

STANISŁAW WYDYMUS

## ZMIENNE SYNTETYCZNE W MODELOWANIU EKONOMETRYCZNYM

### 1. WSTĘP

W kompleksowych badaniach ekonometrycznych dotyczących wielu obiektów ekonomicznych charakteryzowanych zbiorem wielu różnych cech diagnostycznych<sup>1</sup> występuje najczęściej na wstępie problem merytorycznie uzasadnionego ograniczenia zbioru zmiennych pierwotnych, co sprowadza się do określenia podzbioru zmiennych uznanych za najistotniejsze. Celem takiego postępowania jest z jednej strony umożliwienie dokonywania sensownych i przejrzystych analiz, a z drugiej eliminacja informacji powtórzonej lub zbędnej. Problem ten, powszechnie znany w literaturze przedmiotu, dotyczy zarówno badań prowadzonych na podstawie metod wielowymiarowej analizy porównawczej (WAP), jak również modelowania ekonometrycznego, jeszcze bardziej zaś nabiera na znaczeniu, gdy następuje połączenie wyników analizy wielowymiarowej z wynikami modelowania ekonometrycznego.

Problem korzyści płynących z poprawnego doboru zmiennych do modelu ekonometrycznego nie budzi żadnych wątpliwości<sup>2</sup>. Modele, które zawierają zmienne błędnie określone oraz których liczba jest zbyt duża, powodują szereg trudności wynikających z negatywnego zjawiska współliniowości<sup>3</sup>, braku koincydencji<sup>4</sup> itp. W konsekwencji model taki nie jest

<sup>1</sup> Często występujące trudności w uzyskiwaniu kompletnego materiału statystycznego można w dużym stopniu wyeliminować poprzez zastosowanie licznych procedur szacowania brakujących informacji. Por. T. Grabiński, S. Wydymus, A. Zeliaś, *The Problem of Data Completeness in Econometric Modelling*, Materiały konferencji na temat: *Problems of Building and Estimation of Large Econometric Models*, Błażejewko, 6-8. 12. 1978.

<sup>2</sup> Zagadnienie to rozważa szczegółowo R. R. Hocking, *The Analysis and Selection of Variables in Linear Regression*, Biometrics, 9/1976.

<sup>3</sup> Podstawowe informacje na temat zjawiska współliniowości znaleźć można w pracy: A. Zeliaś, *Z problematyki badania współliniowości w jednorównaniowych modelach regresji*, Przegląd Statystyczny, z. 3/1977.

<sup>4</sup> O problemie koincydencji traktuje praca Z. Hellwiga, *Przechodność relacji skorelowania zmiennych losowych i płynące stąd wnioski ekonometryczne*, Przegląd Statystyczny, z. 1/1976.

dobrym narzędziem analizy oraz predykcji. Ograniczanie jednakże liczby zmiennych w modelu prowadzi z drugiej strony do utraty informacji. Dlatego też wybór zmiennych powinien być rozsądnym kompromisem między tymi dwoma sprzecznymi wymogami. Jednostronne ograniczanie bowiem liczby zmiennych przez stosowanie różnych procedur ich doboru prowadzi często do konstrukcji takiego modelu, którego użyteczność praktyczna może być niewielka. Model taki może zawierać duży ładunek informacji wewnętrznej o zmiennej endogenicznej, w sensie wysokiej wartości kwadratu współczynnika korelacji wielowymiarowej ( $R^2$ ) i małych wartości błędów ocen (parametrów, ale jednocześnie model ten może mieć mały ładunek informacji analitycznej, w sensie ograniczonej możliwości wykorzystania go na przykład w prognozowaniu wariantowego rozwoju zmiennej endogenicznej<sup>5</sup>.

Prezentowany artykuł jest próbą konstrukcji modeli ekonometrycznych, których celem jest uwzględnienie jak największej liczby informacji, zwłaszcza w sytuacjach, gdy liczba pierwotnych zmiennych diagnostycznych ( $m$ ) znacznie przekracza liczbę analizowanych, obiektów ( $k$ ) lub liczbę okresów objętych badaniem ( $n$ ), czyli gdy  $m \gg k$  lub  $m \gg n$ . Modele takie dotyczą zmiennych syntetycznych (taksonomicznych mierników rozwoju) konstruowanych na podstawie zasad stworzonych przez wielowymiarową analizę porównawczą. Powstaje na wstępie tak sformułowanego zagadnienia pytanie, czy zmienna syntetyczna, będąca wynikiem odpowiedniego przekształcenia grupy zmiennych diagnostycznych, zawiera większy ładunek informacji niż którakolwiek zmienna z tej grupy. Pomijając problem sposobu mierzenia liczby informacji, odpowiedź na to pytanie jest pozytywna, co wynika chociażby z dosyć już licznych wyników badań empirycznych prowadzonych przy wykorzystaniu WAP. Założenie większej wartości informacyjnej zmiennych syntetycznych umożliwia konstrukcję różnego typu modeli ekonometrycznych, operując na etapie ich specyfikacji tylko zbiorem zmiennych syntetycznych lub zbiorami mieszanymi składającymi się zarówno z pierwotnych zmiennych diagnostycznych, jak również ze zmiennych syntetycznych. Modele takie mogą znaleźć zastosowanie zwłaszcza w analizie i wariantowym prognozowaniu rozwoju społeczno-gospodarczego różnego rodzaju obiektów ekonomicznych (kraje świata, regiony, przedsiębiorstwa itp.).

