	$t$	$Y$	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$
1965	1	104,1	100,0	140,7	131,0	2700	1487	165,7	8,0
1966	2	109,7	108,3	154,7	145,1	2559	1593	187,8	7,9
1967	3	110,0	113,8	168,8	160,6	2825	1700	219,2	8,2
1968	4	113,4	118,0	182,3	180,5	3110	1802	250,7	8,4
1969	5	113,7	124,1	200,5	202,7	3482	1987	269,1	8,7
1970	6	112,4	114,5	220,6	224,5	3899	2177	295,8	9,2
1971	7	119,8	119,4	239,9	248,4	3656	2409	321,7	10,4
1972	8	130,6	120,7	259,5	278,8	4457	2790	327,2	11,9
1973	9	141,3	130,1	267,7	319,2	5376	3227	337,6	13,9
1974	10	147,2	138,6	279,5	364,8	6379	3713	252,7	13,9
1975	11	145,7	137,6	319,6	401,2	7200	4253	360,5	14,5
1976	12	133,0	133,5	375,9	434,0	7593	4858	366,3	13,8
1977	13	151,2	140,2	442,1	472,6	7060	5532	395,3	14,3
1978	14	155,1	130,1	511,6	514,5	8689	6067	411,7	15,3
1979	15	157,1	137,1	518,3	573,1	8771	6754	396,9	15,9
1980	16	148,3	132,0	588,1	619,4	8393	7511	398,0	16,0
1981	17	135,1	111,9	618,1	670,0	7173	8125	389,5	15,4
1982	18	130,8	133,0	595,6	710,2	4950	8707	412,7	13,9

Źródło: Roczniki Statystyczne GUS z lat 1966-1983.

tów jest procedura szacowania modelu na podstawie informacji zwielokrotnionych w proponowany sposób.

Dzięki zaproponowanej procedurze modelowania o (poziomie wartości ocen parametrów strukturalnych  $a_0^*$ ,  $a_1^*$ ,  $a_2^*$ , ...,  $a_K^*$  równania (9) w większym stopniu decydują informacje nowsze, natomiast w mniejszym stopniu ich poziom zależy od danych starszych. Współczynniki  $a_0^*$ ,  $a_1^*$ ,  $a_2^*$ , ...,  $a_K^*$  wykorzystywane są jako narzędzia analizy ekonomicznej. Model (9) może być również podstawą do ustalenia prognoz zjawisk ekonomicznych uwzględniających zasadę „postarzania informacji”.

Proponowana koncepcja zostanie zastosowana do zbudowania liniowego modelu ekonometrycznego wyjaśniającego dynamikę globalnej produkcji zwierzęcej rolnictwa w stosunku do poziomu tej produkcji z lat 1960-1965 ( $Y$ ). Lista pierwotnych zmiennych objaśniających przedstawia się następująco:

$X_1$  — wskaźnik dynamiki globalnej produkcji roślinnej w stosunku do lat 1960 - 1965 wzięty z rocznym opóźnieniem,

$X_2$  — wskaźnik dynamiki nakładów inwestycyjnych w rolnictwie w stosunku do roku 1960 wzięty z trzyletnim opóźnieniem,

$X_3$  — ciągniki w rolnictwie w jednostkach fizycznych w tys. sztuk,

$X_4$  — dostawy pasz treściwych dla rolnictwa w tys. ton,

$X_5$  — zużycie energii elektrycznej w rolnictwie w mln kWh,

$X_6$  — krowy wpisane do ksiąg inwentarza zarodowego w tys. sztuk,

$X_7$  — lochy wpisane do ksiąg inwentarza zarodowego w tys. sztuk.

Przeprowadzone badanie »będzie dotyczyło 18 kolejnych lat okresu 1965-1982. Wartości zmiennej objaśnianej oraz potencjalnych zmiennych objaśniających w poszczególnych latach tego okresu podane są w tabeli 1.

Ponieważ  $T=18$ , dlatego przekształcone zmienne mają

$$T^* = 18(18+1)/2 = 171$$

obserwacji. Wartości wag pierwotnych obserwacji zmiennych z poszczególnych lat okresu 1965-1982 wyznaczone według wzoru (12) podane są w tabeli 2.

