

STUDIA METODOLOGICZNE

Agnieszka PLEŚNIAK

Wybór metody estymacji w budowie skali czynnikowej

Skala jest narzędziem do mierzenia zmiennych (Nowak, 2007) i stanowi regułę przypisania wartości obiektom (Górniak, 2000). Wartość obiektu dla zmiennych ukrytych jest kombinacją wyników uzyskanych na podstawie poszczególnych stwierdzeń (pytań). Skala oparta na czynnikach powstaje przez zsumowanie wartości dla poszczególnych stwierdzeń, zaś skala czynnikowa — w wyniku zastosowania analizy czynnikowej¹. Jednak taka klasyczna analiza nie daje badaczowi sposobności narzucenia z góry określonej struktury powiązań między czynnikami (zmiennymi ukrytymi) a stwierdzeniami (zmiennymi obserwowalnymi)². Możliwości takiej dostarcza modelowanie równań strukturalnych (SEM), a ściślej konfirmacyjna analiza czynnikowa³.

Modelowanie równań strukturalnych, z uwagi na dużą elastyczność i możliwość modelowania związków między zmiennymi, a także możliwość oceny dopasowania modeli pomiarowych, zyskuje coraz większą popularność. Zapro-

¹ Więcej informacji na temat skal czynnikowych i opartych na czynnikach można znaleźć m.in. w Górniak (2000), s. 128, zaś na temat budowy skal czynnikowych w Sagan (2003).

² Zmienne obserwowalne w modelowaniu równań strukturalnych nazywane są również wskaźnikami.

³ Modelowanie równań strukturalnych, obok związków między zmiennymi ukrytymi i obserwowalnymi (model pomiarowy), umożliwia również modelowanie związków między ukrytymi zmiennymi (model strukturalny). Etap polegający na budowie modelu pomiarowego nazywany jest konfirmacyjną analizą czynnikową — więcej na temat konfirmacyjnej analizy czynnikowej (CFA) oraz różnic między CFA i klasyczną eksploracyjną analizą czynnikową w Bollen (1989).

ponowano i rozwinięto wiele metod estymacji, które mają odmienne własności i wymagania co do rozkładu i rozmiaru próby.

W artykule zostaną omówione najbardziej popularne metody estymacji i ich własności, ze wskazaniem, w jakich sytuacjach poszczególne metody najlepiej sobie radzą. Obok własności metod uwzględniono również wyniki badań symulacyjnych dotyczących stopnia ich odporności na odchylenia od rozkładu normalnego, rozmiar próby, błędną specyfikację modelu itp. Następnie przedstawiono przykład empiryczny obrazujący rozważane zagadnienia.

METODY ESTYMACJI I ICH WŁASNOŚCI

Wybór metody estymacji parametrów modelu zależy od skali pomiaru zmiennych obserwowalnych oraz ich rozkładu, a także liczebności próby.

Jeśli zmienne są ciągłe i mają wielowymiarowy rozkład normalny, najczęściej stosuje się metodę największej wiarygodności (ML). Maksymalizuje ona wiarygodność parametrów, co jest równoznaczne z minimalizacją funkcji rozbieżności (Hoyle, 1995)⁴:

$$F_{ML} = \log |\Sigma(\theta)| + \text{tr}(\mathbf{S}\Sigma(\theta)^{-1}) - \log |\mathbf{S}| - q \quad (1)$$

gdzie:

- $\mathbf{S}$ — macierz kowariancji i wariancji obliczona na podstawie próby N -elementowej,
- $\Sigma(\theta)$ — macierz kowariancji i wariancji wynikająca z modelu,
- $\theta = [\theta_1, \theta_2, \dots, \theta_t]$ — wektor parametrów modelu,
- q — liczba zmiennych obserwowalnych.

Założenia, jakie powinny być spełnione w przypadku estymatora GLS (uogólniona metoda najmniejszych kwadratów) są nieco mniej restrykcyjne, zmienne nie muszą mieć wielowymiarowego rozkładu normalnego, ale mogą mieć inny, z niewielką kurtozą⁵. Uogólniona metoda najmniejszych kwadratów minimalizuje funkcję (np. Browne, 1973):

⁴ Dla uproszczenia funkcję rozbieżności $F(\mathbf{S}, \Sigma(\theta))$ będę oznaczać jako F , dodając w dolnym subskrypcie oznaczenie zastosowanej metody, np. F_{ML} oznacza $F(\mathbf{S}, \Sigma(\theta))$ dla metody największej wiarygodności.

