

ALEKSANDRA WITKOWSKA, MAREK WITKOWSKI

METODA GŁÓWNYCH SKŁADOWYCH  
JAKO METODA IDENTYFIKACJI CZYNNIKÓW  
OKREŚLAJĄCYCH POZIOM KOSZTÓW  
PRZEDSIĘBIORSTW HANDLU DETALICZNEGO

I. UWAGI WSTĘPNE

Praca stanowi przyczynek do zagadnienia statystycznej identyfikacji czynników określających poziom kosztów w przedsiębiorstwie handlu detalicznego. W rozumieniu powyższego opracowania pojęcie identyfikacji obejmuje numeryczne określenie danego czynnika oraz siły jego oddziaływania na badane zjawisko. Interesować nas zatem będzie możliwość wydzielenia z wielu apriorycznie ustalonych zmiennych jedynie takich, których wpływ na zmienność kosztów jest istotny tak ze statystycznego, jak i ekonomicznego punktu widzenia. Prawidłowe rozwiązanie zasygnalizowanego problemu jest zadaniem złożonym. Wszystkie podziały, ustalenia i klasyfikacje mają bowiem, jak dotychczas, charakter hipotetyczny i intuicyjny. Wpływ na to posiada z jednej strony charakter wielu determinant (są one z natury niemierzalne), z drugiej zaś wielokierunkowość ich oddziaływania, w wyniku czego jeden z czynników wpływa na drugi, ten z kolei oddziałuje na poziom kosztów, a także na czynnik go determinujący itd. Tworzy się więc system, w którym wpływ jednego czynnika na koszty wiąże się immanentnie z oddziaływaniem pozostałych. Utrudnia to oczywiście postępowanie badawcze, uniemożliwiające precyzyjniejsze ustawienie każdego z czynników na należnym mu miejscu w hierarchii ważności.

Nasze zainteresowanie poruszaną problematyką związane jest z niedorozwojem badań ilościowych w zakresie rozpoznania podstawowych prawidłowości ekonomicznych odnośnie do kształtowania się kosztów przedsiębiorstwa handlowego względem różnorodnych czynników wytwórczych, co powoduje, że nie mogą one być wykorzystane przy podejmowaniu decyzji zmierzających do podwyższenia poziomu świadczonych konsumenten-

tom usług handlowych. Brak rozpoznania wspomnianych prawidłowości dla tak wysoce zagregowanej jednostki statystycznej, jaką jest przedsiębiorstwo, powoduje, że wszelkie dociekania związane z opracowaniem metod analizy kosztów nie mogą się rozwijać z uwagi na hipotetyczność przyjmowanych założeń. Utrudnia to w konsekwencji prowadzenie prawidłowej polityki ekonomicznej w odniesieniu do cen detalicznych i marż handlowych oraz wdrażanie zasad rachunku ekonomicznego w przedsiębiorstwach handlowych.

Podjęta analiza jest więc próbą empirycznego wykrycia zjawisk odgrywających decydującą rolę w wyjaśnianiu mechanizmu kosztów przedsiębiorstwa handlu detalicznego. Jest ona również w pewnym sensie przełamaniem dotychczasowej tendencji do li tylko werbalnego określania wpływu różnych czynników na koszty handlowe.

Egzemplifikacja prezentowanej w opracowaniu metody badawczej dokonana zostanie na przykładzie zbiorowości 57 przedsiębiorstw handlu detalicznego artykułami spożywczymi, działającymi w środowisku miejskim na terenie całego kraju w 1967 r. Wybór tej zbiorowości podyktowany został przede wszystkim względami poznawczymi. Chodzi mianowicie o to, że branża spożywcza charakteryzuje bardzo znaczna kosztochłonność realizacji obrotów połączona ze szczególnie wysokim udziałem czynnika pracy w świadczeniu usług handlowych. Fakt ten powoduje siłą rzeczy stosunkowo dużą wrażliwość kosztów handlowych na zmiany tego czynnika. W tym kontekście interesujące wydaje się określenie czy i jaką rolę, w warunkach silnej presji czynnika pracy na koszty, odgrywają zmienne reprezentujące inne czynniki wytwórcze.

Podjęty temat reprezentuje wycinek długofalowych badań z zakresu analizy kosztów handlu prowadzonych przez Zakład Statystyki i Demografii Akademii Ekonomicznej w Poznaniu. Podstawę źródłową stanowią informacje zawarte w banku kosztów handlu uspołecznionego tegoż Zakładu. Z uwagi na specyfikę przedmiotu badania stosunkowo dużo miejsca poświęcono stronie metodologicznej zagadnienia, przedstawione bowiem propozycje i uzyskane doświadczenia mogą okazać się przydatne w realizacji dalszych prac z tej dziedziny.

## II. CZYNNIKI KSZTAŁTUJĄCE POZIOM KOSZTÓW HANDLOWYCH

Poziom kosztów handlowych może różnić się tak w odniesieniu do poszczególnych podmiotów gospodarujących, jak i w odniesieniu do różnych przyjętych do badania jednostek czasu. Jednakże w każdym zjawisku masowym można wyodrębnić określone prawidłowości, pozwalające wyjaśnić mechanizm kształtowania się tego zjawiska w zależności od czynników go determinujących.