## 2. PROBLEM DOBORU ZMIENNYCH DO MODELI SYNTETYCZNYCH

Procedura doboru optymalnego zbioru zmiennych do modeli ekonometrycznych jest wielostopniowa. W pierwszym etapie określa się zbiór potencjalnych zmiennych objaśniających (pomijamy tu problematykę usta-

<sup>5</sup> Szerokie rozważania nad istotą modelu ekonometrycznego oraz jego walorami poznawczymi i instrumentalnymi zawarte są w pracach: Z. Hellwig, *Rozważania nad istotą modelu ekonometrycznego*, *Ekonomista*, nr 2/1974 oraz Z. Czerwiń-

lania zmiennych endogenicznych) na drodze merytorycznej analizy, na podstawie informacji płynących z teorii ekonomii politycznej, teorii wzrostu gospodarczego, ekonomik branżowych itp. Podstawową metodą służącą do ustalania wstępnego zestawu zmiennych są badania ankietowe, w których wykorzystuje się doświadczenia specjalistów z dziedziny będącej przedmiotem modelowania. Zasadniczym elementem takich badań, jest poprawne określenie celów analizy oraz świadomość kryteriów merytorycznych, które powinny spełniać wstępnie przyjęte zmienne objaśniające<sup>6</sup>. Na etapie określania wstępnego zbioru zmiennych możliwe jest stosowanie zasad formalno-statystycznych, których celem jest badanie zgodności opinii ekspertów, a pośrednio — badanie poprawności ustalenia wstępnego zbioru zmiennych<sup>7</sup>.

Problem określania wstępnego zbioru zmiennych objaśniających jest zasadniczo odmienny w przypadku konstrukcji modeli przyczynowo-skutkowych oraz symptomatycznych. Margines swobody w przypadku modeli przyczynowo-skutkowych jest znacznie mniejszy w porównaniu z modelami symptomatycznymi, a sam dobór wstępnego zbioru zmiennych jest ściśle zdeterminowany wiedzą merytoryczną dotyczącą rozpatrywanych zjawisk. Niemniej jednak, nawet w tym przypadku wstępny zbiór zmiennych może zawierać znaczną ich liczbę. W sytuacji, gdy konstruowane są modele dotyczące rozwoju społeczno-gospodarczego obiektów ekonomicznych, ważną sprawą jest prawidłowe zdefiniowanie mierników i czynników tego rozwoju<sup>8</sup>. Mylenie w praktyce tych pojęć prowadzi często do nieporozumień, a ustalone modele nie mają jasnej interpretacji. Poprawne określenie mierników rozwoju społeczno-gospodarczego prowadzi w efekcie do konstrukcji modeli przyczynowo-skutkowych, których przydatność praktyczna jako narzędzia analizy prawidłowości ekonomicznych i wariantowego prognozowania jest bardzo wysoka. Często jednakże, z różnych powodów (brak ugruntowanej teorii, trudności w uzyskaniu odpowiednich informacji statystycznych), nie dokonuje się ścisłego określenia charakteru zmiennych, rezygnując z pojęcia związku przyczynowo-skutkowego na korzyść związku korelacyjnego, co w efekcie prowadzi do budowy modeli symptomatycznych. Jeśli tylko związek korelacyjny jest obrazem pośredniego powiązania przyczynowego zmiennych, to przydatność takich modeli jest również niekwestionowana.

ski, *Przyczynek do dyskusji nad problemem „dobrego” modelu ekonometrycznego*, Przegląd Statystyczny, z. 4/1976.

<sup>6</sup> Na temat różnych kryteriów doboru zmiennych piszą autorzy „pracy T. Grabińskiego, S. Wydymus, A. Zeliaś, *Metody doboru zmiennych w modelach ekonometrycznych*, Warszawa 1982.

<sup>7</sup> Por. T. Grabiński, S. Wydymus, A. Zeliaś, *Metody prognozowania rozwoju społeczno-ekonomicznego*, Kraków 1982.

<sup>8</sup> Rozważania tego typu są przedmiotem teorii ekonomii politycznej; por. O. Lange, *Ekonomia polityczna*, t. I i II, Warszawa 1978.

Następnym etapem doboru zmiennych do modeli syntetycznych jest grupowanie wstępnego zbioru cech diagnostycznych w jednorodne podzbiory. Ma to bardzo ważne znaczenie przy konstrukcji zmiennych syntetycznych, determinuje bowiem ich walory interpretacyjno-poznawcze. Problem grupowania cech można rozstrzygnąć albo na drodze analizy merytorycznej, albo też poprzez zastosowanie algorytmów taksonomii numerycznej. Prowadząc rozważania na gruncie rozwoju społeczno-gospodarczego obiektów (np. krajów świata) możliwe jest przeprowadzenie grupowania jednorodnego, działowego, problemowego lub też względnie swobodnego. Klasycznym przykładem grupowania jednorodnego może być tworzenie podzbiorów cząstkowych zmiennych, których suma stanowi zmienną podstawową (np. poszczególne składniki majątku trwałego, zatrudnienia, popytu konsumpcyjnego itp.). Grupowanie działowe dotyczy tworzenia podzbiorów zmiennych związanych z działami gospodarki narodowej, a więc obejmuje zmienne cząstkowe charakteryzujące rozwój przemysłu, rolnictwa, budownictwa, handlu wewnętrznego i zagranicznego itp. Grupowanie problemowe związane jest z wyodrębnieniem podstawowych aspektów funkcjonowania obiektów ekonomicznych, a więc przykładem może być tworzenie podzbioru mierników i czynników dotyczących poziomu życia, potencjału ekonomicznego, a przy dalszej dezagregacji — poziomu zdrowotności społeczeństwa, rozwoju infrastruktury społecznej, wielkości i wykorzystania bazy surowcowej, sytuacji demograficznej itp. Grupowanie względnie swobodne występuje wtedy, gdy trudno jest (co może wynikać z charakteru dostępnych pierwotnych cech diagnostycznych) ustalić ścisłe, merytoryczne granice między poszczególnymi grupami zmiennych lub też gdy kryterium tworzenia jednorodnych grup cech jest podobieństwo współzmienności poszczególnych zmiennych. Ten ostatni wariant grupowania zachodzi najczęściej, gdy podstawą analizy są metody taksonomii numerycznej.