⁵ Wystarczy, że asymptotyczna macierz kowariancji $\mathbf{ACOV}(s_{ij}, s_{gh}) = N^{-1}(\sigma_{ig}\sigma_{jh} + \sigma_{ih}\sigma_{jg})$, gdzie s_{ij} są elementami macierzy kowariancji obliczonej na podstawie próby, zaś σ_{ih} — macierzy kowariancji w populacji generalnej. Powyższy warunek spełniony jest również dla innych — niż wielowymiarowy normalny — rozkładów z niewielką kurtozą; por. Browne (1973), s. 2; Bollen (1989), s. 417. Ponadto Bollen (1989), s. 114 zwraca uwagę, że estymator ML należy do rodziny estymatorów GLS i w efekcie również jego asymptotyczne własności nie zależą od tego, czy zmienne obserwowalne użyte jako wskaźniki mają rozkład wielowymiarowy normalny, o ile mają one rozkład wymagany dla GLS. Zaś Browne (1973), s. 17 i 18 udowadnia, że estymatory ML mają asymptotyczne własności BGLS (z ang. *Best Generalized Least Squares*, czyli najlepszych estymatorów GLS; najlepszych — w sensie — najmniejszych wariancji).

$$F_{GLS} = \frac{1}{2} \text{tr}[(\mathbf{S} - \boldsymbol{\Sigma}(\boldsymbol{\theta}))\mathbf{V}^{-1}]^2 \quad (2)$$

gdzie $\mathbf{V}$ — macierz wag⁶.

Najczęściej $\mathbf{V}$ równa się $\mathbf{S}$, co odpowiada minimalizacji sumy kwadratów reszt przeważonych przez odwrotność $\mathbf{S}$, zaś wzór (2) upraszcza się wówczas do:

$$F_{GLS} = \frac{1}{2} \text{tr}[\mathbf{I} - \boldsymbol{\Sigma}(\boldsymbol{\theta})\mathbf{S}^{-1}]^2 \quad (3)$$

GLS dostarcza wyników asymptotycznie równoważnych do wyników ML (pod warunkiem że postać modelu jest poprawnie dobrana), ma podobne właściwości jak ML (por. Olsson, Troye, Howell, 1999), zaś jej zastosowanie wymaga podobnego czasu.

Estymatory uzyskane metodą ML oraz GLS, w przypadku wielowymiarowego rozkładu normalnego oraz innego rozkładu z kurtozą zbliżoną do kurtozy wielowymiarowego rozkładu normalnego, są zgodne i asymptotycznie efektywne, zaś standardowe błędy szacunku oraz statystyka chi-kwadrat w teście dopasowania modelu⁷ są poprawnie oszacowane. W przypadku rozkładu eliptycznego (bez asymetrii i z tym samym stopniem kurtozy) estymatory są nadal zgodne i asymptotycznie efektywne. Warto zaznaczyć, że nawet w przypadku rozkładu znacznie odbiegającego od wielowymiarowego rozkładu normalnego, estymatory uzyskane metodą GLS i ML są zgodne, ale nie są już asymptotycznie efektywne, zaś standardowe błędy szacunku nie są poprawnie oszacowane (są one zaniżone).

W przypadku błędnej specyfikacji ML dostarcza bardziej realistycznych miar dopasowania⁸ i mniej obciążonych wartości parametrów, a to z powodu odmiennej macierzy wag wykorzystanej w funkcjach rozbieżności ML i GLS (por. wzór (2) i przypis 6). Elementy macierzy wag dla ML oparte są na macierzy kowariancji wynikającej z modelu $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ (gdzie $\hat{\boldsymbol{\theta}}$ jest oszacowaniem ML), podczas gdy w GLS — na zaobserwowanej macierzy kowariancji ($\mathbf{S}$) i nie zależą tym samym od postaci modelu. Tak więc w tym ostatnim przypadku $\mathbf{S}$ będzie

⁶ Jeśli macierz wag będzie równa $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$, gdzie $\hat{\boldsymbol{\theta}}$ jest oszacowaniem ML, otrzymamy F_{ML} , zaś w przypadku gdy $\mathbf{V} = \mathbf{I}$ — funkcję rozbieżności dla nieważonej metody najmniejszych kwadratów F_{ULS} ; por. Bollen (1989), s. 429, Olsson, Foss, Troye, Howell, (2000), s. 562.

⁷ Hipoteza zerowa głosi, że macierz kowariancji w populacji generalnej jest równa macierzy kowariancji wynikającej z modelu, więcej informacji na temat statystyki chi-kwadrat w teście dopasowania modelu można znaleźć m.in. w Bollen (1989), s. 263—269 oraz w Hoyle (1995), s. 77—81. W dalszej części artykułu używając określenia statystyka chi-kwadrat będę miała na myśli statystykę w teście dopasowania modelu.

⁸ W badaniach symulacyjnych dla modeli z błędami specyfikacji modelu, modele oszacowane metodą GLS wykazywały zbyt optymistycznie empiryczne dopasowanie, co wiązało się z gorszym dopasowaniem teoretycznym; por. Olsson, Foss, Troye, Howell (2000).

zbiegać do Σ niezależnie od tego, czy postać modelu jest poprawna. W przypadku ML macierz kowariancji $\Sigma(\hat{\theta})$ będzie zbiegać do Σ tylko w wypadku poprawnej specyfikacji.

