ISSN 1507-3866

Piotr Tarka
Uniwersytet Ekonomiczny w Poznaniu

ANALIZA PRZYCZYNOWO-SKUTKOWA W BADANIACH MARKETINGOWYCH Z WYKORZYSTANIEM MODELU RÓWNAŃ STRUKTURALNYCH

Streszczenie: Autor omawia w artykule kluczowe zagadnienia dotyczące badania zależności przyczynowo-skutkowych w kontekście modelu równań strukturalnych SEM wraz z jego praktyczną aplikacją w odniesieniu do czynników determinujących poziom skuteczności badań marketingowych w polskich przedsiębiorstwach.

Słowa kluczowe: model SEM, przyczynowość, analiza ścieżkowa, skuteczność badań marketingowych.

1. Wstęp

W naukach społecznych przyjmuje się, że zmienne będące przyczyną i skutkiem powinny spełniać kilka warunków [Simon 1957; Asher 1976; Sarantakos 1998; Seale 1998; Kozyra 2004]: 1) przyczyna musi poprzedzać skutek w czasie, 2) skutek musi być istotnie powiązany z przyczyną (gdy zmienia się jedna zmienna, to zmienia się też druga), 3) powiązanie to nie może wynikać z wpływu innej zmiennej (wówczas zachodziłaby zależność pozorna), 4) przyczyna i skutek muszą dotyczyć czasu i przestrzeni, 5) musi istnieć racjonalne uzasadnienie, które wyjaśnia relacje przyczynowe między zmiennymi.

Stwierdzenie przyczynowości wymaga warunków na ogół eksperymentalnych, w których zmiennym można nadać określone wartości w sposób kontrolowany, a jednostki można przydzielać w sposób losowy do dwu grup: badanej i kontrolnej. Eksperymenty są szeroko wykorzystywane w naukach przyrodniczych [Angela, Voss 1999]. W badaniach społecznych eksperymenty są trudne do przeprowadzenia, dlatego częściej stosuje się w analizie przyczynowości podejście quasi-eksperymentalne, w którym dowody przyczynowości wyprowadza się wprost z danych. Dla zmiennych mierzonych na skalach słabych przyczynowość bada się zwykle za pomocą tablic kontyngencji [Blalock 1964; 1967]. Dla zmiennych mierzonych na skalach mocnych często stosowaną metodą analizy przyczynowości jest analiza ścieżkowa opracowana przez S. Wrigtha [1960], która jest jednym z głównych elementów modelu równań strukturalnych SEM.

2. Geneza i znaczenie modelu równań strukturalnych (SEM)

Model równań strukturalnych jest mocno ugruntowany na bazie teorii badania związków przyczyno-skutkowych pomiędzy wieloma zmiennymi. W literaturze model ten został szeroko opisany przez wielu badaczy [Everitt 1984; Bentler 1986; Bollen 1989; Mueller 1996; Kaplan 2000]. Proces konstrukcji modelu zakłada pięć następujących po sobie etapów [Bollen 1989]:

- specyfikację/dobór zmiennych w modelu,
- identyfikację (weryfikację) estymowalności modelu,
- estymację parametrów,
- analizę dobroci dopasowania danych do modelu,
- ewentualną modyfikację modelu z uwzględnieniem nowych zmiennych i parametrów, jeśli te nie spełniają oczekiwań na odpowiednim poziomie dopasowania.

Specyfikacja modelu jest dokonywana zgodnie z konstruowaną teorią dotyczącą danego zjawiska, która wyjaśnia, dlaczego pewne zmienne powinny być skorelowane w modelu, a inne nie. Zależności przyczynowe między zmiennymi są określane na podstawie posiadanych hipotez. Zmienne ukryte występujące w modelu są utożsamiane z nieobserwowalnymi zmiennymi będącymi abstrakcyjnymi (*construct*) [Kozyra 2004].

Najpopularniejszym modelem równań strukturalnych SEM (***S**tructural **E**quation **M**odels*) jest LISREL (***Li**near **S**tructural **REL**ationships*), który został spopularyzowany przez K. Jöreskoga i D. Sörboma w roku 1970, a jego nazwa wywodzi się zresztą od oprogramowania komputerowego o tej samej nazwie [Jöreskog 1973; Jöreskog, Sörbom 1979; 1993]. Model ten uwzględnia następującą formę badania relacji przyczyno-skutkowych pomiędzy różnymi zmiennymi. W podejściu tym wykorzystuje się graficzny interfejs, w ramach którego generowane są diagramy – ścieżki zależności. Diagramy te umożliwiają dobór i specyfikację odpowiednich zmiennych do modelu i na ogół służą eksploracji ukrytych związków.

Istnieją w gruncie rzeczy trzy powody, dla których modele równań strukturalnych w zakresie zmiennych ukrytych są wykorzystywane w praktyce. Po pierwsze, w tego rodzaju modelach badacz ma możliwość „podsumowania” (syntetycznego ujęcia) informacji zawartych w wielu zmiennych odpowiedziach do jednej lub kilku kategorii zmiennej/zmiennych ukrytych. Po drugie, prawidłowo zdefiniowane parametry strukturalne modelu pozwalają zminimalizować skutki błędów i odchylenia w rzeczywistych pomiarach przy szacowaniu poszczególnych efektów przyczynowo-skutkowych. Po trzecie, model ten pozwala badać wielkość wpływu i zależności pomiędzy zmiennymi ukrytymi. Po czwarte, do określenia relacji przyczynowo-skutkowych w modelu uwzględnia się różnorodne jego komponenty, w ramach których wykorzystuje się: konfirmacyjną analizę czynnikową w celu analizy ukrytej struktury zmiennych, błędy tzw. obciążenia zmiennych, praktycznie wszystkie formy: klasycznej regresji opartej na jedno- i wielowymiarowych modelach liniowych, w tym odpowiednie testy weryfikujące poziom dobroci dopasowania modelu do danych i zmiennych pojawiających się w określonej strukturze wzajemnych relacji.

3. Diagram analizy ścieżkowej jako graficzna metoda prezentacji zależności przyczynowych w zmiennych

Zmienne ukryte pojawiają się w wielu dziedzinach badań naukowych. W analizach marketingowych występują one chociażby w sferze weryfikacji czynników determinujących poziom konsumpcji, satysfakcji lub wartości wyznawanych przez konsumentów czy w odniesieniu do ukrytych czynników wpływających na skuteczność realizowanych badań marketingowych w firmach. Złożoność postaw i zachowań ludzkich oraz subiektywizm wyrażanych ocen w zakresie poszczególnych cech wewnętrznych sprawia, że badanie relacji pomiędzy zmiennymi obserwowalnymi generującymi zmienne ukryte staje się koniecznością. Wiele z tych zmiennych mierzy jeden wspólny wymiar.