Najczęściej w praktyce badawczej dominuje pierwsze podejście, opierające się na analizie merytorycznej. Jednakże coraz częściej przy tworzeniu podzbiorów relatywnie jednorodnych cech stosuje się dualne algorytmy taksonomiczne<sup>9</sup>. Pamiętać jednakże należy, iż przy takim podejściu kryterium jednorodności zmiennych jest wielkość ustalonej metryki odległości. W związku z tym najbardziej poprawna sytuacja występuje wtedy, gdy metody taksonomiczne znajdują zastosowanie do wyodrębniania podgrup podobnych zmiennych, ale w ramach wcześniej ustalonych — drogą analizy merytorycznej — szerokich ich grup (grupowanie działowe lub problemowe).

<sup>9</sup> Taksonomiczne zasady doboru zmiennych przedstawione są np. w pracy T. Grabiński, S. Wydymus, A. Zeliaś, *Metody doboru zmiennych*, a także U. Siedlecka, *Taksonomiczna metoda wyboru zmiennych*, Prace Naukowe Akademii Ekonomicznej we Wrocławiu, 4/1976.

Przy stosowaniu dualnych algorytmów taksonomicznych przedmiotem klasyfikacji jest  $m$ -elementowy zbiór  $\Omega$  wstępnych zmiennych, rozpatrywany w  $k$ -wymiarowej przestrzeni obserwacji (obiektów),  $n$ -wymiarowej przestrzeni okresów lub  $(k \times n)$ -wymiarowej przestrzeni tzw. obiekt-okresów. Dla ustalenia uwagi w dalszych rozważaniach przyjęto, że realizacje poszczególnych zmiennych określone są na zbiorze  $n$ -elementowym o dowolnym charakterze, przy czym wśród zmiennych podlegających analizie wyróżnia się zmienną endogeniczną  $Y$  oraz potencjalne zmienne objaśniające  $X_j$  ( $j=1, \dots, m$ ). Wartości tych zmiennych przedstawione są w postaci następującego wektora i macierzy:

Punktem wyjścia procedur taksonomicznych jest odpowiednio zdefiniowana funkcja odległości, przyporządkowująca każdej parze elementów klasyfikowanego zbioru miarę ich wzajemnej odległości lub podobieństwa. Kryterium klasyfikacji jest funkcjonal określony na zbiorze wszystkich możliwych podzbiorów zbioru  $\Omega$  i mierzący — w kategoriach wynikających z przyjętej metryki odległości — stopień wewnętrznej jednorodności (homogeniczności) poszczególnych podzbiorów oraz stopień niejednorodności (heterogeniczności) między wyodrębnionymi podzbiarami.

Centralnym zagadnieniem w dualnych algorytmach taksonomicznych jest poprawne zdefiniowanie metryki odległości służącej do ustalenia podzbiorów zmiennych objaśniających. Spośród różnych mierników odległości najczęściej używa się miernik odległości Euklidesa zdefiniowany wzorem

$$d_{jp}^2 = \frac{1}{n} \sum_{i=1}^n (x'_{ij} - x'_{ip})^2 \quad (j, p=1, \dots, m), \quad (2)$$

gdzie  $d_{jp}$  to odległość między  $j$ -tą oraz  $p$ -tą zmienną, natomiast  $x'_{ij}$  oraz  $x'_{ip}$  to standaryzowane na (0, 1) wartości tych cech dla  $i$ -tej obserwacji.

Stosowanie w prowadzonych rozważaniach metryki określonej wzorem (2) nie znajduje jednakże uzasadnienia. Można bowiem wykazać<sup>10</sup>, że zachodzi następujący związek między odległością Euklidesa  $d_{jp}^2$  a współczynnikiem korelacji liniowej  $r_{jp}$ :

$$d_{jp} = \sqrt{2(1 - r_{jp})}. \quad (3)$$

Jak łatwo zauważyć, podzbiór ostatecznych zmiennych reprezentujących poszczególne podgrupy, otrzymany w wyniku zastosowania miernika (2) stanowi zarazem grupę zmiennych najbardziej skorelowanych ze sobą, tylko że w sposób ujemny.