Gdy dane są ciągłe, a rozkład nie jest wielowymiarowy normalny, może być zastosowana odporna metoda największej wiarygodności (RML)⁹ lub metoda ADF (ang. *asymptotically distribution free*) — asymptotycznie niewrażliwa na rozkład lub według tłumaczenia zaproponowanego przez A. Sagana — nieparametryczna, niezależna od rozkładu. Estymacja metodą ADF polega na minimalizacji funkcji rozbieżności wykorzystywanej w ważonej metodzie najmniejszych kwadratów (WLS), przy czym macierzą wag $\mathbf{W}$ jest asymptotyczna macierz kowariancji (Kaplan, 2000):

$$F_{WLS} = (s - \sigma)^T \mathbf{W}^{-1} (s - \sigma) \quad (4)$$

gdzie:

- s — wektor zawierający elementy górnej części macierzy $\mathbf{S}$, łącznie z przekątną,
- σ — wektor zawierający analogiczne elementy macierzy $\Sigma(\theta)$,
- $\mathbf{S}$ — macierz kowariancji i wariancji obliczona na podstawie próby N -elementowej (w przypadku zmiennych porządkowych macierz korelacji polichorycznej),
- $\Sigma(\theta)$ — macierz kowariancji i wariancji wynikająca z modelu i jego założeń,
- $\theta = [\theta_1, \theta_2, \dots, \theta_t]$ — wektor parametrów modelu.

Badania symulacyjne wskazują, że w przypadku ADF próba powinna wynosić 5000 obserwacji¹⁰ (np. Yuan, Bentler, 1997), w przypadku większej liczby zmiennych obserwowalnych i większej kurtozy odpowiednio więcej (Boomsma, Hoogland, 2001). Warto zaznaczyć, że jeśli wskaźniki mają wielowymiarowy rozkład normalny i specyfikacja modelu jest poprawna, wyniki otrzymane metodą ADF są asymptotycznie zbliżone do otrzymanych metodą ML i GLS. W sytuacji błędnej specyfikacji modelu (i rozkładu wielowymiarowego normalnego), z uwagi na macierz wag użytą w funkcji rozbieżności, wyniki uzyskane metodą ADF będą zbliżone do uzyskanych metodą GLS.

Skala pomiaru rzutuje na wybór miary korelacji i pośrednio na wybór metody estymacji. W przypadku zmiennych ciągłych współczynnik korelacji liniowej Pearsona jest naturalnym wyborem, natomiast dla zmiennych mierzonych na skalach dychotomicznych lub porządkowych powinno wybrać się korelację te-

⁹ Niewiele jest badań poświęconych RML i trudno określić, jaki rozmiar próby jest wystarczający. Fakt, że konieczne jest oszacowanie asymptotycznej macierzy kowariancji wskazuje na to, że nie powinien być on zbyt mały.

¹⁰ Dla prostych modeli przyjmuje się, że wystarczy 1000 obserwacji (Hoyle, 1995, s. 68; Curran i in., 1996, s. 16—29).

trachoryczną (dla skal dychotomicznych) lub polichoryczną (dla skal porządkowych), więcej na temat korelacji polichorycznej i tetrachorycznej można znaleźć m.in. u U. Olssona (1979a) i A. Pleśniak (2005).

W praktyce współczynnik korelacji liniowej Pearsona stosuje się często również dla skal porządkowych. Postępowanie takie może być uzasadnione, jeśli asymetria rozkładów poszczególnych stwierdzeń nie jest wysoka, a skala wystarczająco długa (minimum 5-stopniowa, jak podaje tutorial Amosa (por. też Labovitz, 1967), choć im dłuższa skala, tym lepiej, gdyż zjawisko tłumienia korelacji jest słabsze (Olsson, 1979b)¹¹. Poszczególne kategorie skali porządkowej traktowane są jako środki przedziałów. Zakłada się, że wyższa pozycja skali odpowiada większemu natężeniu badanej cechy (postawy) oraz że odległości między kolejnymi kategoriami skali są równe. Dalsza procedura nie odbiega od procedury stosowanej w przypadku zmiennych ciągłych.

Opisane podejście jest jednak krytykowane (Jöreskog, 2002) z uwagi na fakt, że założenie o równych odległościach między poszczególnymi kategoriami skali może nie być spełnione w praktyce. Ponadto często występuje silna asymetria, co w połączeniu z małą liczbą kategorii może zniekształcać korelację Pearsona (model oszacowany na podstawie błędnie oszacowanych współczynników korelacji będzie błędny).

W przypadku modelowania na podstawie macierzy korelacji polichorycznej, Muthén i Jöreskog (Jöreskog, 2002; Bollen, Long, 1993; Muthén, 1984) rozwinęły metodologię dla zmiennych kategorialnych CVM (ang. *categorical variable methodologies*), opartą na estymatorze asymptotycznie nieparametrycznym, niezależnym od rozkładu ADF.