W badaniach ankietowych przyjmuje się, że zmienne obserwowalne określane są przez werbalną ekspresję i zawierają się w prostych stwierdzeniach – pozycjach testowych. Na przykład w analizie wartości konsumentów (określonych przez stwierdzenia) możemy wykazać istnienie ukrytej zmiennej. Przyjmijmy np. cztery stwierdzenia, które będą definiować ukryty wymiar, jakim jest „hedonizm konsumpcyjny":

x_1 – „Spędzam czas na bezustannych rozrywkach",

x_2 – „Potrzebuję ciągle nowych, rzeczy np. nowych ubrań",

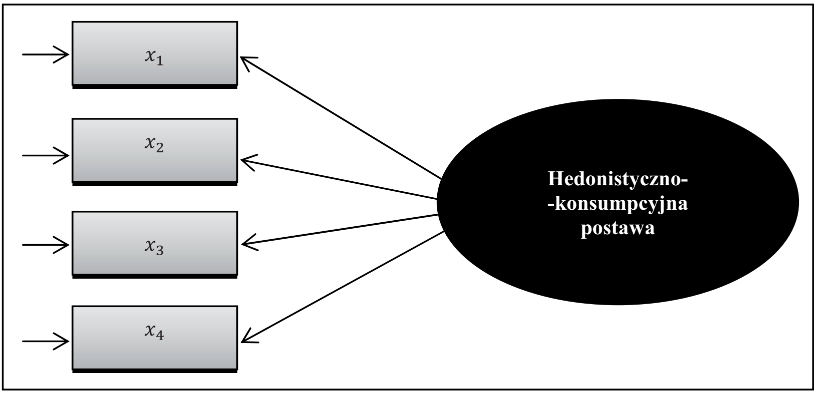
x_3 – „Cenię sobie luksusowe warunki życia",

x_4 – „Poszukuję w swoim życiu systematycznie nowych i niepowtarzalnych bodźców emocjonalnych".

Hipotetyczny model SEM tej struktury zaprezentowano na rys. 1. Model ze zmiennymi ukrytymi i obserwowalnymi powstaje przy użyciu tzw. diagramu ścieżek i ma on szczególne znaczenie przy wizualizacji graficznej wyników, która jest nierozerwalnie związana z procesem analizy i stanowi podstawę w interpretacji wyników uzyskanych w procesie rzutowania w zredukowanej przestrzeni o mniejszej liczbie wymiarów [Sagan 2009].

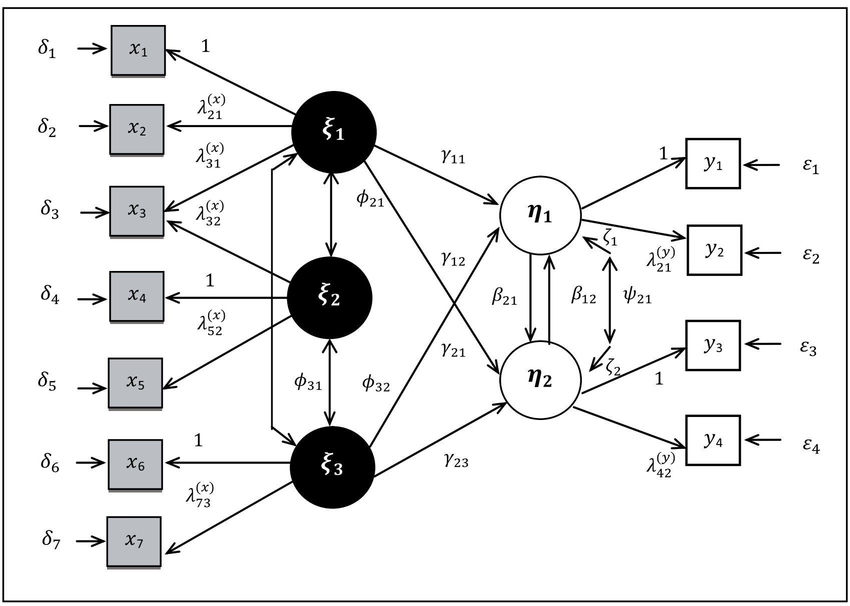
Jak można zaobserwować, ukryta zmienna „Hedonizm konsumpcyjny" stanowi niemierzalną bezpośrednio właściwość, która kształtuje się dopiero pod wpływem czterech zmiennych obserwowalnych – zewnętrznie determinujących jej charakter. Zmienna ukryta służy zatem do syntetycznego ujęcia ukrytej informacji przez analizę wartości/odpowiedzi skorelowanych wzajemnie ze sobą w obrębie zmiennych obserwowalnych. W ten sposób następuje redukcja zestawu odpowiedzi do jednego wymiaru.

W innym z kolei przykładzie z uwzględnieniem tym razem siedmiu zmiennych obserwowalnych (tzw. wskaźników) – oznaczonych jako x – możemy badać ich



Rys. 1. Hipotetyczny model postawy hedonistycznej

Źródło: opracowanie własne.



Legenda: Pola w kolorze czarnym oznaczają zmienne niezależne ukryte – tzw. zmienne egzogeniczne. Pola w kolorze białym określają zmienne zależne ukryte – tzw. zmienne endogeniczne. Kwadraty w kolorze szarym to wskaźniki zmiennych niezależnych ukrytych. Kwadraty w kolorze białym to wskaźniki zmiennych zależnych ukrytych. Wariancje błędów pomiaru znajdują się przy strzałkach dochodzących do zmiennych obserwowalnych. Wartości ładunków czynnikowych znajdujących się na strzałkach określających relacje między zmiennymi ukrytymi a wskaźnikami. Parametr ścieżkowy leży na strzałce wskazującej na relacje między zmiennymi ukrytymi.

Rys. 2. Diagram analizy ścieżkowej na przykładzie siedmiu zmiennych obserwowalnych

Źródło: [Jöreskog, Sörbom 1993].

wpływ na trzy zmienne ukryte – ξ. Na diagramie (rys. 2) zmienna x_3 wpływa zarówno na zmienną ukrytą ξ_1, jak i na zmienną ξ_2. Widoczne są również zmienne ukryte η, z których każda jest powiązana z dwoma niezależnymi wskaźnikami y. Diagram uwzględnia także ς błędy w równaniach modelowych i błędy występujące w zmiennych x oznaczone jako ε i δ.

Diagram ścieżkowy umożliwia zatem przedstawienie modelu równań strukturalnych graficznie, gdzie elementy/symbole takich diagramów mają różne znaczenia. Zaletą modelu SEM jest więc graficzna prezentacja standaryzowanych i niestandaryzowanych równań strukturalnych, gdzie występują zmienne obserwowalne (w postaci prostokątów) i zmienne ukryte (elipsy) [Sagan 2009].

Diagramy ścieżkowe są także szczególnie użyteczne na etapie specyfikacji modelu. Jeżeli w modelu strukturalnym reprezentującym zależności przyczynowe między zmiennymi nie występują zmienne ukryte, to modeli pomiarowych się nie stosuje. Podobnie jeżeli nie bada się żadnych zależności przyczynowych między zmiennymi ukrytymi, to ogólną postać modelu równań strukturalnych upraszcza się do modelu konfirmacyjnej analizy czynnikowej. Według C. Jöreskoga i D. Sörboma [1993], „istnieją dwa podstawowe problemy, które są ważne dla nauk społecznych i behawioralnych. Pierwszy problem dotyczy własności pomiarowych (trafności i rzetelności) narzędzi pomiaru. Drugi problem dotyczy zależności przyczynowych między zmiennymi i względnej siły wyjaśniania tych zależności".