Z teoretycznego punktu widzenia lista potencjalnych czynników kosz-

totwórczych jest bardzo znaczna<sup>1</sup>. Identyfikacja wszystkich w drodze analizy statystycznej jest praktycznie niemożliwa, gdyż z jednej strony wiele z nich jest ilościowo niewymiernych (nieskwantyfikowanych), a z drugiej — barierą ograniczającą uwzględnienie innych jest oddany do użytku podmiotu badającego materiał empiryczny dotyczący kosztów i efektów działalności wybranej zbiorowości przedsiębiorstw. Biorąc te fakty pod uwagę wybrano jako potencjalne zmienne objaśniające<sup>2</sup>:

1) wartość obrotu towarowego  $X_1$ ; 2) liczba punktów sprzedaży detalicznej  $X_2$ ; 3) liczba sklepów  $X_3$ ; 4) liczba pracowników ogółem  $X_4$ ; 5) liczba pracowników operatywnych  $X_5$ ; 6) powierzchnia lokali sklepowych  $X_6$ ; 7) powierzchnia sal sprzedażowych  $X_7$ ; 8) wartość zapasu końcowego  $X_8$ .

Uzasadnienie wyboru tych, a nie innych czynników oparte było także na ogólnych przesłankach ekonomicznych. Rozwój przedsiębiorstwa handlowego jest, jak wiadomo, uwarunkowany z jednej strony rozmiarami świadczonych usług handlowych oraz z drugiej — przyrostem zaangażowanych czynników wytwórczych. Przyjęcie wartości obrotu towarowego uznano za wskazane, ponieważ w praktyce jest on uważany powszechnie za symptom rozmiaru usługi handlowej i podstawowy efekt działalności przedsiębiorstwa handlu detalicznego. Jest to również wielkość będąca podstawowym elementem programowania w przedsiębiorstwie handlowym. Równie niezbędne było wprowadzenie liczby zatrudnionych, jako że udział czynnika pracy w wytwarzaniu usług handlowych, podobnie jak usług w ogóle, jest bardzo duży, znacznie większy niż w innych dziedzinach działalności gospodarczej. Spostrzeżenie to jest tym bardziej uzasadnione, jeśli zważy się, że przedmiotem badania są przedsiębiorstwa wchodzące w skład branż spożywczych. W branżach tych koszty związane z wykorzystaniem czynnika pracy są szczególnie wysokie, wahają się bowiem w granicach od 35 do 40% kosztów ogólnych<sup>3</sup>. Czynnikiem pracy stanowi też podstawową determinantę określającą rozmiary działalności handlowej. Uwzględnienie powierzchni (tak lokali sklepowych, jak i sal sprzedażowych) podyktowane było z kolei chęcią określenia efektu substytuowania pracy innymi czynnikami produkcji. Wychodzono z założenia, że wpływ postępu technicznego wyraża się właśnie zmianą w stru-

<sup>1</sup> Zainteresowanych odsyłamy do literatury przedmiotu. Por. m. in. R. Peretiatkowicz, *Ekonomika i organizacja przedsiębiorstwa handlowego*, Warszawa 1971 oraz Z. Zakrzewski, *Ekonomika handlu wewnętrznego*, Warszawa 1970.

<sup>2</sup> Wybór wyjściowego zespołu zmiennych jest szczególnie trudny i ograniczony w przypadku badań o charakterze retrospektywnym cechujących się wysokim stopniem agregacji badania.

<sup>3</sup> *Handel Wewnętrzny 1976*, Statystyka Polski nr 87, Warszawa 1977, tab. 4 (53), s. 84; S. Wierzchosławski, *Zmiany w strukturze kosztów przedsiębiorstw handlu detalicznego w okresie powojennym*, Ruch Prawniczy, Ekonomiczny i Socjologiczny nr 4/1965, s. 162 - 165.

kturze kosztów, idącą w kierunku wzrostu nakładów o charakterze rzeczowym. Rozmiary powierzchni wydają się podstawowym elementem charakteryzującym bazę materialno-techniczną. Z punktu widzenia substytucyjności pracy żywej pracą uprzedmiotowioną korzystniejsze byłoby przyjęcie jako zmiennej objaśniającej wartość środków trwałych. Niestety brak danych statystycznych uniemożliwił takie postępowanie. Stąd też powierzchnia, szczególnie sal sprzedażowych, musi stanowić niejako synonim majątku trwałego. Liczba punktów sprzedaży detalicznej określa natomiast wielkość sieci detalicznej, kształtując równocześnie zatrudnienie, a co za tym idzie także potencjalną zdolność sieci do realizowania określonych usług handlowych. Ostatnim elementem przyjętym w badaniu jest wartość zapasu końcowego. Uwzględniono ją wychodząc z założenia, że determinanta ta reprezentuje w gruncie rzeczy czynnik kapitałowy, a jej wielkość decyduje z jednej strony o tym w jakim stopniu zapewniona jest ciągłość sprzedaży, z drugiej natomiast wpływa na rozmiary zużycia czynników wytwórczych, a więc w konsekwencji i na poziom kosztów.