Wykorzystanie asymptotycznej macierzy kowariancji, jako macierzy wag $\mathbf{W}$ we wzorze (4), w przypadku modeli bardzo rozbudowanych może nastroczać trudności obliczeniowych ze względu na wymiar macierzy, gdyż wymaga ona oszacowania momentów centralnych czwartego rzędu. W przypadku zbyt małej próby asymptotyczna macierz kowariancji może nie być dobrze oszacowana, co pogłębia się przy obliczaniu jej odwrotności, wpływając na wyniki otrzymane ważoną metodą najmniejszych kwadratów — WLS (por. wzór (4), które nie są stabilne. Ponadto, jeśli asymptotyczna macierz kowariancji nie jest dodatnio określona, to zastosowanie estymatora ADF nie jest możliwe.

K. Jöreskog sugeruje, by w sytuacji małej liczebności próby wykorzystać metodę największej wiarygodności do oceny modelu, a następnie poprawić standardowe błędy szacunku i wartości statystyki chi-kwadrat (por. przypis 7), używając asymptotycznej macierzy kowariancji (Jöreskog, 2002).

W. Wothke (Bollen, Long, 1993) proponuje zaś wykorzystanie nieważonej metody najmniejszych kwadratów (ULS). Metoda ta jest techniką odporną i nie wymaga założeń dotyczących rozkładu, ale nie szacuje standardowych błędów szacunku i jest wrażliwa na skalę. Zaletą nieważonej metody najmniejszych

¹¹ Jednak zastosowanie zbyt dużej liczby kategorii może prowadzić do dzielenia skali przez respondentów i w efekcie — do wielomodalnych rozkładów.

kwadratów jest fakt, że może być zastosowana w przypadku macierzy korelacji, które nie są dodatnio określone (co często zdarza się w przypadku macierzy korelacji polichorycznych)¹². Metoda ULS dostarcza zgodnych punktowych ocen parametrów modelu, ale nie efektywnych. Może zatem być stosowana, gdy standardowe błędy szacunku nie są potrzebne (czyli np. w eksploracyjnej analizie czynnikowej lub do obliczenia wartości początkowych). Funkcja rozbieżności dla nieważonej metody najmniejszych kwadratów (ULS) jest opisana wzorem (2), z tym że macierz wag jest macierzą jednostkową, wobec czego wspomniany wzór upraszcza się do:

$$F_{ULS} = \frac{1}{2} \text{tr}[(S - \Sigma(\theta))^2] \quad (5)$$

PRÓBA KONSTRUKCJI SKALI CZYNNIKOWEJ

Dla zobrazowania rozważań posłużono się przykładem. Dane wykorzystane w opracowaniu pochodzą z badania ankietowego przeprowadzonego w 2006 r. przez CEBOS w Instytucie Statystyki i Demografii SGH. Próba losowa liczyła 1320 gospodarstw domowych¹³. W analizie uwzględniono jedynie gospodarstwa domowe, których głowa nie osiągnęła wieku emerytalnego (dla kobiet 60 lat, dla mężczyzn 65 lat)¹⁴. Było ich 982.

Analizie poddano stwierdzenia oceniające ważność potrzeby zabezpieczenia własnej starości na tle innych potrzeb, według klasyfikacji V. Schraba (omówionej w Rudnicki, 2000).

Stwierdzenia respondentów oceniające znaczenie potrzeby zabezpieczenia własnej starości na tle innych potrzeb określały zmienne stanowiące odpowiedzi na pytania zawarte w kwestionariuszu — Czy zabezpieczenie warunków na starość jest mniej ważne, tak samo ważne czy ważniejsze niż...

*I*1 — zdobywanie pieniędzy i osiągnięcie wyższych dochodów,

*I*2 — zdobywanie popularności,

*I*3 — zdobywanie szacunku,

¹² Problem braku dodatniej określoności może pojawić się również na etapie ustalania wartości początkowych. Większość metod to metody iteracyjne, które wymagają ustalenia początkowych wartości parametrów, bowiem jeśli macierz $\Sigma(\theta)$, wynikająca z początkowych wartości parametrów, nie jest dodatnio określona, wówczas dalsza estymacja staje się niemożliwa. Szczególnie dotyczy to metody największej wiarygodności, inne metody, jak uogólniona metoda najmniejszych kwadratów, radzą sobie z tym problemem; por. Bollen, Long (1993), s. 270. W takim przypadku można wykorzystać nieważoną metodę najmniejszych kwadratów do obliczenia wartości początkowych.

¹³ Dobór próby dwustopniowy, warstwowy.

¹⁴ Rozwiązanie takie ma swoje wady, gdyż emerytami mogą być osoby, które nie osiągnęły wieku emerytalnego (wcześniejsze emerytury). W 2006 r. spośród gospodarstw domowych, których głowa nie osiągnęła wieku emerytalnego, 9% zaliczało się do emerytów, natomiast 11% pobierało emeryturę.