4. Istota i założenia modelu równań strukturalnych SEM

Modele równań strukturalnych SEM stanowią metodę statystycznej analizy wielu zmiennych, polegającą na jednoczesnej estymacji parametrów zbioru wzajemnie powiązanych równań przedstawiających zależności przyczynowo-skutkowe między zmiennymi. Równaniami strukturalnymi nazywa się równania liniowe ze względu na parametry, które są wykorzystywane do analizy wpływu przyczynowego zmiennych oraz pomiaru zmiennych ukrytych [Kozyra 2004]. W zapisie macierzowym dla jednorodnych prób model równań strukturalnych zawiera w sobie wielorównaniowy model regresji liniowej (tzw. model analizy ścieżkowej) i konfirmacyjnej analizy czynnikowej – określanej mianem modelu pomiarowego. Najprostszy zapis modelu SEM przyjmuje postać trzech równań [Bollen 1989]:

$$\eta = \mathrm{A} + \mathrm{B}\eta + \Gamma\xi + \zeta. \tag{1}$$

$$Y = \Lambda_y \eta + \varepsilon. \tag{1a}$$

$$X = \Lambda_x \xi + \delta. \tag{1b}$$

gdzie: η jest $m \times 1$ wektorem ukrytych zmiennych endogenicznych, przy założeniu, że $n \times 1$ wektor ξ ukrytych zmiennych egzogenicznych ma średnią κ i macierz kowariancji Φ oraz że $m \times 1$ wektor ζ błędów losowych ma średnią równą zeru i macierz kowariancji Ψ, a także $\operatorname{cov}(\xi, \zeta') = 0$.

Poza tym: B to macierz $m \times m$ współczynników zależności (z zerami na przekątnej), między zmiennymi endogenicznymi; Γ to macierz $m \times n$ współczynników zależności zmiennych endogenicznych od zmiennych egzogenicznych; $\xi = [\xi_1 \ldots \xi_n]'$ – wektor ukrytych zmiennych egzogenicznych; $\mathrm{Y} = [y_1 \ldots y_n]'$ – wektor zmiennych obserwowalnych, na które wpływają zmienne endogeniczne; $\mathrm{X} = [x_1 \ldots x_n]'$ – wektor zmiennych obserwowalnych, na które wpływają zmienne egzogeniczne; Λ_y – macierz $p \times m$ współczynników zależności (ładunków czynnikowych) zmiennych obserwowalnych od endogenicznych zmiennych ukrytych; Λ_x – macierz $q \times n$ współczynników zależności (ładunków czynnikowych) zmiennych obserwowalnych od egzogenicznych zmiennych ukrytych; $\varepsilon_{p\times 1}$ – wektor błędów pomiaru dla zmiennych obserwowalnych mierzących zmienne endogeniczne; $\delta_{q\times 1}$ – wektor błędów pomiaru dla zmiennych obserwowalnych mierzących zmienne egzogeniczne.

Jeśli $|\mathrm{I} - \mathrm{B}| \neq 0$, i $\mathrm{A} = (\mathrm{I} - \mathrm{B})^{-1}$, wówczas [Bollen 1989]

$$\mu_\eta = \mathrm{A}(\alpha + \Gamma_k), \tag{2}$$

oraz

$$\operatorname{cov}(\eta) = \mathrm{A}(\Gamma\Phi\Gamma' + \Psi)\mathrm{A}'. \tag{3}$$

Założenia w modelu SEM są następujące: błędy pomiaru δ zmiennych x są nieskorelowane z ξ, η i ε, błędy pomiaru ε zmiennych y są nieskorelowane z ξ, η i δ, a błędy losowe ζ są nieskorelowane z ξ. Błędy losowe ξ, η i δ, cechuje homoskedastyczność i brak autokorelacji, natomiast mogą być skorelowane błędy z różnych równań. Ponadto macierz $(\mathrm{I} - \mathrm{B})$ jest nieosobliwa. W celu wyeliminowania z równań wyrazów wolnych przyjmuje się też zazwyczaj, że wektory zmiennych obserwowalnych x oraz y maja zerową wartość oczekiwaną.

Macierz kowariancji zmiennych obserwowalnych wynikająca z modelu równań strukturalnych ma następującą postać [Bollen 1989]:

$$\Sigma = \begin{bmatrix} \Lambda_y (\mathrm{I}-\mathrm{B})^{-1}(\Gamma\Phi\Gamma' + \Psi)\left[(\mathrm{I}-\mathrm{B})^{-1}\right]'\Lambda_y' + \Theta_\varepsilon & \Lambda_y(\mathrm{I}-\mathrm{B})^{-1}\Gamma\Phi\Lambda_x \\ \Lambda_x\Phi\Gamma'\left[(\mathrm{I}-\mathrm{B})^{-1}\right]'\Lambda_y' & \Lambda_x\Phi\Lambda_x' + \Theta_\delta \end{bmatrix}, \tag{4}$$

gdzie: $\Psi = E(\zeta\zeta')$ oznacza macierz kowariancji błędów losowych ζ, $\Phi = E(\xi\xi')$ oznacza macierz kowariancji zmiennych ukrytych ξ, $\Theta_{\varepsilon} = E(\varepsilon\varepsilon')$ oznacza macierz kowariancji błędów pomiaru ε, a $\Theta_{\delta} = E(\delta\delta')$ oznacza macierz kowariancji błędów pomiaru δ.

A zatem w odniesieniu do modeli SEM przyjmuje się, że p obserwowalne zmienne endogeniczne są reprezentowane przez wektor y, a zmienne q obserwowalne egzogeniczne (zawarte w wektorze x) odnoszą zmienne obserwowalne do czynników (ukrytych zmiennych). Obie kategorie zmiennych można określić jako [Bollen 1989]:

$$y = \tau_y + \Lambda_y \eta + \varepsilon,\ E(\varepsilon) = 0,\ Cov(\varepsilon) = \Theta_{\varepsilon} \tag{5}$$

$$x = \tau_x + \Lambda_x \xi + \delta,\ E(\delta) = 0,\ Cov(\delta) = \Theta_{\varepsilon}. \tag{6}$$

Średnie wektorów dla zmiennych obserwowalnych to z kolei

$$\mu_y = \tau_y + \Lambda_y \mathrm{A}(\alpha + \Gamma\kappa), \tag{7}$$

W badanej (jednorodnej) populacji τ_y, τ_x, α i κ nie będą identyfikowalne, jeśli nie zostaną przyjęte pozostałe uwarunkowania, takie jak

$$\Sigma_y = \Lambda_y \left[\mathrm{A}\left(\Gamma\Phi\Gamma' + \Psi\right)\mathrm{A}'\right]\Lambda_y' + \Theta_{\varepsilon}, \tag{8}$$

$$\Sigma_x = \Lambda_x \Phi \Lambda_x' + \Theta_{\delta}, \tag{9}$$

oraz

$$\Sigma_{yx} = \Lambda_y \mathrm{A}\Gamma\Phi\Lambda_x'. \tag{10}$$

Z powyższych założeń przyjętych w (7), (8), (9), (10) wynika, że struktura kowariancji dla zmiennych obserwowalnych w populacji w modelu SEM określana jest jako:

$$\Sigma = \mathrm{cov}\begin{bmatrix} y \\ x \end{bmatrix} = \begin{bmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{bmatrix}. \tag{11}$$