Aby otrzymać poprawne wyniki grupowania przy minimalizacji stop-

<sup>10</sup> Por. w tej sprawie pracę U. Siedlecka, *Taksonomiczna metoda*.

nia wzajemnego skorelowania cech, należy skorzystać z innych, zmodyfikowanych mierników nie spełniających warunków trójkąta, jak np.<sup>11</sup>:

$$d_{jp}^{(1)} = 1 - r_{jp}^2, \quad (4)$$

$$d_{jp}^{(2)} = \sqrt{1 - r_{jp}^2}, \quad (5)$$

$$d_{jp}^{(3)} = 1 - |r_{jp}|. \quad (6)$$

Jak się okazuje, największe zastosowanie w problematyce ustalania podzbiorów zmiennych objaśniających mogą znaleźć techniki grafowe, procedury aglomeracyjne, algorytmy obszarowe, metody czynnikowe, jak również metoda Czekanowskiego. Należy w tym miejscu podkreślić, iż wymienione metody prowadzą tylko do ustalenia bardziej jednorodnych grup zmiennych<sup>12</sup>.

W wyniku zastosowania analizy merytorycznej lub też metod taksonomii numerycznej, otrzymujemy następujący ciąg macierzy wartości zmiennych pierwotnych, tworzących ustalone ich podzbiory

$$\{\mathbf{X}_1^{(1)}, \mathbf{X}_2^{(1)}, \dots, \mathbf{X}_p^{(1)}\}. \quad (7)$$

Macierze takie mogą mieć oczywiście różne wymiary, w zależności od liczby potencjalnych zmiennych objaśniających wchodzących w skład poszczególnych podgrup.

W dalszym etapie specyfikacji zmiennych do modelu syntetycznego proponuje się wykorzystać pewną modyfikację metody pojemności informacji Z. Hellwiga<sup>13</sup>, zwaną dalej dwustopniową metodą pojemności informacji. W pierwszym etapie<sup>14</sup> rozważanego problemu dokonuje się wyboru optymalnego podzbioru zmiennych objaśniających w ramach poszczególnych podgrup zmiennych potencjalnych<sup>15</sup>. Następuje w ten sposób

<sup>11</sup> Mierniki te zapewniają właściwy dobór zmiennych objaśniających, ponieważ ich maksymalne odległości (równe jedności) odpowiadają minimalnym wartościom współczynników korelacji (równym zero) i odwrotnie. Można także w tym względzie stosować inne miary odległości (jak np. bezwzględna wartość współczynnika korelacji  $r$ ), jednakże niezbędna jest wówczas modyfikacja reguł decyzyjnych w procedurach taksonomicznych.

<sup>12</sup> Dalsza analiza może polegać na wyborze zmiennych reprezentujących poszczególne podgrupy, jednakże ten sposób podejścia jest sprzeczny z podstawową tezą pracy.

<sup>13</sup> Por. w tej sprawie prace: Z. Hellwig, *Problem optymalnego wyboru predykant. Przegląd Statystyczny*, z. 3-4/1969 jak również A. Zeliaś, *Uwagi o problemie optymalnego wyboru wektora zmiennych objaśniających*, *Przegląd Statystyczny*/z. 2/1970.

<sup>14</sup> Zwróćmy uwagę, iż w pierwszym etapie doboru zmiennych możliwe jest również stosowanie zasad analizy czynnikowej w celu redukcji zbioru zmiennych w poszczególnych podgrupach.

<sup>15</sup> Zauważmy, że dwustopniową metodą pojemności informacji można stosować bez uprzedniego podziału wstępnego zbioru zmiennych na jednorodne podgrupy. W takim przypadku wyłania się w pierwszym etapie określoną liczbę podzbiorów

wstępna redukcja zmiennych, które są słabo skorelowane ze zmienną endogeniczną, a jednocześnie silnie skorelowane między sobą<sup>16</sup>. W wyniku zastosowania metody Z. Hellwiga otrzymujemy w konsekwencji ciąg wartości pojemności integralnych optymalnych nośników informacji w ramach poszczególnych grup zmiennych, który możemy zapisać następująco:

$$\{H_1, H_2, \dots, H_p\}. \quad (8)$$

Każdy wyraz tego ciągu jest — jak wiadomo — sumą pojemności indywidualnych nośników informacji  $h$ , z których każdy charakteryzuje zasób informacji, jaki wnosi o zmiennej endogenicznej określona zmienna cząstkowa wchodząca w skład optymalnej kombinacji zmiennych wyłonionych w ramach poszczególnych podgrup. Wartości tych lokalnie optymalnych zmiennych można przedstawić w postaci następującego ciągu macierzy, z których każda jest odpowiednią częścią macierzy stanowiących ciąg (7):

$$\{\mathbf{X}_1^{(2)}, \mathbf{X}_2^{(2)}, \dots, \mathbf{X}_p^{(2)}\}.$$

Następny etap analizy «polega na konstrukcji agregatowych zmiennych syntetycznych w ramach każdego (podzbioru optymalnych zmiennych objaśniających otrzymanych przy zastosowaniu pierwszego kroku dwustopniowej metody pojemności informacji. W ten sposób unika się straty informacji wynikającej z faktu oparcia się tylko na pojedynczych zmiennych reprezentujących poszczególne ich podzbiory. Zmienna syntetyczna zawierająca w sobie informacje wnoszone przez poszczególne zmienne cząstkowe ma większe walory informacyjne, o czym wspomniano już na wstępie niniejszej pracy.