### III. PRÓBA ZASTOSOWANIA KLASYCZNEJ REGRESJI LINIOWEJ

Najbardziej oczywistym sposobem określenia oddziaływania wybranych czynników na koszty jest wykorzystanie modelu regresji liniowej ze wszystkimi zmiennymi zbioru wyjściowego. Celem takiego postępowania jest wstępne rozpoznanie siły oddziaływania przyjętych determinant na poziom kosztów handlowych. Model ma postać następującą:

$$\hat{Y} = a_0 + \sum_{i=1}^8 a_i X_i,$$

gdzie:  $\hat{Y}$  — zmienna objaśniana (tj. koszty handlowe),  $X_i$  — zmienne objaśniające ( $i = 1, 2, \dots, 8$ ),  $a_i$  — współczynniki regresji cząstkowej określające „czysty” wpływ zmiennej  $X_i$  na koszty handlowe.

Estymacji parametrów przyjętego modelu regresji, dla wybranej zbiorowości przedsiębiorstw handlu detalicznego, dokonano stosując klasyczną metodę najmniejszych kwadratów (KMNiK). Otrzymane oszacowania przedstawia tabela 1<sup>4</sup>.

Wśród otrzymanych rezultatów uwagę zwracają bardzo duże wartości liczbowe błędów średnich szacunku większości oszacowanych parametrów strukturalnych, a co za tym idzie niskie wartości statystyki  $t$ -Studenta. Na podstawie tylko tego modelu można by więc sądzić, że statystycznie istotnymi, z punktu widzenia oddziaływania na koszty, są jedynie: liczba

<sup>4</sup> Wszystkie obliczenia prezentowane w opracowaniu przeprowadzono w Uczelnianym Ośrodku Przetwarzania Informacji przy Akademii Ekonomicznej w Poznaniu.

Tabela 1

Parametry opisowe regresyjnego modelu kosztów handlowych względem czynników je określających

Parametr $a_i$	Ocena parametru $a_i$	Błąd średni parametru $s(a_i)$	Wartość statysty- ki $t$ -Studenta
$a_0$	-16,001	—	—
$a_1$	0,006	0,005	1,11
$a_2$	10,501	16,539	0,63
$a_3$	-22,893	22,331	1,03
$a_4$	63,905	20,737	3,08
$a_5$	-18,932	21,578	0,82
$a_6$	0,181	0,425	0,43
$a_7$	0,623	0,856	0,73
$a_8$	0,148	0,059	2,51
Parametry struktury stochastycznej modelu		$R=0,995$ ${}^{48}t_{0,05}=2,01$	$R^2=0,990$ $S_y=15,033$

Tabela 2

Współczynniki  $\beta$  regresyjnego modelu kosztów handlowych względem czynników je określających

Parametr $a_i$	Odchylenie standardowe zmienniej $X_i$	Współczynniki $\beta_i$
$a_1$	2245,648	0,826
$a_2$	0,548	0,382
$a_3$	0,490	-0,745
$a_4$	2,171	9,227
$a_5$	1,954	-2,416
$a_6$	34,674	0,418
$a_7$	21,258	0,882
$a_8$	82,491	0,814

pracowników ogółem i wartość zapasu końcowego. Stwierdzenie tego faktu nie jest jednak tak oczywiste, jeżeli dokonamy analizy ważności poszczególnych czynników ze względu na wpływ jaki one mają na wartość zmiennej zależnej  $Y$ . Obiektywnej miary w tym zakresie dostarczają współczynniki  $\beta$  zdefiniowane następująco<sup>5</sup>:

$$\beta_i = a_i \cdot \frac{\sigma_i}{S_y},$$

gdzie:  $a_i$  — współczynnik regresji przy zmiennej  $X_i$ ,  $\sigma_i$  — odchylenie standardowe zmiennej  $X_i$ ,  $S_y$  — błąd standardowy modelu regresji.

Obliczone współczynniki  $\beta$  zamieszczone w tabeli 2 wskazują, że

<sup>5</sup> A. Goldberger, *Teoria ekonometrii*, Warszawa 1976, s. 259 - 260.

Tabela 3

Macierz współczynników korelacji między zmiennymi objaśniającymi modelu regresji liniowej

$X_i$	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$X_8$
$X_1$	1,000	0,888	0,898	0,975	0,971	0,946	0,956	0,862
$X_2$	0,888	1,000	0,971	0,842	0,834	0,879	0,893	0,685
$X_3$	0,898	0,971	1,000	0,865	0,858	0,914	0,927	0,696
$X_4$	0,975	0,842	0,865	1,000	0,998	0,955	0,964	0,894
$X_5$	0,971	0,834	0,858	0,998	1,000	0,954	0,963	0,896
$X_6$	0,946	0,879	0,914	0,955	0,954	1,000	0,990	0,838
$X_7$	0,956	0,893	0,927	0,964	0,963	0,990	1,000	0,835
$X_8$	0,862	0,685	0,696	0,894	0,896	0,838	0,835	1,000

wprawdzie najważniejszą zmienną jest liczba pracowników ogółem, ale już znaczenie wartości zapasu końcowego jest mniejsze niż wartości obrotu towarowego i powierzchni sal sprzedażowych, które były nieistotne w. sensie statystyki t-Studenta.