Z kolei dla struktury średnich zmiennych obserwowalnych

$$\mu = E\begin{bmatrix} y \\ x \end{bmatrix} = \begin{bmatrix} \mu_y \\ \mu_x \end{bmatrix}. \tag{12}$$

Reasumując, należy stwierdzić, że model SEM dopasowuje strukturę kowariancji i średnich (określonych w (11) i (12)) do danych na podstawie obserwowalnych zmiennych. W ten sposób umożliwia on zarówno analizę danych na prostej próbie losowej, jak i analizę bardziej złożonych danych.

5. Identyfikacja, estymacja i dopasowanie parametrów w modelu równań strukturalnych

Określenie zależności między zmiennymi w modelu powinno być dokonane w taki sposób, aby wszystkie parametry w modelu były identyfikowalne. Identyfikacja parametrów liniowych modeli wielorównaniowych jest zagadnieniem dobrze znanym w ekonometrii [Goldberger 1975]. W modelu równań strukturalnych parametr określa się jako identyfikowalny, jeśli można go przedstawić jako funkcję jednego elementu lub większej liczby elementów macierzy Σ [Bollen 1989]. Jeżeli wszystkie estymowane parametry modelu będą identyfikowalne, to cały model określa się jako **identyfikowalny**. Oznacza to, że dla określonej macierzy kowariancji zmiennych obserwowalnych wynikającej z modelu istnieje tylko jeden zbiór wartości parametrów modelu. Model jest identyfikowalny tylko wtedy (warunek konieczny), gdy liczba szacowanych parametrów t jest mniejsza od liczby elementów leżących na głównej przekątnej i pod główną przekątną macierzy kowariancji lub jest jej równa, czyli jest spełniona nierówność [Kozyra 2004]:

$$t \leq \frac{1}{2}(q+p)(q+p+1). \tag{13}$$

Warunki wystarczające identyfikacji całego modelu otrzymuje się przez dwuetapową procedurę: w pierwszym etapie przeformułowuje się model tak, aby można było ustalić identyfikację modelu konfirmacyjnej analizy czynnikowej dla ukrytych zmiennych endogenicznych i ukrytych zmiennych egzogenicznych. W drugim etapie ustala się identyfikację modelu analizy ścieżkowej, w którym zmienne ukryte traktuje się tak jak zmienne obserwowalne. Jeśli model jest identyfikowalny na każdym z tych dwu etapów, to jest identyfikowalny również w całości.

W celu **dopasowania modelu** pożądane jest zapewnienie nadidentyfikalności, gdy liczba **szacowanych parametrów** jest mniejsza niż liczba elementów macierzy kowariancji leżących na głównej przekątnej i pod główną przekątną, gdyż umożliwia to wybór spośród dopuszczalnych rozwiązań takiego, dla którego funkcja dopasowania osiąga minimum, oraz testowanie dopasowania modelu do danych empirycznych.

Estymacja parametrów modelu równań strukturalnych polega z kolei na znalezieniu takich parametrów, aby różnica między zaobserwowaną macierzą kowariancji S a macierzą wynikającą z modelu Σ była jak najmniejsza zgodnie z przyjętym kryterium [Bollen 1989]. Estymacja parametrów jest dokonywana zwykle metodami iteracyjnymi, które znajdują minimum funkcji dopasowania macierzy kowariancji Σ do macierzy S. Najczęściej wykorzystuje się do estymacji *metodę naj-*

większej wiarygodności (ML) lub *uogólnioną metodę najmniejszych kwadratów (GLS)*. Funkcje dopasowania tych metod mają następującą postać [Bollen 1989; Kozyra 2004]:

$$F_{ML} = \ln\det(\Sigma) + tr(S\Sigma^{-1}) - \ln\det(S) - (p+q), \tag{14}$$

$$F_{GLS} = \frac{1}{2} tr\left(\left(\mathrm{I} - \Sigma S^{-1}\right)^2\right), \tag{15}$$

gdzie: $tr(A)$ oznacza ślad macierzy A, $\det(A)$ zaś reprezentuje wyznacznik macierzy A.

Minimum funkcji dopasowania *F* otrzymane w wyniku estymacji podanymi wyżej metodami jest wykorzystywane do **oceny dobroci dopasowania modelu**. Jeżeli zmienne obserwowalne mają wielowymiarowy rozkład normalny, to statystyka $(N-1)F$ ma symptotyczny rozkład chi-kwadrat χ^2 z $k = 1/2(q+p)(q+p+1) - s$ stopniami swobody. Stosunek wielkości próby do liczby estymowanych parametrów powinien zawierać się między 5:1 a 10:1 [Kelloway 1998], przy czym minimalna liczebność próby N powinna wynosić co najmniej 100 obserwacji. Oprócz testu chi-kwadrat opracowano bardzo wiele miar dopasowania modelu. Ich omówienie można znaleźć w literaturze [Bollen 1989; Kelloway 1998; Mueller 1996; Miszczak 2003; 2004].

Ocena dopasowania modelu służy do podjęcia decyzji o ewentualnej **modyfikacji przyjętej specyfikacji modelu** w celu poprawy dopasowania lub w celu zmniejszenia liczby zmiennych w modelu. Modyfikacja, która jest dokonywana na podstawie uzyskanych zależności empirycznych, polega na ustaleniu parametrów nieistotnie różnych od zera jako równych zeru lub zezwoleniu na estymację parametrów wcześniej ustalonych jako równych zeru. Zmodyfikowany model należy testować na nowej próbie w celu weryfikacji poprawy dopasowania. Modyfikacja modelu stanowi eksploracyjne, a nie konfirmacyjne podejście w analizie, gdyż model zmodyfikowany oznacza zmianę teorii na temat danego zjawiska na podstawie uzyskanych danych empirycznych [Kozyra 2004].

6. Badanie przyczynowości w modelu analizy ścieżkowej

Model analizy ścieżkowej jest wielorównaniowym modelem regresji liniowej, przy czym zmienna niezależna w jednym równaniu może się stać zmienną zależną w innym. Kolejne równania odzwierciedlają zależności przyczynowe, gdyż zakłada się zależność przyczynową zmiennej zależnej od zmiennych niezależnych. Analiza ścieżkowa służy więc do oszacowania bezpośredniego i pośredniego wpływu przyczynowego zmiennych niezależnych na zmienne zależne.