Spotykane w literaturze propozycje konstruowania zmiennych syntetycznych różnią się w zależności od<sup>17</sup>:

— sposobu uwzględniania w obliczeniach tzw. stymulant, destymulant i nominant<sup>18</sup>,

zmiennych mających wystarczająco dużą wartość pojemności integralnej informacji  $H$ , lecz różniących się elementami.

<sup>16</sup> Postulat silnego skorelowania zmiennych objaśniających ze zmienną endogeniczną w połączeniu z jednoczesnym postulatem słabego skorelowania wzajemnego zmiennych objaśniających należy rozpatrywać w rozsądny sposób, mając na względzie uwagi zawarte w pracy J. Hozer, *Własność macierzy  $R_n$  a dobór zmiennych objaśniających do modelu ekonometrycznego*, Przegląd Statystyczny, z. 3-4/1981.

<sup>17</sup> Nie będziemy w tym miejscu prowadzić dokładnych rozważań w zakresie wszystkim problemów związanych z konstrukcją zmiennych syntetycznych, odsyłając czytelnika do licznych już pozycji literatury fachowej dostępnej w tym względzie.

<sup>18</sup> Pojęcia te przyjmujemy za Z. Hellwigiem. *Zastosowanie metody taksonomicznej do typologicznego podziału krajów ze względu na poziom ich rozwoju oraz zasoby i strukturę wykwalifikowanych kadr*, Przegląd Statystyczny, z. 4/1968 i T. Borysem, *Metody normowania cech w statystycznych badaniach porównawczych*, Przegląd Statystyczny, z. 2/1978.

- sposobu określania współrzędnych wzorca rozwoju,
- sposobu normowania zmiennych w celu doprowadzenia ich do porównywalności,
- charakteru wykorzystywanej postaci analitycznej funkcji agregującej,
- systemu wag nadawanych zmiennym pierwotnym.

Warunkiem ustalenia zmiennych syntetycznych jest na ogół doprowadzenie wszystkich wyjściowych cech do wzajemnej porównywalności. Nie wnikając głębiej w tę problematykę<sup>19</sup> trzeba stwierdzić, iż możliwe sposoby normowania zmiennych można sprowadzić do następujących operacji:

- rangowanie, polegające na zastąpieniu wyjściowych realizacji cech ich rangami wynikającymi z uporządkowania obserwacji (obiektów) zgodnie z rosnącymi lub malejącymi wartościami cechy,
- standaryzacja przebiegająca zgodnie z powszechnie znanymi regułami,
- unitaryzacja, przekształcająca wartości absolutne na relatywne, przy zachowaniu stałego i jednostkowego obszaru zmienności,
- przekształcenia ilorazowe, przyjmujące za układ odniesienia wartość minimalną, maksymalną bądź też dowolną wartość zmiennej ustaloną arbitralnie.

Dysponując macierzami unormowanych<sup>20</sup> w dowolny sposób wartości zmiennych optymalnych ustalonych w pierwszym kroku dwustopniowej metody pojemności informacji, które można (przedstawić w postaci następującego ciągu:

$$\{\mathbf{X}_1^{(u)}, \mathbf{X}_2^{(u)}, \dots, \mathbf{X}_P^{(u)}\}, \quad (10)$$

przystępuje się do agregowania cech, przy czym wykorzystać można w tym celu różnie określone funkcje addytywne lub moltiplicatywne<sup>21</sup>.

Rozważając problematykę zmiennych syntetycznych, zauważmy, iż istnieje możliwość stosowania przy ich konstrukcji systemu wag opartych na wartościach pojemności indywidualnych informacji  $h$  charakteryzujących walory informacyjne poszczególnych zmiennych cząstkowych. Nadawanie większej wagi tym zmiennym, które w większym stopniu

<sup>19</sup> Por. również pracę: T. Borys, *Metody normowania cech*.

<sup>20</sup> Trzeba w tym miejscu podkreślić, że w przypadku dysponowania szeregami przekrojowo-czasowymi można zapewnić porównywalność czasową wartości zmiennych syntetycznych, jeśli normowanie, prowadzone na przykład przy wykorzystaniu zasad standaryzacji, operować będzie' wartościami średnich i odchyłeń standardowych ustalonych w całym obszarze czasowym badań. W metodzie np. wzorca rozwoju jego współrzędne również muszą być ustalone wspólnie dla całego badanego okresu.

<sup>21</sup> Różne typy takich funkcji przedstawione są np. w pracy: T. Grabiński, S. Wydymus, A. Zeliaś, *Metody prognozowania*.



wyjaśniają kształtowanie się zmiennej endogenicznej, ma "bardzo istotne znaczenie z punktu widzenia przydatności praktycznej konstruowanych modeli syntetycznych. Jest to — zdaniem autora — bardzo ważna cecha dwustopniowej metody pojemności informacji.