Tabela 2

Wagi obserwacji z kolejnych lat okresu 1965 - 1982

Lata	$t$	$w_t$	Lata	$t$	$w_t$
1965	1	1,000	1974	10	3,162
1966	2	1,414	1975	11	3,317
1967	3	1,732	1976	12	3,464
1968	4	2,000	1977	13	3,606
1969	5	2,236	1978	14	3,742
1970	6	2,449	1979	15	3,873
1971	7	2,646	1980	16	4,000
1972	8	2,828	1981	17	4,123
1973	9	3,000	1982	18	4,243

Źródło: Obliczenia własne.

Wektor współczynników korelacji przekształconej zmiennej objaśnianej z przekształconymi potencjalnymi zmiennymi objaśniającymi przedstawia się następująco:

$$R_0^* = \begin{bmatrix} 0,692 \\ 0,403 \\ 0,492 \\ 0,882 \\ 0,462 \\ 0,742 \\ 0,870 \end{bmatrix} .$$

Macierz współczynników korelacji między przekształconymi potencjalnymi zmiennymi objaśniającymi jest natomiast postaci:

$$R^* = \begin{bmatrix} 1 & 0,174 & 0,236 & 0,491 & 0,204 & 0,522 & 0,496 \\ 0,174 & 1 & 0,688 & 0,463 & 0,685 & 0,617 & 0,586 \\ 0,236 & 0,688 & 1 & 0,593 & 0,998 & 0,873 & 0,799 \\ 0,491 & 0,463 & 0,593 & 1 & 0,572 & 0,748 & 0,895 \\ 0,204 & 0,685 & 0,998 & 0,572 & 1 & 0,850 & 0,771 \\ 0,522 & 0,617 & 0,873 & 0,748 & 0,850 & 1 & 0,896 \\ 0,496 & 0,586 & 0,799 & 0,895 & 0,771 & 0,896 & 1 \end{bmatrix} .$$

Do doboru zmiennych objaśniających została zastosowana metoda wskaźników pojemności informacji Z. Hellwiga<sup>6</sup>. Optymalna kombinacja zawiera zmienne  $X_1^*$ ,  $X_4^*$  i  $X_7^*$ , dla (których wskaźnik integralnej pojemności informacji

$$H(X_1^*, X_4^*, X_7^*) = 0,883.$$

Oznacza to, że w liniowym modelu ekonometrycznym wyjaśniającym dynamikę zwierzęcej produkcji rolniczej w charakterze zmiennych objaśniających powinny znaleźć się następujące wielkości: dynamika roślinnej produkcji rolniczej wzięta z rocznym opóźnieniem ( $X_1$ ), dostawy pasz treściwych dla rolnictwa ( $X_4$ ) oraz obsada loch wpisanych do ksiąg inwentarza zarodowego ( $X_7$ ).

Po oszacowaniu parametrów strukturalnych ważoną metodą najmniejszych kwadratów z wagami zawartymi w tablicy 2 zbudowany model przyjął następującą postać:

$$\hat{Y} = 30,6 + 0,468X_1 + 0,00335X_4 + 1,91X_7.$$

Współczynnik zbieżności powyższego modelu obliczony dla danych przekształconych wynosi 0,117.

O wartościach ocen parametrów strukturalnych zbudowanego modelu w mniejszym stopniu zdecydowały dane sprzed kilkunastu lat niż dane nowsze. Można więc przypuszczać, że informacje jakie niosą te współczynniki o wpływie wyodrębnionych czynników na dynamikę produkcji zwierzęcej są bardziej aktualne od tych, jakie uzyskano by klasyczną procedurą modelowania ekonometrycznego. Zastosowanie proponowanej procedury w badaniach ekonomiczno-rolniczych jest tym bardziej uzasadnione, ponieważ cechą charakterystyczną rolnictwa jest zachodząca w miarę upływu czasu zmienność zjawisk: warunków produkcji rolniczej oraz ich wpływu na rezultaty produkcji.

## MODELLING OF ECONOMIC PROCESSES UNDER CONDITIONS OF OUTDATING INFORMATION

### Summary

The article is devoted to econometric modelling of economic phenomenae interrelations in a situation when statistical data being the basis for modelling are becoming out of date in process of time. The article presents a concept of creating econometric model (choice of describing variables and assessing parameters) which would attribute lesser weight to statistical data of earlier periods of time, while more weight to the data originating in periods closer to the present time. The idea can be summarized as multiplying initial information proportionally to numbers of respective units of time. The proposed concept was applied in construction of the econometric model explaining formation of dynamics of total animal output of Polish agriculture in the years 1965 - 1982.

<sup>6</sup> Porównaj pracę Z. Hellwiga *Problem optymalnego wyboru predyktant*, Przegląd Statystyczny 1963. nr 3-4.