Ogólnie rzecz biorąc przedstawiony model wskazał niskie prawdopodobieństwa zachodzenia pewnych związków o charakterze przyczynowo-skutkowym, co z ekonomicznego punktu widzenia budzić musi duże wątpliwości i zastrzeżenia. Na tym tle powstaje więc pytanie, co jest przyczyną takiego stanu rzeczy; trudno oczywiście na to jednoznacznie odpowiedzieć. Bardziej wnikliwa analiza przedstawionego modelu, przeprowadzona z punktu widzenia teorii estymacji, wskazuje, że nie jest on odpowiedni do badania omawianego zagadnienia. Okazuje się mianowicie, że bezpośrednią przyczyną niskiego poziomu istotności przedstawionych estymatorów jest istnienie bardzo silnego wzajemnego skorelowania wszystkich przyjętych w badaniu zmiennych objaśniających to jest występowanie zjawiska współliniowości zmiennych objaśniających (tab. 3).

#### IV. OGRANICZENIE WPLYWU WSPÓLLINIOWOŚCI ZMIENNYCH

Bardzo wysokie wartości liczbowe współczynników korelacji między parami zmiennych wskazują, że uzyskano model zmiennych współliniowych. Stosowanie w tej sytuacji klasycznej metody estymacji (KMNK) nie jest w pełni uzasadnione. Okazuje się bowiem, że parametry stałe modelu odzwierciedlają nie tylko siłę oddziaływania na zmienną objaśnianą tych zmiennych, do których się odnoszą, lecz również pozostałych czynników-argumentów. Komplikuje to oczywiście proces wydzielania najbardziej istotnych czynników kosztotwórczych<sup>6</sup>. Ponadto obniża się wartość poznawcza współczynników regresji przy próbie ich interpretacji

<sup>6</sup> G. Kildiszew, A. Frenkel, *Analiza szeregów czasowych i prognozowanie*, Warszawa 1977, s. 100.

ekonomicznej. Dzieje się tak dlatego, że otrzymane oszacowania parametrów, przy różnym zestawie zmiennych, mogą sugerować różną siłę oddziaływania danego czynnika na poziom badanego zjawiska ekonomicznego. Po trzecie wreszcie pojawiają się trudności natury numerycznej związane ze złym uwarunkowaniem macierzy obserwacji dokonanych na zmiennych objaśniających. Następuje mianowicie zmniejszenie jej określoności (tj. zbliżenie się wyznacznika tej macierzy do zera), co powoduje w konsekwencji, że błędy średnie szacunku parametrów są bardzo duże, przekraczające nawet wartości bezwzględne samych oszacowań. Takie wyniki numeryczne są więc w gruncie rzeczy niezbyt, użyteczne od strony poznawczej<sup>7</sup>. Istnieje bowiem tendencja do nieuzasadnionego wykluczania niektórych istotnych zmiennych z modelu regresji.

Poczynione uwagi wskazują, że współliniowość jest zjawiskiem, którego nie można lekceważyć, jeśli chce się w sposób możliwie dokładny określić przyczynowo-skutkowe uwarunkowania kosztów i czynników je determinujących. Postanowiono więc podjąć próbę ograniczenia jej ujemnych dla procesu estymacji skutków. Najbardziej oczywistym sposobem zmniejszenia efektu współliniowości wydaje się być tak zwana selekcja zmiennych. Polega ona na usunięciu ze zbioru zmiennych objaśniających, zmiennych najsilniej skorelowanych. Do tego celu można zastosować statystyczny test istotności, zaproponowany przez R. R. Glaubera i D. E. Farrara<sup>8</sup>. Test ten służy do weryfikacji hipotezy zerowej  $H_0: R_i^2 = 0$  wobec hipotezy alternatywnej  $H_1: R_i^2 > 0$ , a jako kryterium oceny stopnia współliniowości badanych zmiennych wprowadza się statystykę

$$F_i = (r_{ii} - 1) \cdot \frac{n-k}{k-1},$$

gdzie  $R_i$  — jest współczynnikiem korelacji wielokrotnej między zmienną „i” objaśniającą a pozostałymi  $k-1$  zmiennymi objaśniającymi (tj. bez zmiennej o numerze i),  $r_{ii}$  —  $i$ -ty element diagonalny macierzy odwrotnej współczynników korelacji między zmiennymi objaśniającymi.

Statystyka ta przy założeniu prawdziwości  $H_0$  ma rozkład  $F$ -Fischera o  $n-k$  i  $k-1$  stopniach swobody. Zmienna o numerze „i” podlega wpływowi współliniowości, jeśli przy danym poziomie istotności  $\alpha$   $F_i \geq F_{\alpha}$ . Postępowanie takie w konsekwencji prowadzi do wyodrębnienia takiego zespołu zmiennych, w stosunku do którego można twierdzić, że charakteryzuje się współliniowością. Rezultaty zastosowania testu  $F$  zawarto w tabeli 4. Przytoczone parametry wskazują jednoznacznie, że zakresem współliniowości objęte są wszystkie uwzględnione w badaniu czynniki-

<sup>7</sup> M. Gruszczyński, M. Kolupa, E. Leniewska, G. Napiórkowski, *Miary zgodności, metody doboru zmiennych, problemy współliniowości*, Warszawa 1979, s. 116.

<sup>8</sup> D. E. Farrar, R. R. Glauber, *Multicollinearity in Regression Analysis: The Problem Revisited*, w: *Review of Economics and Statistics*, 1967/49, s. 92 - 107.