- I/4 — urządzenie sobie obecnie wygodnych warunków życia,
- I/5 — zdobywanie awansu społecznego,
- I/6 — poprawienie wyglądu zewnętrznego,
- I/7 — zdobycie uznania wśród współpracowników i znajomych,
- I/8 — uzyskanie lub zachowanie dobrej kondycji fizycznej,
- I/9 — zapewnienie rozrywki i przyjemności.

Respondenci oceniali ważność potrzeby zabezpieczenia własnej starości względem pozostałych posługując się 11-stopniową skalą Stapela (od -5 do +5, gdzie -5 oznaczało zdecydowanie mniej ważne, 0 tak samo ważne, +5 zdecydowanie ważniejsze). Z kolei 11,2% respondentów wybrało kategorię „trudno powiedzieć” w przypadku co najmniej jednej pary potrzeb, nie ujawniając tym samym preferencji potrzeb, natomiast nie wystąpiły odmowy odpowiedzi.

Z jednej strony pominięcie braków odpowiedzi w analizach zniekształca ich wyniki, zaś z drugiej, jak słusznie zauważyli G. King, J. Honaker, A. Joseph i K. Scheve (2001), nie ma sensu imputować wartości, które nie istnieją. Imputacja jest uzasadniona jedynie w sytuacji, gdy respondent ma postawę, ale jej nie ujawnia. W praktyce trudno jest jednoznacznie określić, czy respondent rzeczywiście nie ma opinii na dany temat czy jedynie nie chce jej ujawnić, maskując ten fakt stwierdzeniem „trudno powiedzieć”¹⁵. Dlatego odpowiedzi „trudno powiedzieć” nie zostały imputowane, zaś szacując macierz kowariancji braki danych wyłączono parami.

Do konstrukcji skali czynnikowej w celu pomiaru zabezpieczenia się na starość gospodarstw domowych oraz oceny jej trafności zastosowano modelowanie równań strukturalnych (konfirmacyjną analizę czynnikową). Obliczenia zostały dokonane w programach SPSS (Amos) i LISREL.

Przy wyborze metody estymacji parametrów modelu pomiarowego uwzględniono wymagania metod co do rozkładów i skali pomiaru oraz wyniki badań symulacyjnych nad wrażliwością poszczególnych metod estymacji na odstępstwa od założeń oraz skutkami ich niespełnienia.

Zmienne służące jako wskaźniki w modelu pomiarowym orientacji na starość mierzone były na skali porządkowej. W takiej sytuacji można obliczyć współczynniki korelacji polichorycznej (lub potraktować skalę jako przedziałową) i współczynniki korelacji Pearsona. W artykule przyjęto drugi sposób z dwóch powodów. Efekt tłumienia korelacji dla skal 11-stopniowych jest słaby (por. Olsson, 1979 b)¹⁶. Ponadto asymetria rozkładów odpowiedzi dla poszcze-

¹⁵ Sytuacja w rzeczywistości jest bardziej skomplikowana. Obok respondentów, którzy nie ujawnili opinii (niezależnie od tego, czy ją mieli czy też nie), można wyobrazić sobie respondentów niemających opinii na dany temat i ukrywających ten fakt przez wybieranie kategorii neutralnej czy też, zdaniem respondenta, zgodnej z oczekiwaniami ankietera. Grynberg i Gilliam (1993), s. 348—357 wyodrębniają rzeczywiste i pozorne postawy oraz rzeczywisty i pozorny brak postawy w zależności od tego, czy respondent ma postawę i czy ją ujawnia.

¹⁶ Średnia różnica między współczynnikami korelacji liniowej Pearsona i polichorycznej wyniosła -0,03.

gólnych stwierdzeń występuje jedynie w przypadku pięciu z dziewięciu stwierdzeń (przy poziomie istotności 0,05) i nie jest ona silna, ponadto ma ten sam kierunek¹⁷, z wyjątkiem jednego stwierdzenia (tabl. 1).

TABL. 1. KLASYCZNE WSPÓŁCZYNNIKI ASYMETRII I KURTOZY ORAZ WARTOŚCI MINIMALNE I MAKSYMALNE ZMIENNYCH

Zmienna	Asymetria		Kurtoza		Minimum		Maksimum	
		<i>p</i>		<i>p</i>		liczebność		liczebność
<i>V</i> 1	-0,150	0,060	-0,276	0,050	-5	28	5	158
<i>V</i> 2	<u>0,309</u>	0,000	<u>-1,152</u>	0,000	-5	202	5	116
<i>V</i> 3	-0,181	0,024	-0,474	0,000	-5	79	5	100
<i>V</i> 4	-0,244	0,002	-0,169	0,270	-5	28	5	110
<i>V</i> 5	-0,137	0,088	-0,430	0,001	-5	95	5	58
<i>V</i> 6	-0,008	0,917	-0,921	0,000	-5	92	5	101
<i>V</i> 7	-0,038	0,629	-0,512	0,000	-5	72	5	67
<i>V</i> 8	-0,226	0,005	-0,408	0,001	-5	48	5	111
<i>V</i> 9	-0,068	0,390	-0,643	0,000	-5	65	5	104

p — krytyczny (empiryczny) poziom istotności.