W modelu tym występują zmienne ukryte, odzwierciedlające uproszczoną strukturę modelu z (1, 1*a* – *b*). Model analizy ścieżkowej przedstawia się za pomocą równania [Bollen 1989]:

$$Y = \mathrm{B}Y + \Gamma X + \zeta. \tag{16}$$

Modele analiz ścieżkowych przyjmują z reguły albo postać **modeli rekurencyjnych** bądź są to **modele nierekurencyjne**. W *modelu rekurencyjnym* związki występujące pomiędzy zmiennymi y są jednokierunkowe (gdzie występuje strzałka z jednym końcem „grotem strzały"). Z kolei w *modelu nierekurencyjnym* związki są dwukierunkowe (dwie strzałki po obu końcach) z tzw. sprzężeniami zwrotnymi. W modelu nierekurencyjnym zależności przyczynowe mogą być wzajemne, czyli dwie zmienne mogą być równoczesne dla siebie swoimi przyczynami i skutkami. W modelu rekurencyjnym macierz B określana jest przez dolną część macierzy (z Ψ jako przekątną). Macierz tę można przedstawić za pomocą macierzy trójkątnej, tj. takiej, w której niezerowe elementy występują tylko pod główna przekątną, a oprócz tego macierz Ψ musi być diagonalna, czyli błędy losowe nie mogą być ze sobą skorelowane. Często wektor y jest podzielony na dwa podzbiory, które generują strukturę blokową dla B i Γ. Dla przykładu dla dwóch podzbiorów zmiennych struktura składająca się z B i Ψ jest następująca

$$\mathrm{B} = \begin{bmatrix} \mathrm{B}_{11} & 0 \\ \mathrm{B}_{21} & \mathrm{B}_{22} \end{bmatrix} \text{ i } \Psi = \begin{bmatrix} \Psi_{11} & 0 \\ 0 & \Psi_{22} \end{bmatrix}. \tag{17}$$

Modele z tego typu strukturą określane są jako blokowo-powtarzalne.

Macierz kowariancji zmiennych x i y wynikająca z modelu analizy ścieżkowej jest uproszczeniem macierzy (4) i ma następującą postać

$$\Sigma = \begin{bmatrix} (\mathrm{I}-\mathrm{B})^{-1}(\Gamma\Phi\Gamma' + \Psi)\left[(\mathrm{I}-\mathrm{B})^{-1}\right]' & (\mathrm{I}-\mathrm{B})^{-1}\Gamma\Phi \\ \Phi\Gamma'\left[(\mathrm{I}-\mathrm{B})^{-1}\right]' & \Phi \end{bmatrix}. \tag{18}$$

Strukturę modelu w analizie ścieżkowej można przekształcić w postać zredukowaną, co z kolei pozwala traktować zmienne endogeniczne jako losowe zmienne egzogeniczne z błędem losowym. Zazwyczaj macierz $(\mathrm{I}-\mathrm{B})$ musi być nieosobliwa w celu uzyskania tzw. zredukowanej postaci modelu [Timm 2002]:

$$y = (\mathrm{I}-\mathrm{B})^{-1}\Gamma x + (\mathrm{I}-\mathrm{B})^{-1}\zeta = \Pi x + \varepsilon. \tag{19}$$

Przyjmuje się również, że zmienne $x_i\,(i = 1 \ldots q)$ oraz $y_i\,(i = 1 \ldots p)$ mają zerową wartość oczekiwaną w celu wyeliminowania wyrazów wolnych w równaniach. Macierz Π zawiera zredukowane współczynniki i wektor ze zredukowaną liczbą błędów losowych ε. Porównując model (16) z ogólnym modelem równań strukturalnych (1, 1a – b), można zauważyć, że wszystkie zmienne w modelu są obserwowalne (poza błędami losowymi) i są mierzone bez błędów pomiaru. Zmienne ukryte w ogólnym modelu są w modelu analizy ścieżkowej (16) równe obserwowalnym $\eta = y\ (p = m),\ \xi = x,\ (q = n)$ [Kozyra 2004].

Warunkiem koniecznym identyfikacji modelu analizy ścieżkowej jest podobne jak w ogólnym modelu równań strukturalnych spełnienie nierówności (17). W ramach <u>identyfikacji modelu</u> mogą się pojawić także inne podobne modele pomiarowe. Jeśli w zredukowanej formie modelu strukturalnego przyjmiemy, że $\mathrm{B} = 0$, wówczas model analizy ścieżkowej jest niczym innym jak wspomnianym poprzednio wielowymiarowym modelem regresji. Stąd Γ, Φ i Ψ określa się jako obserwowane macierze kowariancji S

$$S = \begin{bmatrix} S_{yy} & S_{yx} \\ S_{xy} & S_{xx} \end{bmatrix}, \tag{20}$$

a parametry mają strukturę

$$\Phi = S_{xx}, \Gamma' = S_{xx}^{-1} S_{xy} \ \text{ i } \ \Psi = S_{yy} - S_{yx} S_{xx}^{-1} S_{xy}. \tag{21}$$

W praktyce jednak ta reguła nie do końca się sprawdza w samej analizie, ponieważ większość modeli ma strukturę, gdzie $\mathrm{B} \neq 0$. Znacznie korzystniej jest, jeśli w modelu rekurencyjnym B występuje jako dolna podstawa trójkąta, a Ψ jest przekątną. Wyrównując Σ z S, otrzymujemy [Timm 2002]

$$\begin{aligned} S_{yy} &= (\mathrm{I} - \mathrm{B})^{-1} \left[(\Gamma \Phi \Gamma' + \Psi)(\mathrm{I} - \mathrm{B})^{-1} \right]', \\ S_{xy} &= \Phi \Gamma' \left[(\mathrm{I} - \mathrm{B})^{-1} \right]', \\ S_{yy} &= \Phi. \end{aligned} \tag{22}$$

Ponieważ Φ jest określone, stąd dalej możemy zdefiniować pozostałe elementy Γ, Ψ, i B, wprowadzając każde równanie jednocześnie do analizy. Także ponieważ $(\mathrm{I} - \mathrm{B})^{-1}$ jest podstawą prostokąta, rozwiązanie pierwszego równania zależy tylko od pierwszego wiersza w Γ. W ten sposób można odnaleźć γ_1', pierwszy wiersz w Γ i następnie ψ_{11}. Z drugiego równania otrzymujemy drugi wiersz w Γ i następnie w drugim wierszu B i ψ_{22}. Kontynuując w ten sposób obliczenia, identyfikujemy wszystkie po kolei parametry [Timm 2002].

Po zakończeniu identyfikacji parametrów analizujemy następnie **dopasowanie zmiennych względem rozpatrywanego modelu**. W analizie ścieżkowej tworzone są dwie macierze $\mathrm{B} = [\beta_{ij}]$ i $\Gamma = [\gamma_{ij}]$. Zawierają one *efekty bezpośrednie* w odniesieniu wpływu przyczynowego y_i na y_j lub x_i na x_j oraz *efekty pośrednie* wpływu przyczynowego zmiennej x_i na y_i i *efekty łączne*. Efekt pośredni bazuje na relacji y_i w y_j i x_i w x_j i jest rozdzielony w tym samym czasie przez zmienne pośredniczące oraz odzwierciedla iloczyny efektów bezpośrednich. Łączny efekt jest zaś sumą wszystkich pośrednich i bezpośrednich efektów. Model jest dopasowany idealnie do danych, przy $k = 0$ stopni swobody.