-argumenty. Łatwo zauważyć, że zmiennymi o najsilniejszej współliniowości okazały się generalnie zmienne reprezentujące czynnik pracy, a najsłabiej zmienna reprezentująca czynnik kapitałowy, to jest wartość zapasu końcowego. Selekcja zmiennych w tym przypadku nie może więc dać siłą rzeczy pozytywnych rezultatów. W ogóle technikę tę należy stosować z dużą ostrożnością, gdyż zubaża ona zbiór informacji służących, do wyjaśniania zmiennej objaśnianej. Z reguły jest to postępowanie zawodne.

Tabela 4

Określenie zakresu współliniowości na zbiorze zmiennych objaśniających (test  $F$ )

Zmienne objaśniające $X_i$	Wartość sprawdzianu hipotezy $F_i$
Wartość obrotu towarowego $X_1$	211,4
Liczba punktów sprzedaży detalicznej $X_2$	135,8
Liczba sklepów $X_3$	200,2
Liczba pracowników	
Ogółem $X_4$	3507,0
Liczba pracowników operatywnych $X_5$	3080,0
Powierzchnia lokali sklepowych $X_6$	369,6
Powierzchnia sal sprzedażowych $X_7$	567,0
Wartość zapasu końcowego $X_8$	34,3
Wartość krytyczna sprawdzianu hipotezy ${}_{7}^{49}F=3,32$ $\alpha=0,05$	

W praktyce bowiem zdarza się, że nieistotność wyeliminowanych czynników jest wynikiem wysokiego udziału skorelowania ich z innymi. Stąd w wyniku selekcji można usunąć pewne zmienne o dużym, choć pośrednim wpływie na zmienną zależną. W pracy zdecydowano się na inne podejście polegające na wykorzystaniu metod ortogonalizujących zmienne objaśniające. Wśród metod tych największego znaczenia nabiera analiza czynnikowa, a szczególnie jeden z jej wariantów zwany analizą głównych składowych.

Metoda czynnikowa obejmuje grupę metod matematyczno-statystycznych pozwalających na sprowadzenie pierwotnego zbioru zmiennych, które charakteryzują jednostki statystyczne poddane badaniu, do znacznie mniejszej liczby ukrytych hipotetycznych zmiennych zwanych czynnikami. Te nowe zmienne zawierają podstawową informację o zmiennych oryginalnych i mogą być traktowane jako przyczyny zaobserwowanej zmienności (wariancji). Analiza czynnikowa jest liniowym modelem matematycznym i przyjmuje się w niej dwa założenia dotyczące zmiennych i czynników, a mianowicie: 1) zmienne pierwotne i czynniki powinny być znormalizowane, 2) wszystkie czynniki powinny być nieskorelowane ze

sobą. Biorąc te założenia pod uwagę, każdą ze zmiennych wyjściowych można wyrazić jako kombinację liniową hipotetycznych czynników<sup>9</sup>:

$$X_i = a_{i1} Q_1 + a_{i2} Q_2 + \dots + a_{ir} Q_r + \gamma_i F_i,$$

gdzie:  $Q_1, \dots, Q_r$  — są to tak zwane czynniki wspólne,  $F_i$  — czynnik swoisty, właściwy dla zmiennej o numerze  $i$ ,  $a_{ir}$  — ładunki czynnikowe określające wagę danego czynnika wspólnego w opisie zmiennych oryginalnych,  $\gamma_i$  — ładunek czynnikowy czynnika swoistego  $F_i$ ,  $i=1, 2, \dots, k$ , przy czym z założenia  $r < k$ .

W metodzie głównych składowych zakłada się natomiast, że poszczególne cechy nie zawierają czynników specyficznych, co jest do przyjęcia w przypadku występowania wysokich korelacji między nimi<sup>10</sup>. Uzasadnia to więc możliwość zastosowania tej metody w badaniu interesującego nas problemu kosztów handlowych. W metodzie głównych składowych (w skrócie MGS) wykrywa się czynniki w określonym porządku, stosownie do ich udziału w całkowitym zasobie zmienności wspólnej. Analizę rozpoczyna się od czynnika  $Q_{11}$ , którego udział w ogólnej zmienności danego zbioru cech jest największy<sup>11</sup>. Uwzględniając powyższe, problem sprowadza się, formalnie rzecz biorąc, do znalezienia takich dwóch wektorów  $q_1$  i  $a_1$  o wymiarach odpowiednio  $n \times 1$  i  $k \times 1$ , aby macierz  $q_1 a_1'$  była najlepszą aproksymacją macierzy zmiennych objaśniających w postaci znormalizowanej  $z$  o wymiarach  $n \times k$ .

W konsekwencji można więc zapisać, że:

$$q_1 = z a_1,$$

gdzie:

$$q_1 = \begin{bmatrix} q_{11} \\ q_{21} \\ \vdots \\ q_{n1} \end{bmatrix} \quad z = \begin{bmatrix} z_{11} & z_{12} & \dots & z_{1k} \\ z_{21} & z_{22} & \dots & z_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ z_{n1} & z_{n2} & \dots & z_{nk} \end{bmatrix} \quad a_1 = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{k1} \end{bmatrix}$$

oraz, że

$$q_1' q_1 = a_1' (z' z) a_1.$$

Składowe wektora  $a_1$  są wartościami nieznanymi i oblicza się je maksymalizując wyrażenie

$$q_1' q_1 = a_1' (z' z) a_1 \quad (1)$$

<sup>9</sup> Z. Chojnicki, T. Czyż, *Zmiany struktury regionalnej Polski w świetle przepływów towarowych w latach 1958 - 1966*, Warszawa 1972, s. 11 - 13.