Ź r ó d ł o: obliczenia własne.

Jak wspomniano, traktowanie skali porządkowej jako przedziałowej wiąże się z przyjęciem założenia o istnieniu ukrytej zmiennej ciągłej, która jest mierzona za pomocą skali porządkowej oraz — jeśli do obliczeń posłużyły liczby odpowiadające poszczególnym stopniom skali — o równych odległościach między poszczególnymi stopniami skali.

Przy wyborze metody estymacji istotne jest czy zmienne obserwowalne mają wielowymiarowy rozkład normalny. Założenie to można sprawdzić za pomocą testu opartego na kurtozie Mardii dla rozkładu wielowymiarowego lub sprawdzając, czy jednowymiarowe rozkłady są normalne. Jednak drugi sposób może służyć jedynie do wykrycia odstępstw od wielowymiarowego rozkładu normalnego, gdyż spełnienie założeń o jednowymiarowych rozkładach normalnych nie oznacza spełnienia założenia o wielowymiarowym rozkładzie normalnym, jest to warunek konieczny, ale niewystarczający. Założenie o wielowymiarowym rozkładzie normalnym nie jest spełnione, gdyż rozkłady brzegowe nie są normalne (na co wskazuje istotna asymetria i/lub kurtoza).

Z teoretycznego punktu widzenia, z uwagi na rozkład wskaźników, najlepsza byłaby metoda asymptotycznie niewrażliwa na rozkład (ADF). Jak wspomniano, wymaga ona jednak dużej liczebności próby ($n > 5000$), aby poprawnie oszaco-

¹⁷ Silna asymetria w różnych kierunkach jest przeciwwskazaniem do traktowania skali porządkowej jako przedziałowej.

wać asymptotyczną macierz kowariancji. Dla zbyt małych prób stosowanie metody ADF jest nieefektywne, gdyż daje ona gorsze oceny parametrów niż metoda największej wiarygodności czy uogólniona metoda najmniejszych kwadratów.

Można też spotkać się z poglądem, że w przypadku jeśli asymetria i kurtoza nie wykraczają poza przedział $(-1, 1)$, odstępstwa od wielowymiarowego rozkładu normalnego nie zniekształcają wyników. Warunek ten spełniają wszystkie wskaźniki poza jednym, dla którego kurtoza nieznacznie przekracza -1 .

Według tych rozważań najbardziej odpowiednia wydaje się metoda największej wiarygodności. Estymacji dokonano metodą największej wiarygodności na podstawie macierzy kowariancji. Ponadto, przedstawiono wyniki uzyskane uogólnioną metodą najmniejszych kwadratów (w przypadku poprawnej specyfikacji modelu powinna dać ona zbliżone wyniki). Obliczeń dokonano w pakiecie Amos.

ALTERNATYWNE MODELE POMIAROWE

Konceptualny model pomiarowy zakłada istnienie zmiennej ukrytej (jednej lub więcej) wyjaśniającej obserwowane korelacje między zmiennymi obserwowalnymi.

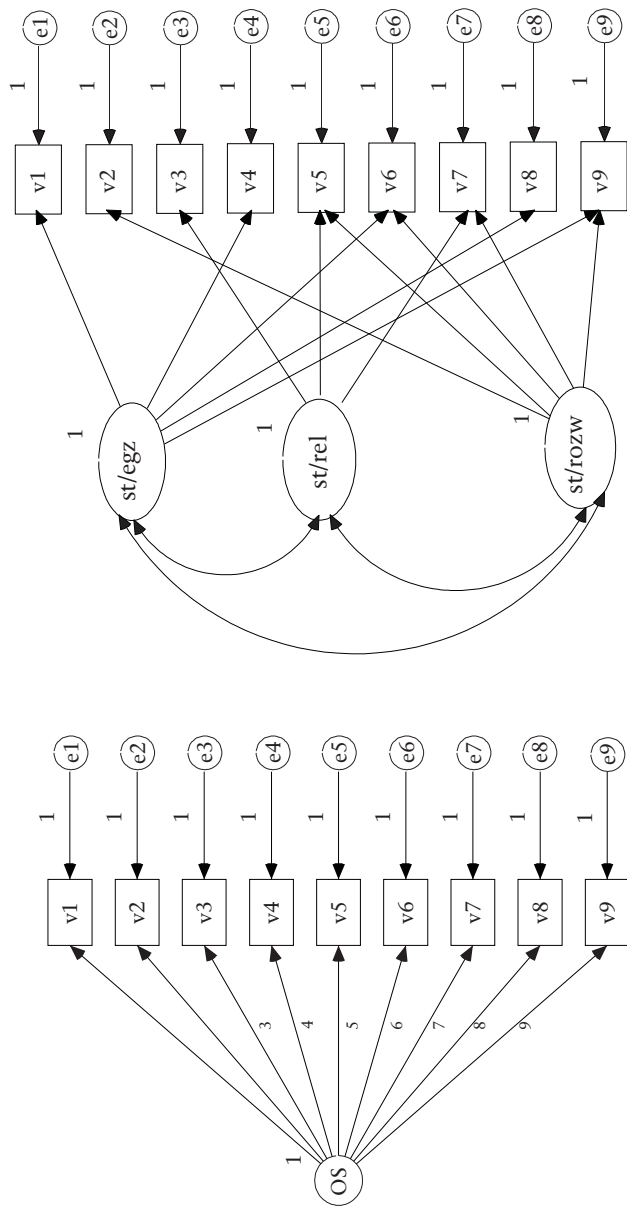
Budując model pomiarowy należy zastanowić się nad wymiarowością, tzn. czy jeden wymiar w pełni odzwierciedla strukturę zjawiska czy też konieczne jest uwzględnienie więcej wymiarów, by w pełni opisać badane zjawisko. Model zakładający istnienie jednego wymiaru jest prostszy i przez to bardziej pożądany z punktu widzenia badacza. Jeśli można bowiem jakieś zjawisko wyjaśnić w prostszy sposób, po co budować model bardziej skomplikowany, skoro model ze swej natury jest uproszczeniem.