Biorąc na przykład pod uwagę pierwszą podgrupę optymalnych zmiennych ustalonych w pierwszym kroku dwustopniowej metody (pojemności informacji (założmy, że liczba tych zmiennych jest równa  $L$ ), możemy na jej podstawie utworzyć zmienną syntetyczną według wzoru

$$S_i = \sum_{l=1}^L w_l x_{il}^{(u)} \quad (i=1, \dots, n), \quad (11)$$

przy czym wagi  $w_l$  zdefiniowane są następująco:

$$w_l = \frac{h_l}{H}, \quad (12)$$

natomiast  $h_l$  oznacza pojemność indywidualną nośnika informacji  $X_l$ , a  $H$  pojemność integralną całego podzbioru zmiennych:

$$H = \sum_{l=1}^L h_l. \quad (13)$$

Kontynuując ten tok postępowania dla każdego podzbioru optymalnych zmiennych, których wartości zawierają macierze tworzące ciąg (10), otrzymujemy w konsekwencji  $P$  zmiennych syntetycznych  $S_p$  ( $p=1, \dots, P$ ). Wartości tych zmiennych można przedstawić w postaci następującej macierzy:

$$S = \begin{bmatrix} S_{11} & S_{12} & \dots & S_{1P} \\ S_{21} & S_{22} & \dots & S_{2P} \\ \dots & \dots & \dots & \dots \\ S_{n1} & S_{n2} & \dots & S_{nP} \end{bmatrix}. \quad (14)$$

W następnym etapie analizy przeprowadza się wybór optymalnego podzbioru zmiennych syntetycznych do modelu syntetycznego poprzez zastosowanie drugiego kroku dwustopniowej metody pojemności informacji. Po obliczeniu wartości współczynników korelacji liniowej  $r_p$  między poszczególnymi zmiennymi  $S_p$  a zmienną endogeniczną oraz pomiędzy wszystkimi parami zmiennych  $S_p$  ustala się pojemności indywidualnych nośników informacji w ramach każdej kombinacji zmiennych<sup>22</sup>

$$h_p^* = \frac{r_p^2}{1 + \sum_{p \neq q} |r_{pq}|} \quad (p, q=1, \dots, P; p \neq q), \quad (15)$$

<sup>22</sup> Liczba tych kombinacji jest, jak wiadomo, równa  $2P - 1$ .

a na tej podstawie wartości pojemności integralnych

$$H_d^* = \sum_{p=1}^P h_p^* \quad (d=1, 2, \dots, 2^P-1). \quad (16)$$

Przedstawiona metoda postępowania pozwala na wybór optymalnej (w sensie przyjętego kryterium) kombinacji zmiennych syntetycznych charakteryzujących się maksymalną wartością parametru  $H$ . Zmienne te stanowią optymalny podzbiór zmiennych objaśniających w modelu syntetycznym.

Na zakończenie niniejszych rozważań zwróćmy uwagę, iż zakładaliśmy dobór zmiennych syntetycznych do modelu opisującego wahania określonej zmiennej endogenicznej  $Y$  (np. dochód narodowy poszczególnych krajów). W analizach porównawczych rozwoju obiektów ekonomicznych zmienną endogeniczną może również być określona zmienna syntetyczna  $S^*$ , która dotyczy przykładowo całego zespołu mierników rozwoju, czy to społecznego czy gospodarczego, lub też pewnych aspektów takiego rozwoju. Zauważmy, że przy założeniu dysponowania endogeniczną zmienną syntetyczną  $S^*$  nie ulega zmianie tok postępowania w dwustopniowej metodzie pojemności informacji.

### 3. MODELE SYNTETYCZNE W ANALIZIE ROZWOJU SPOŁECZNO-GOSPODARCZEGO OBIEKTÓW EKONOMICZNYCH

W dotychczasowym ekonometrycznym modelowaniu zmiennych syntetycznych charakteryzujących rozwój społeczno-gospodarczy poszczególnych obiektów ekonomicznych lub ich grup wykorzystywane są najczęściej modele tendencji rozwojowych<sup>23</sup> lub też modele jednorównaniowe, w których zmienna syntetyczna jest funkcją cząstkowych zmiennych diagnostycznych<sup>24</sup>. W pierwszym przypadku wynikiem wykorzystania modelu trendu jest prognoza ekstrapolacyjna, uwarunkowana szeregiem założeń wynikających z klasycznej teorii predykcji, której wartość (poznawcza jest jednakże znacznie ograniczona. W drugim przypadku zmienna syntetyczna jest funkcją swoich składowych, co stwarza określone trudności interpretacyjne. Ponadto wydaje się, iż wstępuje wtedy mieszanie pojęć (mierników i czynników wzrostu (lub też nakładów i wyników), co zmniejsza walory użytkowe takich modeli.

Proponowane modele syntetyczne pozbawione są powyższych niedogodności, a jednocześnie zbiór syntetycznych zmiennych objaśniających  $S_p$  charakteryzuje się — co wynika z rozważań przeprowadzonych wcześ-

<sup>23</sup> Por. np. pracę: T. Grabiński, S. Wydymus, A. Zeliaś, *Niektóre metody konstrukcji prognoz globalnych*, Folia Oeconomica Cracoviensia, vol. XXII, 1979.

<sup>24</sup> Por. np. pracę: R. Pluta, *Wielowymiarowa analiza porównawcza w badaniach ekonomicznych*, Warszawa 1977.

niej — dużym ładunkiem informacji o kształtowaniu się syntetycznej zmiennej endogenicznej  $S^*$ . Modele syntetyczne mogą znaleźć zastosowanie w analizie prawidłowości zachodzących między wielkością poziomu rozwoju społeczno-gospodarczego obiektów a grupami czynników determinujących ten rozwój, a zwłaszcza mogą być podstawą konstrukcji wariantowych prognoz tego rozwoju przy założeniu różnych poziomów czynników. Celem sporządzania prognoz na podstawie modeli syntetycznych jest określenie zarówno miejsca danego obiektu na mapie osiągniętego poziomu rozwoju, jak również zmian w hierarchicznym uszeregowaniu wszystkich obiektów lub ich grup.