Aby ocenić, jaki jest efekt *y* na *y*, należy ponownie powrócić do macierzy B będącej macierzą bezpośrednich efektów. Mnożąc macierz B przez siebie (tę samą macierz), otrzymujemy macierz B^2 zawierającą efekty pośrednie z jedną zmienną oraz macierz B^3 reprezentującą pośrednie efekty z dwiema zmiennymi interweniującymi. Ponieważ suma całkowita (T) efektu jest sumą efektów pośrednich (I) i efektów bezpośrednich (D), stąd całkowita suma efektu *y* na *y* to [Timm 2002]

$$\mathrm{T}_{yy} = D_{yy} + \mathrm{I}_{yy} = \mathrm{B}\sum_{k=2}^{\infty}\mathrm{B}^k. \tag{23}$$

Model rekurencyjny, gdzie B stanowi dolną podstawę trójkąta, zawiera zera na przekątnej. W tej sytuacji całkowity efekt z *y* na *y* przybiera formułę [Timm 2002]

$$\mathrm{T}_{yy} = \mathrm{B} + \mathrm{B}^2 + \ldots \mathrm{B}^k = \mathrm{B}^0 + \mathrm{B}^1 + \ldots + \mathrm{B}^k - \mathrm{I} = (\mathrm{I} - \mathrm{B})^{-1} - \mathrm{I}. \tag{24}$$

stosując

$$\sum = \sum_{j=0}^{m}\left(\prod_{i=1}^{j}\mathrm{B}_i\right) L_j \left(\prod_{i=1}^{j}\mathrm{B}_i\right), \tag{25}$$

z $\mathrm{B}_i = \mathrm{I}$ i pośrednim efekcie przy y na y otrzymujemy [Timm 2002]

$$\mathrm{I}_{yy} = \mathrm{T}_{yy} = \mathrm{B} = (\mathrm{I} - \mathrm{B})^{-1} - \mathrm{I} - \mathrm{B}. \tag{26}$$

Aby z kolei ocenić efekt z *x* na *y*, należy ponownie wyznaczyć model rekurencyjny przez określenie łącznej sumy efektu jako sumy bezpośrednich i pośrednich efektów

$$\mathrm{T}_{xy} = D_{xx} + \mathrm{I}_{xy} = \Gamma + \Gamma\mathrm{B} + \Gamma\mathrm{B}^2 + [\mathrm{I} + \mathrm{B} + \mathrm{B}^2 + \ldots +]\Gamma = (\mathrm{I} - \mathrm{B})^{-1}\Gamma, \tag{27}$$

który jest identyczny z macierzą współczynników dla zredukowanej postaci modelu. Stąd macierz niebezpośrednich efektów przy x na y to [Timm 2002]

$$\mathrm{I}_{xy} = \mathrm{T}_{xy} - \Gamma = (\mathrm{I} - \mathrm{B})^{-1} - \Gamma. \tag{28}$$

Analiza przyczynowości za pomocą analizy ścieżkowej jest przeprowadzana zwykle z wykorzystaniem danych przekrojowych pobieranych mniej więcej w jednym czasie. Przy tak pobieranych danych można nie zarejestrować efektu przyczynowego, jeśli czas ujawnienia się skutków określonych przyczyn jest długi [Miszczak 2003]. W takim przypadku wskazane jest korzystanie z danych uzyskiwanych w różnych okresach. Przyczynowości nie można dowieść statystycznie [Kelloway 1998; Bollen 1989], można co najwyżej weryfikować zgodność modelu zależności przyczynowych z danymi. Nawet dla modelu dobrze dopasowanego do danych empirycznych wymagane jest, w celu stwierdzenia przyczynowości, zapewnienie trzech warunków w fazie budowy modelu [Kozyra 2004]:

1) odłączenia od innych zmiennych,
2) uwzględnienia powiązań między zmiennymi,
3) ustalenia kierunku zależności przyczynowych.

Przyjęcie kolejności czasowej zmian w zmiennych za warunek konieczny zależności przyczynowej uniemożliwia badanie modeli nierekurencyjnych zawierających obustronne zależności przyczynowe. Dlatego modele analizy ścieżkowej służące do badania zależności przyczynowych nazywa się często modelami strukturalnymi po to, by nie uwypuklać zagadnienia przyczynowości [Miszczak 2003]. Analiza ścieżkowa może być dokonywana nie tylko dla zmiennych obserwowalnych, lecz także dla zmiennych ukrytych. W tym drugim przypadku model zależności przyczynowych musi być uzupełniony o modele pomiaru zmiennych ukrytych.

7. Wykorzystanie modelu SEM do oceny skuteczności badań marketingowych w polskich przedsiębiorstwach

W celu prezentacji najistotniejszych założeń wyróżnionych w modelu SEM dobrano przykład, w którym zarówno sam zbiór danych, jak i skonstruowany na jego podstawie model dotyczył analizy wybranych czynników oddziałujących na skuteczność badań marketingowych w polskich przedsiębiorstwach. Przykład ten pochodził z przeprowadzonych przez autora w roku 2010 badań empirycznych na próbie 250 jednostek (przedsiębiorstw) [Tarka 2011]. Wybrane stwierdzenia badawcze, które zostały włączone do modelu, podano w tab. 1. Postawione w pracy hipotezy utworzyły model z udziałem 9 zmiennych obserwowalnych, które miały bezpośredni wpływ na zmienne ukryte. Model ten obejmował relacje przyczynowo-skutkowe o bardzo prostej strukturze, odbiegając tym samym znacznie od założeń konstrukcji modelu SEM o zaawansowanych parametrach. Poszczególne relacje pomiędzy zmiennymi były testowane przez analizę ścieżki w programie LISREL, badające zależności pomiędzy zmiennymi obserwowalnymi a zmiennymi ukrytymi endogenicznymi i egzogenicznymi [Tarka 2011].

Ze względów praktycznych w modelu założono w pierwszej fazie model zupełny (ze wszystkimi możliwymi ścieżkami i ich współczynnikami) pomiędzy zmiennymi. Następnie wyeliminowano niepotrzebne ścieżki i sprawdzono, w jaki sposób zredukowana postać modelu odtwarzała macierz kowariancji. Reprodukcja ta została przeprowadzona za pomocą testu chi-kwadrat. Liczba iteracji potrzebna do przeprowadzenia analizy wyniosła 26. Do obliczeń wybrano metodę maksymalnej wiarygodności. Strukturę modelu SEM zaprezentowano na rys. 3. Jak można zaobserwować, w modelu tym zachodzą także interakcje pomiędzy grupami zmiennych zależnych endogenicznych. Zakładany model odzwierciedla dobre dopasowanie do danych wyjściowych do modelu zgodnie z poziomem $\chi^2(23) = 45,94$ przy stosunkowo niskim indeksie RMSEA = 0,054 oraz wysokim GFI = 0,93 i AGFI = 0,87 (biorąc pod uwagę określone granice przedziałów indeksów GFI i AGFI od 0 do 1 i przedziały ufności: 90, 95, 99 dla RMSEA).