Jednak na wymiarowość orientacji na starość będzie wpływała struktura potrzeb, które zostały użyte, by ocenić relatywną ważność potrzeby zabezpieczenia własnej starości. Potrzeby z tego samego poziomu hierarchii lub z tej samej grupy mogą być podobnie oceniane pod względem ich ważności. Kierując się z jednej strony prostotą, zaś z drugiej strukturą potrzeb zaproponowano model jednoczynnikowy i trójczynnikowy.

Model trójczynnikowy wynika z teorii ERG C. P. Alderfera¹⁸. Zakłada on istnienie trzech ukrytych czynników mierzących znaczenie potrzeby zabezpieczenia własnej starości na tle potrzeb egzystencjalnych (st/egz), na tle potrzeb związanych z poziomem relacji (st/rel) oraz z poziomem rozwoju (st/rozw). Konceptualny model jedno- i trójczynnikowy przedstawia wykr. 1.

¹⁸ Model ERG Alderfera dzieli potrzeby na trzy poziomy: egzystencji (E), relacji (R) i rozwoju (G). Poziom egzystencji odpowiada potrzebom fizjologicznym i bezpieczeństwa w hierarchii A. Masłowa, poziom relacji — potrzebom miłości oraz szacunku i uznania, wreszcie poziom rozwoju tworzą potrzeby szacunku i uznania oraz samorealizacji.

KONCEPTUALNY MODEL POMIARÓW ORIENTACJI NA STAROŚĆ JEDNO- I TRÓJCZYNNIKOWY



OS — orientacja na starość, zmienna ukryta, v_1 — v_9 — stwierdzenia dotyczące znaczenia potrzeby zabezpieczenia własnej starości na tle innych potrzeb (por. tabl. 1), i_{ij} — ładunki ścieżkowe (czynniki) w modelu jedno- i trójczynnikiem, e_1 — e_9 — błędy pomiaru, st/egz, st/rel, st/rozw — zmienne ukryte w modelu trójczynnikiem: ocena ważności potrzeby zabezpieczenia własnej starości na tle potrzeb egzystencjalnych, relacji oraz rozwoju — odpowiednio, i_k — współczynniki korelacji między zmiennymi ukrytymi.

Źródło: opracowanie własne.

ANALIZA RZETELNOŚCI SKAL

Eksploracyjna analiza czynnikowa (czynniki wyodrębniono metodą głównych składowych) pozwoliła sprawdzić, czy istnieją podstawy do wyodrębniania ukrytych struktur na podstawie zaobserwowanych korelacji pomiędzy zmiennymi obserwowalnymi. Do oceny zasadności wyodrębniania ukrytych czynników wykorzystano test sferyczności macierzy korelacji Bartletta, statystykę Kaisera-Meyera-Olkina oraz procent wariancji wskaźników tworzących skalę — odtworzonej przez pierwszą główną składową (por. tabl. 2).

TABL. 2. NIEKTÓRE WŁASNOŚCI SKAL

Skale	Liczba stwierdzeń	Wartość statystyki KMO	λ	ss	Wyjaśniane wariacje w %	Współczynnik α -Cronbacha
Orientacja na starość (skala jednoczynnikowa)	9	0,834	2825,24	36	43,59	0,830
Ocena ważności potrzeby zabezpieczenia własnej starości na tle potrzeb egzystencjalnych	5	0,736	920,07	10	47,68	0,724
Ocena ważności potrzeby zabezpieczenia własnej starości na tle potrzeb relacji	3	0,667	826,73	3	69,84	0,783
Ocena ważności potrzeby zabezpieczenia własnej starości na tle potrzeb rozwoju	5	0,807	1451,39	10	57,02	0,805

λ — wartość statystyki w teście Bartletta dla $p < 0,001$, ss — stopnie swobody.