<sup>10</sup> B. Podolec, K. Zając, *Ekonometryczne metody ustalania rejonów konsumpcji*, Warszawa 1978, s. 30.

<sup>11</sup> A. Zeliaś, *Badania współliniowości w modelach ekonometrycznych*, Przegląd Statystyczny 1977/2, s. 218-225; Z. Hellwig, *Zastosowanie przekształcenia ortogonalnego do wyznaczania dopuszczalnych wartości zmiennych objaśniających w modelu ekonometrycznym*, Przegląd Statystyczny 1974/3, s. 307 - 315.

przy założeniu, że  $\mathbf{a}'_1 \mathbf{a}_1 = 1$ . Prowadzi to w efekcie do poszukiwania ekstremum (max) funkcji określonej wzorem:

$$\psi = \mathbf{a}'_1 (\mathbf{z}' \mathbf{z}) \mathbf{a}_1 - \lambda_1 (\mathbf{a}'_1 \mathbf{a}_1 - 1)$$

gdzie  $\lambda_i$  jest tak zwanym mnożnikiem Lagrange'a. W wyniku tego ostatecznie otrzymuje się:

$$(\mathbf{z}' \mathbf{z}) \mathbf{a}_1 = \lambda_1 \mathbf{a}_1 \quad (2)$$

przy czym  $\mathbf{a}_1$  jest wektorem własnym macierzy  $(\mathbf{z}' \mathbf{z})$ , któremu zawsze odpowiada pewna liczba  $\lambda_1 > 0$  zwana wartością własną lub pierwiastkiem charakterystycznym macierzy  $(\mathbf{z}' \mathbf{z})$ . Uwzględniając z kolei w (1), że  $(\mathbf{z}' \mathbf{z}) \mathbf{a}_1 = \lambda_1 \mathbf{a}_1$  otrzymuje się, iż  $\mathbf{q}'_1 \mathbf{q}_1 = \lambda_1$ . Widać więc, że  $\lambda_1$  dobiera się jako największy z pierwiastków charakterystycznych macierzy  $(\mathbf{z}' \mathbf{z})$ .

Następnie wyznacza się drugi czynnik  $Q_{i2}$ , ortogonalny względem pierwszego według tej samej zasady, z maksymalnym udziałem w nowo znalezionej zmienności wspólnej. Oblicza się więc  $q_2 = z a_2$  i  $\mathbf{q}'_2 \mathbf{q}_2 = \mathbf{a}'_2 (\mathbf{z}' \mathbf{z}) \mathbf{a}_2$  zakładając, że  $\mathbf{a}'_2 \mathbf{a}_2 = 1$  oraz  $\mathbf{a}'_1 \mathbf{a}_2 = \mathbf{a}'_2 \mathbf{a}_1 = 0$ .

Drugi warunek wynika z faktu, że wektory  $q_2$  i  $q_1$  z definicji powinny być ortogonalne względem siebie. Postępując analogicznie jak uprzednio znajduje się, że

$$\mathbf{q}'_2 \mathbf{q}_2 = \lambda_2 \mathbf{a}'_2 \mathbf{a}_2 = \lambda_2.$$

Zauważyć wypada, że teraz  $\lambda_2$  dobieramy jako drugi z największych pierwiastków charakterystycznych macierzy  $(\mathbf{z}' \mathbf{z})$  co oznacza, iż  $\lambda_2 < \lambda_1$ . Proces ten można kontynuować aż do całkowitego wyczerpania zasobu zmienności wspólnej wszystkich badanych zmiennych. W rezultacie otrzymuje się macierz ortogonalną  $A = [a_1, a_2, \dots, a_k]$ , której kolumny reprezentują unormowane wektory macierzy  $(\mathbf{z}' \mathbf{z})$  i spełniają warunki  $\mathbf{a}'_i \mathbf{a}_j = \delta_{ij}$  gdzie:

$$\delta_{ij} = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases} \quad (i, j = 1, 2, \dots, k)$$

oraz macierz  $Q$  o wymiarach  $n \times k$ , złożoną z wartości poszczególnych nowych cech  $Q_{i1}, Q_{i2}, \dots, Q_{ik}$  obliczając  $Q = Z \times A$ .

Wyznaczone w ten sposób zmienne określa się mianem głównych składowych, a ich dyspersję określają wielkości  $\lambda_i = \mathbf{q}'_i \mathbf{q}_i$  ( $i = 1, 2, \dots, k$ ).

W ten sposób wyrażenia  $p_i = \frac{\lambda_i}{\sum \lambda_i}$  reprezentują udział  $i$ -tego komponenta w całkowitej zmienności zmiennych oryginalnych<sup>12</sup>. Stosunki te wska-

<sup>12</sup> Oznacza to więc, że w ogólnym przypadku głównych składowych jest tyle co zmiennych objaśniających a więc  $k$ . W praktyce jednak analizę przeprowadza się do chwili przebadania od 95 do 98% zasobu zmienności wspólnej wszystkich wyjściowych zmiennych, w wyniku czego czynników  $Q_i$  jest z reguły mniej niż  $k$ , np.  $r$ .