Specyfikacja parametrów obejmowała następujące konfiguracje powiązań pomiędzy zmiennymi obserwowalnymi (p5, p6, p7, p8, p9, p10) i zmiennymi endogenicznymi („Systematyczność", „Potrzeba", „Procesy") a zmiennymi obserwowalnymi (p11, p12, p13) i pojedynczą zmienną egzogeniczną („Świadomość").

Tabela 1. Wybrane stwierdzenia badawcze włączone do modelu

SYSTEMATYCZNOŚĆ BADAŃ

- **P5.** Nasza firma przeprowadza systematycznie badania marketingowe, przez co poprawia skuteczność samych badań
- **P6.** Systematycznie przeprowadzane analizy satysfakcji konsumentów wpływają korzystnie na poprawę skuteczności badań

POTRZEBA INFORMACYJNA I AKTUALIZACJI BAZY DANYCH

- **P7.** Wyniki z badań marketingowych (jakie przeprowadziliśmy do tej pory) przechowujemy w bazie danych naszej firmy
- **P8.** Częste przeprowadzanie badań marketingowych w celu uaktualnienia naszej marketingowej bazy danych o konsumentach przyczynia się do poprawy skuteczności badań marketingowych

PROCESY KOMUNIKACJI

- **P9.** Badacze marketingowi i analitycy nie mają problemów w zakresie komunikacji, tj. sformułowania czytelnego przekazu istotnych informacji z wyników badań dla menedżerów naszej firmy, co z kolei wpływa korzystnie na skuteczność samych badań
- **P10. Nasz personel z departamentu ds. marketingu omawia często z innymi departamentami wewnątrz naszej firmy analizę przyszłych stanów rynku i zachowań naszych konsumentów i tym samym poprawia odbiór wyników z przeprowadzonych badań marketingowych**

ŚWIADOMOŚĆ KADRY KIEROWNICZEJ I ZARZĄDU FIRMY WOBEC SKUTECZNOŚCI BADAŃ MARKETINGOWYCH

- **P11.** Zarząd oraz kierownictwo naszej firmy jest przekonane o tym, że sukces i powodzenie firmy na rynku zależy w dużej mierze od przeprowadzonych badań marketingowych
- **P12.** Zarząd oraz kierownictwo naszej firmy chętnie przeznacza środki finansowe na doskonalenie podejmowanych przez firmę procesów badawczych
- **P13.** Zarząd i kierownictwo naszej firmy angażuje się osobiście w sferę projektowania nowych metod i technik badawczych zwiększających skuteczność badań marketingowych.

Źródło: opracowanie własne na podstawie przeprowadzonych badań empirycznych w programie LISREL (wersja 8.80).

Oszacowana macierz (uwzględniająca ładunki czynnikowe dla poszczególnych powiązań w modelu) wskazywała na silne i istotne związki pomiędzy zmiennymi ukrytymi (oznaczonymi przez Y) – „Systematyczność", a zmiennymi obserwowalnymi: p5 i p6; oraz „Procesy komunikacyjne" a zmienną p9. Z kolei w zakresie zmiennej oznaczonej symbolem X najsilniejszy związek występował pomiędzy „Świadomością" a zmienną obserwowalną – p11. W tabelach 2 i 3 zamieszczono wartości współczynników między zmiennymi ukrytymi a ich wskaźnikami oraz wartości współczynników między zmiennymi ukrytymi endogenicznymi i egzogenicznymi.

Tabela 2. Wartości współczynników (zmienne ukryte a wskaźniki tych zmiennych)

Lambda – Y		Lambda – X	
Wskaźniki zmiennych zależnych ukrytych wobec zmiennych zależnych ukrytych (endogenicznych)		Wskaźniki zmiennych niezależnych ukrytych wobec zmiennej niezależnej – ukrytej egzogenicznej	
P5 → *Systematyczność*	0,41	P11 → *Świadomość*	0,81
P6 → *Systematyczność*	0,44	P12 → *Świadomość*	0,47
P7 → *Potrzeba*	0,23	P13 → *Świadomość*	0,50
P8 → *Potrzeba*	0,37		
P9 → *Procesy*	0,72		
P10 → *Procesy*	0,28		

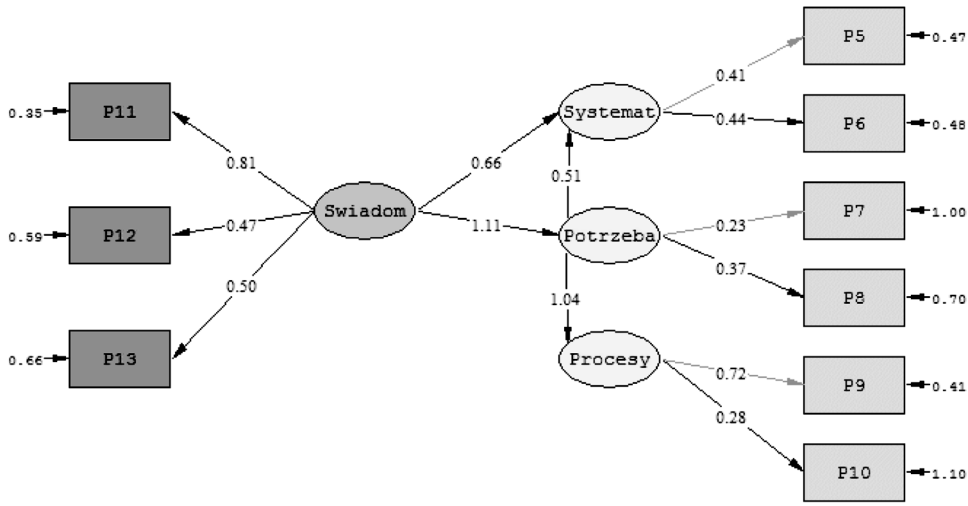
Źródło: opracowanie własne w programie LISREL (wersja 8.80).

Tabela 3. Wartości współczynników między zmiennymi ukrytymi

Kierunek oddziaływania	Wartości
Świadomość → Systematyczność	0,66
Świadomość → Potrzeby	1,11
Potrzeby → Procesy	1,04
Potrzeby → Systematyczność	0,51

Źródło: opracowanie własne w programie LISREL (wersja 8.80).

Obserwacja badanych relacji pomiędzy wybranymi zmiennymi ukrytymi endogenicznymi i zmienną egzogeniczną wskazuje na silne oddziaływanie zmiennej „Świadomość" w kierunku takich zmiennych, jak: „Procesy komunikacyjne" – 1,04, „Potrzeby informacyjne i aktualizacja danych" – 1,11 oraz „Systematyczność badań" – 0,66. Świadomość w znacznej mierze oddziałuje na potrzeby informacyjne firm, które w pierwszej kolejności determinują procesy wewnętrznej komunikacji w zakresie przepływu informacji marketingowych, a następnie w dalszej kolejności (w mniejszym stopniu) determinują systematyczność badań. Istotny jest w tym względzie fakt, że procesy komunikacji nie są bezpośrednio uwarunkowane poziomem świadomości kadry zarządzającej, lecz kształtują się one niejako pośrednio przez potrzeby informacyjne firm, które opierają się na wcześniejszym założeniu, tj. „świadomości menedżerskiej".