W omawianej analizie znaleźć mogą zastosowanie różne typy modeli syntetycznych, poczynając od następującego, dotyczącego badania relacji w grupie  $k$  obiektów w określonym momencie<sup>25</sup>:

$$S_i^* = \alpha_0 + \alpha_1 S_{1i} + \dots + \alpha_p S_{pi} + u_i \quad (i=1, \dots, k), \quad (17)$$

przy czym  $u_i$  jest składnikiem losowym.

W przypadku gdy dysponujemy danymi przekrojowo-czasowymi, istnieje możliwość konstrukcji ciągu modeli typu (17) dla każdego okresu

$$\begin{aligned} S_{1i}^* &= \alpha_{01} + \alpha_{11} S_{11i} + \dots + \alpha_{p1} S_{1pi} + u_{1i} \\ &\dots \\ S_{ni}^* &= \alpha_{0n} + \alpha_{1n} S_{n1i} + \dots + \alpha_{pn} S_{npi} + u_{ni} \end{aligned} \quad (18)$$

lub też jednego modelu globalnego operującego danymi przekrojowo-czasowymi.

Analiza ciągu modeli (18), a zwłaszcza badanie ewentualnej tendencji rozwojowej poszczególnych wartości ocen parametrów, może dostarczyć szeregu interesujących informacji na temat prawidłowości rozwoju społeczno-gospodarczego analizowanych obiektów, a w szczególności może być podstawą ustalenia hipotetycznego modelu prognostycznego jako narzędzia wariantowych analiz predyktywnych.

Dysponowanie danymi przekrojowo-czasowymi, czyli w efekcie macierzą trójwymiarową wartości zmiennych syntetycznych

$$S = [S_{ipt}] \quad (i=1, \dots, k, p=1, \dots, P, t=1, \dots, n), \quad (19)$$

pozwala konstruować również modele wielorównaniowe proste, z których każdy dotyczy rozwoju społeczno-gospodarczego określonego obiektu<sup>26</sup>.

<sup>25</sup> Dla ustalenia uwagi załóżmy, iż mamy do czynienia z «modelami liniowymi.

<sup>26</sup> Zwróćmy uwagę, iż w przypadku każdego obiektu może występować taki sam lub różny zestaw objaśniających zmiennych syntetycznych. W modelach typu (20) mogą się znaleźć opóźnione zmienne syntetyczne lub też cząstkowe zmienne diagnostyczne, jeśli np. w wyniku zastosowania pierwszego kroku dwustopniowej metody pojemności informacji wyłoniony zostanie podzbiór jednoelementowy.

$$\begin{aligned}
 S_{1t}^* &= \beta_{10} + \beta_{11} S_{11t} + \dots + \beta_{1P} S_{1Pt} + e_{1t}, \\
 &\dots \dots \dots \dots \dots \dots \dots \dots \\
 S_{kt}^* &= \beta_{k0} + \beta_{k1} S_{k1t} + \dots + \beta_{kP} S_{kPt} + e_{kt}.
 \end{aligned}
 \tag{20}$$

Modele te mogą być podstawą budowy wariantowych prognoz rozwoju społeczno-gospodarczego obiektów ekonomicznych. Wartości zmiennych  $S_i$  w okresie prognozowanym mogą być przy tym w różny sposób ustalone. W szczególności mogą być one określone przez prowadzącego badanie, mogą wynikać z zastosowania np. modeli trendu do prognozowania wartości  $S_i$ , a także mogą być uzyskane poprzez zakładanie różnych poziomów zmiennych cząstkowych tworzących zmienną syntetyczną<sup>27</sup>.

Prowadząc badania na przykładzie krajów świata jako obiektów ekonomicznych, poszczególne zmienne syntetyczne mogą charakteryzować:

$S_t^*$  — grupę mierników poziomu rozwoju społecznego, gospodarczego lub społeczno-gospodarczego,

$S_{1t}$  — grupę cech dotyczących czynników demograficznych,

$S_{2t}$  — grupę cech dotyczących zatrudnienia w gospodarce narodowej,

$S_{3t}$  — grupę cech dotyczących majątku narodowego,

$S_{4t}$  — grupę cech dotyczących wielkości i struktury bazy surowcowej,

$S_{5t}$  — grupę cech dotyczących stopnia udziału danego kraju w międzynarodowym podziale pracy itp.

Zwróćmy uwagę, iż modele (20) o takim samym zbiorze syntetycznych zmiennych objaśniających mogą być również podstawą analiz porównawczych dotyczących podobieństwa systemu współzależności ilościowych w badanym układzie społeczno-gospodarczym<sup>28</sup>.

Analizę prawidłowości rozwoju społeczno-gospodarczego obiektów ekonomicznych można także prowadzić przy założeniu istnienia wielokierunkowych powiązań między kilkoma zmiennymi syntetycznymi  $S^*$  dotyczącymi różnych aspektów rozwoju społeczno-gospodarczego. W konsekwencji narzędziem analizy jest wielorównaniowy model syntetyczny, którego formę strukturalną można zapisać następująco<sup>29</sup>:

<sup>27</sup> W tym ostatnim przypadku wariantowe zakładanie a priori określonych poziomów zmiennych diagnostycznych musi poprzedzać w czasie tworzenie zmiennej syntetycznej, a ustalone warianty prognostyczne zespołu zmiennych  $X$  muszą być uwzględnione przy normowaniu zmiennych w całym obszarze czasowym badań lub np. przy ustaleniu współrzędnych wzorca rozwoju wspólnego, w tym przypadku dla całego okresu historycznego i prognozowanego.