Tabela 5

Parametry opisowe modelu kosztów handlowych z uwzględnieniem głównych składowych

Parametr $\mu_i$	Ocena parametru	Błąd średni parametru	Wartość statystyki $t$ -Studenta
$\mu_0$	263,88	—	—
$\mu_1$	50,14	1,11	45,15
$\mu_2$	29,62	4,47	6,63
$\mu_3$	12,69	8,78	1,44
Parametry struktury stochastycznej modelu		$R^2=0,9761$	$S_e=22,58$
		${}^{53}t_{0,05}=2,01$	

zują więc jaka jest relatywna wartość poznawcza poszczególnych zmiennych  $Q_{ib}$ , a więc także ich wartość jako nośników informacji. Te nowe zmienne można z kolei interpretować jako przyczyny odpowiedzialne za zachowanie się zmiennej objaśnianej  $Y$  i traktować jako zmienne objaśniające. Takie postępowanie daje w konsekwencji podstawę do estymacji parametrów nowego modelu kosztów handlowych w postaci<sup>13</sup>:

$$Y = \mu_0 + \mu_1 Q_1 + \dots + \mu_r Q_r,$$

gdzie:  $Q_i$  — zmienne reprezentujące główne składowe ( $i=1, 2, \dots, r$ ),  $\mu_i$  parametry stałe,  $Y$  — koszty handlowe, prowadząc tym samym do całkowitej eliminacji współliniowości w trakcie estymacji modelu kosztów, a równocześnie dając możliwość dodatkowego wyeliminowania tych głównych składowych, które okażą się nieistotne dla opisu zmienności kosztów<sup>14</sup>.

Zastosowanie MGS w badaniu interesującego nas problemu doprowadziło do wyróżnienia trzech głównych składowych, których wariancje są równe odpowiednio:

**składowa  $Q_1$  8,2452**

**składowa  $Q_2$  0,4797**

**składowa  $Q_3$  0,1220**

Łatwo sprawdzić, że  $Q_1$  wyjaśnia aż 91,61% wariancji całkowitej, komponent  $Q_2$  5,33%, a trzeci  $Q_3$  tylko 1,34% zmienności ogólnej. Łącznie wy-

<sup>13</sup> Parametry te mogą być już, rzecz jasna, szacowane klasyczną metodą najmniejszych kwadratów.

<sup>14</sup> Wyeliminowania nieistotnych ze statystycznego punktu widzenia zmiennych  $Q_i$  można dokonać stosując znane testy na istotność parametrów  $\mu_i$  (m. in. test  $t$ -Studenta na istotność współczynników regresji liniowej). Por. F. Massy, *Principal Components Regression in Exploratory Statistical Research*, Journal of the American Statistical Association 1965/60, s. 234 - 256.

różnione główne składowe wyjaśniają 98,30% wariancji całkowitej, przy czym udział dwóch pierwszych komponentów jest decydujący. Potwierdzenie tego faktu uzyskano również sprawdzając istotność zmiennych  $Q_1$ ,  $Q_2$ ,  $Q_3$ . W tym celu oszacowane zostały parametry modelu kosztów handlowych, w którym zmiennymi objaśniającymi są wyróżnione główne składowe. Otrzymano w konsekwencji następujące oszacowania (tab. 5).

Uzyskane wyniki świadczą wyraźnie, że stopień wyjaśniania kosztów przez główne składowe jest bardzo wysoki. Z drugiej strony warto zauważyć, że składowa  $Q_3$  jest zmienną statystycznie nieistotną, a więc zmienną o niewielkim znaczeniu dla opisu zmienności kosztów handlowych. Uzyskane rezultaty, jakkolwiek bardzo cenne z punktu widzenia wymogów estymacji, nie dają jednak odpowiedzi na najważniejsze pytanie: które ze zmiennych oryginalnych w sposób istotny oddziałują na poziom kosztów handlowych? Odpowiedzi nie można uzyskać na podstawie otrzymanych głównych składowych, gdyż są one zmiennymi ukrytymi, bezpośrednio nieobserwowalnymi, a samo stwierdzenie, że określona składowa występuje w sposób zdecydowany w kilku zmiennych pierwotnych, nie jest wystarczające. Zaproponujemy w związku z tym następujące postępowanie. Ponieważ  $Q_1$  jest pierwszą główną składową, a jej udział w wyjaśnianiu zmienności całkowitej wynosi aż 91,61%, przeto można użyć jej jako kryterium wyboru najbardziej istotnych zmiennych kosztotwórczych. W tym celu przeprowadza się oszacowanie parametrów modelu

$$Q_1 = b_{11}X_1 + b_{21}X_2 + \dots + b_{k1}X_k,$$

a następnie eliminuje się z pełnego zbioru zmiennych te, którym odpowiadają współczynniki  $b_{i1}$  nieistotnie różniące się od zera<sup>15</sup>. Rezultaty zastosowania proponowanego postępowania zaprezentowano w tabeli 6. Krytyczna wartość sprawdzianu hipotezy wynosi <sup>49</sup> $t_{0,05}=2,012$ . Okazuje się więc, że optymalny zestaw zmiennych kosztotwórczych stanowić będą: wartość obrotu towarowego  $X_1$ , liczba punktów sprzedaży  $X_2$ , liczba pracowników operatywnych  $X_5$ , powierzchnia sal sprzedażowych  $X_7$  i wartość zapasu końcowego  $X_8$ . Z czysto ekonomicznego punktu widzenia otrzymane wyniki są bardzo zachęcające. Jest jednak rzeczą ważną stwierdzić, że postępowanie takie kryje w sobie pewne niebezpieczeństwa, a mianowicie:

- 1) składowa  $Q_1$  może stanowić kryterium wyboru tylko wtedy, gdy  $P_i$  jest odpowiednio duże (w praktyce  $p_i$  powinno być większe od 0,8),
- 2) procedura ta prowadzi zwykle do wyodrębnienia cech silnie skorelowanych ze zmienną  $Q_1$  (co jest wskazane), ale często również cech skorelowanych znacznie wzajemnie ze sobą (co już nie świadczy zbyt pozytywnie o metodzie),

<sup>15</sup> Z. Hellwig, op. cit., s. 307 - 321.

Tabela 6

Parametry opisowe modelu głównej składowej  $Q_1$  względem oryginalnych zmiennych objaśniających

Parametr $b_{i1}$	Ocena parametru	Błąd średni parametru	Wartość statystyki $t$ -Studenta
$b_{11}$	0,00026	0,000086	3,03
$b_{21}$	0,66833	0,283407	2,36
$b_{31}$	0,31279	0,382649	0,82
$b_{41}$	0,63046	0,355331	1,77
$b_{51}$	1,00246	0,369744	2,71
$b_{61}$	0,00028	0,007283	0,04
$b_{71}$	0,04224	0,014674	2,88
$b_{81}$	0,00255	0,001014	2,51

3) jeżeli wszystkie zmienne objaśniające  $X_i$  są silnie skorelowane ze sobą i ze zmienną  $Q_1$ , a ich liczba jest duża, może się zdarzyć, że wszystkie  $b_{i1}$  będą mało różnić się między sobą i mało różnić się od zera (zwłaszcza, gdy liczba obserwacji  $n$  jest niewiele większa od liczby zmiennych  $k$ ).

4) w wyniku jego zastosowania mogą też zostać wybrane cechy, które mogą okazać się w części nieistotne (tj. błędy średnie szacunku parametrów przy tych zmiennych będą bardzo duże, przekraczające nawet wartości samych oszacowań), jeśli na ich podstawie będziemy chcieli zbudować model regresji między nimi a zmienną objaśnianą  $Y$ .

W tym kontekście optymistyczne wnioski poczynione uprzednio muszą ulec osłabieniu, tym bardziej, że procedura ta nie gwarantuje wyeliminowania współliniowości między zmiennymi objaśniającymi. Konkludując wypada zauważyć, że MGS nie zawsze jest zbyt efektywną metodą identyfikacji czynników kosztotwórczych. Jest natomiast dobrą metodą eliminacji współliniowości na etapie estymacji modeli za pomocą KMNK.

## V. UWAGI KOŃCOWE

W przedstawionym opracowaniu postawiono sobie dwa zadania. Pierwsze, o charakterze metodycznym, polegające na poszukiwaniu efektywnych metod identyfikacji czynników kosztotwórczych w warunkach istnienia efektu współliniowości zmiennych objaśniających, drugie natury poznawczej, dotyczące określenia czynników wyznaczających poziom kosztów przedsiębiorstw handlu detalicznego. Podjęto więc próbę wykorzystania często ostatnio zalecanej w literaturze MGS, jako że zastosowanie klasycznej regresji, jak wykazano, nie dało pozytywnych rezultatów. Pozytywnych wyników nie przyniósł również najczęściej stosowany w praktyce zabieg określony mianem selekcji zmiennych. Empiryczna weryfi-

kacja metody głównych składowych doprowadziła jednak do stwierdzenia, że w przypadku wysokiego skorelowania zmiennych objaśniających oraz niewielkiej różnicy między liczbą obserwacji a liczbą zmiennych objaśniających, nie jest ona najbardziej efektywną metodą prowadzącą do wydzielenia najistotniejszych cech kosztotwórczych. W tym kontekście zachodzi konieczność poszukiwania innych metod badawczych. Wyniki dociekań w tym zakresie stanowiąc będą przedmiot osobnych rozważań.

THE MAIN FACTORS METHOD AS A METHOD OF FACTORS  
IDENTIFICATION REFLECTING COST LEVEL OF RETAIL  
TRADE ENTERPRISES

Summary

In the elaboration an attempt has been taken up to identify factors determining cost level of retail trade enterprises. At the initial stage of research choice of potential explanatory variables has been done as well as quantitative analysis of their influence on costs when classical linear regression used. It turned out, however, that many explanatory variables are then statistically insignificant. The appraisal of used regression model revealed that it could not be used as a tool of problem solution. The main reason was a strong correlation among explanatory variables, i.e. phenomenon of collinearity.

An attempt to eliminate the collinearity thanks to factor analysis in form of the main factors method has been taken up in further part of the article. It has been shown that the method gives also — along with the complete elimination of collinearity at the stage of costs model estimation — a possibility of separation of the most important determinants from the starting set of potential cost-creating factors. Limitations to the utilization of the presented procedure have also been pointed out.

Empirical exemplification of the presented method has been carried out on\* the ground of source material concerning costs and results of activities of retail trade enterprises of food products in Poland.