Dane w zakresie istotności modelu: Chi-kwadrat = 45,94, df = 23, wartość p = 0,072, RMSEA = 0,054.

Rys. 3. Diagram ścieżek

Źródło: opracowanie własne w programie LISREL 8.80.

Reasumując, można stwierdzić, że zasadność prowadzenia badań marketingowych przez firmy, a ściślej rzecz ujmując – stopień ich zaangażowania w zwiększanie skuteczności badań, wynika w znacznej mierze ze świadomości menedżerów. Należy sądzić, że im większa świadomość i wiedza „odbiorców" w sferze przydatności badań i pozyskiwanych na ich podstawie informacji marketingowych, tym większe zainteresowanie i realnie podejmowane działania firm w sferze poprawy skuteczności badań oraz analitycznego myślenia w zakresie otoczenia zewnętrznego.

8. Podsumowanie

Model równań strukturalnych SEM zaprezentowany w niniejszej pracy posłużył do opisu związków występujących pomiędzy zmiennymi obserwowalnymi oraz zmiennymi latentnymi (ukrytymi) w zakresie problematyki skuteczności badań marketingowych w polskich firmach. Niewątpliwie model SEM (jako strategia badawczo-analityczna) jest w tym względzie o tyle istotny, że pozwala w znacznej mierze wpłynąć na sposób „konstrukcji" teorii naukowej. Z literatury wiadomo, że koncepcja modelu SEM jest powszechnie stosowana w wielu różnych dyscyplinach naukowych, m.in. w biologii, medycynie, genetyce, psychologii czy socjologii. Sam model pomaga badaczowi w uchwyceniu właściwej ścieżki klarownego myślenia i zrozumieniu całego mechanizmu generującego dane. Jednakże stosowanie tego modelu statystycznego zmusza badacza do zdefiniowania, przynajmniej w minimalnym stopniu, mechanizmu przyczynowego, który leży u podstaw zaobserwowanych korelacji/kowariancji między zmiennymi obserwowalnymi modelu SEM.

Rozwój SEM w obecnym czasie powoduje wiele zmian w sposobie stosowania statystyki w nauce. Aplikowanie SEM stymuluje budowanie i wykorzystywanie modeli zamiast prostego opisu i eksploracji statystycznej, co zwiększa z kolei zainteresowanie badaczy wyjaśnieniami przyczynowymi, filozofią przyczynowości oraz warunkami pozwalającymi na wnioskowanie przyczynowe. Dzięki pracy wielu badaczy i naukowców SEM staje się ogólnym modelem zmiennej latentnej (ukrytej), który w połączeniu z liniowym modelem mieszanym lub modelem hierarchicznym jest dzisiaj najbardziej ogólnym modelem statystycznym. SEM jako model ogólny zmiennej ukrytej obejmuje m.in.: tradycyjną analizę czynnikową, analizę czynnikową dla zmiennych dyskretnych, analizę latentnych profili czy analizę latentnych grup.

Literatura

Angela D., Voss D.T., *Design and Analysis of Experiments*, Springer-Verlag, New York 1999.

Asher H.B., *Causal Modeling*, CA: Sage Publications, Beverly Hills 1976.

Bentler P.M., *Structural modeling and psychometrika: An historical perspective on growth and achievements*, „Psychometrika" 1986, no 51.

Blalock H.M., *Causal Inferences in Non-Experimental Research*, University of North Carolina Press, Chapel Hill 1964.

Blalock H.M., *Causal inference, closed populations, and measures of association*, „American Political Science Review" 1967, no 61 (March).

Bollen K.A., *Structural Equations with Latent Variables*, John Wiley, New York 1989.

Everitt B.S., *An Introduction to Latent Variable Method*s, Chapman and Hall, London 1984.

Goldberger A.S., *Teoria ekonometrii*, PWE, Warszawa 1975.

Jöreskog K.G., *A General Method for Estimating a Linear Structural Equation System*, [w:] A.S. Goldberger, O.D. Duncan, *Structural Equation Models in the Social Sciences*, Academic Press, New York 1973a.

Jöreskog K.G., Sörbom D., *Advances in Factor Analysis and Structural Equation Models*, MA: Abt Associates Inc., Cambridge 1979.

Jöreskog K.G., Sörbom D., *LISREL* 8: *User's Reference Guide*, Scientific Software International, Chicago 1993.

Kaplan D., *Structural Equation Modeling Foundations and Extensions*, CA: Sage Publications, Thousand Oaks 2000.

Kelloway E.K., *Using LISREL for Structural Equation Modeling – a Researchers' Guide*, Sage, Thousand Oaks 1998.

Kozyra C., *Modele analizy i oceny jakości usług*, Praca doktorska, AE, Wrocław 2004.

Miszczak W., *Analiza przyczynowości*, [w:] *Pomiar statystyczny*, W. Ostasiewicz (red.) Wrocław 2003.

Miszczak W., *LISREL w analizie jakości życia*, [w:] *Ocena i analiza jakości życia*, W. Ostasiewicz (red.), AE, Wrocław 2004.

Mueller R.O., *Basic Principles of Structural Equation Modeling*, Springer-Verlag, New York 1996.

Sagan A., *Prezentacja graficzna modeli równań strukturalnych*, [w:] *Wizualizacja wyników badań marketingowych – podejścia, metody i zastosowania*, M. Walesiak (red.), Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu nr 86, UE, Wrocław 2009.
Sarantakos S., *Social Research*, 2nd ed., MacMillan Press, London 1998.
Simon H.A., *Spurious Correlation: A Causal Interpretation*, John Wiley&Sons, Inc., New York 1957.
Seale C., *Statistical Reasoning – Casual Arguments and Multivariate Analysis*, [w:] *Researching Society and Culture*, C. Seale (red.), Sage, London 1998.
Tarka P., *Zastosowanie modelu LISREL do oceny skuteczności badań marketingowych*, Materiały z konferencji „Metody pomiaru i analizy rynku usług – przydatność i efektywność", Poznań, 17 czerwca 2011.
Timm N.H., *Applied Multivariate Analysis*, Springer-Verlag, New York 2002.
Wright S., *Patch coefficients and patch regressions – alternative or complementary concepts?*, "Biometrics" 1960, vol. 16, no 2.

ANALYSIS OF CAUSE AND EFFECT IN MARKETING RESEARCH BASED ON STRUCTURAL EQUATION MODEL (SEM)

Summary: In the article the author discusses key problems referring to exploration and research within cause and effect events in the context of model or structural equations describing factors that determine marketing research effectiveness in Polish companies.

Keywords: SEM model, causality, marketing research effectiveness.