<sup>28</sup> Zasady tego typu analizy przedstawione są w pracy: S. Wydymus, *Wielowymiarowa analiza porównawcza systemu ilościowych współzależności w gospodarce światowej*, Folia Oeconomica Cracoviensia vol. XXVII/1984.

<sup>29</sup> Dobór zmiennych syntetycznych do wielorównaniowego modelu syntetycznego można prowadzić przy różnych założeniach, transponując odpowiednio propozycje przedstawione np. w pracy: M. Gruszczyński, M. Kolupa, E. Leniewska, G. Napiórkowski, *Miary zgodności, metody doboru zmiennych, problemy współli-*

$$\mathbf{BS}^* + \mathbf{\Gamma S} = \boldsymbol{\varepsilon}, \quad (21)$$

przy czym  $\mathbf{S}^*$  jest wektorem zawierającym  $G$  zmiennych syntetycznych łącznie współzależnych,  $\mathbf{S}$  jest wektorem zawierającym  $P$  zmiennych z góry ustalonych,  $\mathbf{B}$  i  $\mathbf{\Gamma}$  są macierzami ocen parametrów, a  $\boldsymbol{\varepsilon}$  jest wektorem składników losowych poszczególnych równań.

Model (21) może dotyczyć zarówno danych przekrojowych, jak również czasowych (a więc każdego obiektu oddzielnie) lub też może być modelem, którego estymację przeprowadza się na podstawie danych przekrojowo-czasowych (obiektookresów).

W badaniach np. związanych z rozwojem społeczno-gospodarczym krajów świata logicznym wydaje się wprowadzenie założenia o istnieniu sprzężenia zwrotnego między rozwojem gospodarczym a rozwojem społecznym<sup>30</sup>, co implikuje następujący, prosty model o dwóch równaniach współzależnych:

$$\begin{aligned} S_1^* &= f(S_2^*, S_1, \dots, S_P, \varepsilon_1), \\ S_2^* &= f(S_1^*, S_1, \dots, S_P, \varepsilon_2), \end{aligned} \quad (22)$$

gdzie  $S_1^*$  — zmienna syntetyczna charakteryzująca grupę mierników dotyczących rozwoju gospodarczego,  $S_2^*$  — zmienna syntetyczna charakteryzująca grupę mierników dotyczących rozwoju społecznego,  $S_p$  ( $p = 1, \dots, P$ ) — zmienne syntetyczne dotyczące różnych grup czynników wzrostu społeczno-gospodarczego.

Przedstawione powyżej przykłady ekonometrycznych modeli syntetycznych nie wyczerpują oczywiście wszystkich problemów, jakie się w tym względzie pojawiają. Są one jedynie wstępem do dalszych studiów nad ekonometrycznym modelowaniem złożonej problematyki rozwoju społeczno-gospodarczego obiektów ekonomicznych, zwłaszcza krajów świata lub regionów. Modele te są pewną próbą rozwiązania zagadnienia bardzo dużej liczby informacji, jakie powinien badacz uwzględnić, gdy chce przeprowadzić wszechstronną, a jednocześnie syntetyczną analizę tego problemu.

Modele takie mogą znaleźć szerokie zastosowanie zwłaszcza w analizie porównawczej i wariantowym prognozowaniu rozwoju społeczno-gospodarczego różnego rodzaju obiektów ekonomicznych. Aczkolwiek rozważania prowadzone były głównie na przykładzie gospodarki krajów świata, to jednak wszystkie propozycje zachowują swoje znaczenie w odniesieniu do innych obiektów ekonomicznych, jak np. regionów, przedsiębiorstw określonej branży lub nawet stanowisk pracy czy też wyrobów, jeśli badaniem objęte są na przykład problemy kształtowa-

*niowości*, Warszawa 1979; por. również w tej sprawie uwagi zawarte w pracy T. Grabiński, S. Wydymus, A. Zeliaś, *Metody doboru zmiennych*.

<sup>30</sup> Por. np. pracę O. Lange, *Ekonomia polityczna*.

nia wydajności pracy lub też zagadnienia kosztów. Szeroka gama zagadnień, w których mogą znaleźć zastosowanie modele ze zmiennymi syntetycznymi w połączeniu z nową propozycją zasad tworzenia takich zmiennych, daje podstawy do przypuszczeń, iż modele te powinny stanowić pewien krok na drodze rozwoju metod modelowania ekonometrycznego, a także zwiększenia stopnia ich praktycznej użyteczności.

## SYNTHETIC VARIABLES IN ECONOMETRIC MODELS

### Summary

The present article is an attempt to construct econometric models with synthetic variables, created on the grounds of multidimensional comparative analysis. The author discusses a problem of creating these variables having time-sectional data provided, and next presents a proposition of defining an optimal sub-set of synthetic describing variables utilizing the so-called two-stage method of capacity of information. Information facilitating construction of new weighted synthetic variable is also provided by this method, weights being measures of stock of information of fractional diagnostic variables. The article is concluded with a review of various types of model propositions which can be applied in comparative analysis and in forecasting of socioeconomic development of all analysed economic objects or their groups.