Źródło: obliczenia własne.

Wartości statystyki w teście Bartletta pozwalają stwierdzić, że macierze korelacji nie są macierzami jednostkowymi w przypadku wszystkich analizowanych skal. Warto dodać, że modelowanie równań strukturalnych, jako oparte na zaobserwowanych kowariancjach, również nie będzie miało sensu, jeśli macierz korelacji będzie jednostkowa. Z kolei wysokie wartości statystyki Kaisera-Meyera-Olkina świadczą o tym, że zaobserwowane macierze korelacji są produktem oddziaływania wspólnych czynników.

Dla wszystkich skal współczynniki rzetelności (α -Cronbacha) przyjmują akceptowalne wartości, pozwalając uznać skale za rzetelne. Analiza rzetelności nie daje więc podstaw do wyboru między modelem jednoczynnikowym a trójczynnikowym.

WERYFIKACJA MODELI POMIAROWYCH

W celu weryfikacji zaproponowanego modelu (wykr. 1) przeprowadzono konfirmacyjną analizę czynnikową. Pozwala ona również ocenić trafność skal stworzonych na podstawie wspomnianego modelu pomiarowego. Ponadto, porównano oszacowania uzyskane metodą największej wiarygodności (ML) i uogólnioną metodą najmniejszych kwadratów (GLS). Jak wspomniano, przy

poprawnej specyfikacji modelu obie metody powinny dawać zbliżone wyniki, zaś duże rozbieżności między wynikami uzyskanymi każdą z tych metod mogą wskazywać na błędy specyfikacji modelu.

Model jednoczynnikowy bez korelacji błędów pomiaru jest źle dopasowany — wartość statystyki chi-kwadrat jest wysoka¹⁹, średniokwadratowy błąd aproksymacji RMSEA²⁰ (ang. *root mean square error of approximation*) (Pleśniak, 2003) przekracza wartość 0,1, uznaną za dyskwalifikującą, zaś miary dopasowania nie osiągają akceptowalnej wartości 0,9, z kolei wartości bezwzględne reszt standaryzowanych przekraczały często 2, sugerując brak dopasowania macierzy kowariancji wynikającej z modelu i macierzy obliczonej na podstawie próby, N Hoeltera wyniosło 65 — znacznie poniżej wartości uznawanej za dopuszczalną, wyniki uzyskane metodą GLS również nie pozwalają uznać modelu za dopasowany (por. tabl. 3).

TABL. 3. MIARY DOPASOWANIA DLA MODELI PRZY ZASTOSOWANIU METODY ML I GLS

Statystyka	Jednoczynnikowy		Trójczynnikowy		Trójczynnikowy z kbp	
	ML	GLS	ML	GLS	ML	GLS
chi-kwadrat	612,898	423,736	193,95	171,06	46,53	44,104
(chi-kwadrat)/ss	22,7	15,7	9,7	8,6	2,9	2,8
n_R	15	25	4	6	4	5
RMR	0,670	0,913	0,302	0,360	0,278	0,325
GFI	0,869	0,904	0,958	0,961	0,990	0,990
AGFI	0,782	0,840	0,906	0,913	0,970	0,972
RMSEA	0,149	0,122	0,094	0,088	0,044	0,042
N Hoeltera	65	93	159	181	555	585

kbp — korelacje błędów pomiaru, chi-kwadrat — wartość statystyki w teście dopasowania modelu (por. przypis 7), ss — stopnie swobody, n_R — liczba reszt standaryzowanych, których wartość bezwzględna przekracza 2, RMR — odchylenie standardowe reszt, GFI — wskaźnik dopasowania modelu Jöreskoga i Sörboma, AGFI — skorygowany wskaźnik dopasowania modelu²¹, RMSEA — średniokwadratowy błąd aproksymacji. Miary dopasowania modelu opisane są m.in. w Bollen (1989), s. 257—279, Hoyle, 1995, rozdz. 5, Pleśniak (2003), s. 17—23.

Źródło: obliczenia przeprowadzono w pakiecie Amos.

¹⁹ Wysoka wartość statystyki chi-kwadrat implikuje wysokie wartości opartych o nią miar opartych na teorii informacji, takich jak: AIC (kryterium informacyjne Akaike’a), BIC (kryterium informacyjne Bayesa), BCC (kryterium Browna-Cudecka), ECVI (z ang. *expected cross-validation index* — wartość oczekiwana wskaźnika krosvalidacji) oraz niskie wartości absolutnych miar dopasowania. Miary oparte na teorii informacji są opisane w Pleśniak (2003), s. 22 i 23, ponadto ECVI w Bollen, Long (1993), s. 148—150.

²⁰ $RMSEA = \sqrt{\frac{F_{\min}(\mathcal{Z}(\theta), \mathcal{Z})}{ss}}$, gdzie: $F_{\min}(\mathcal{Z}(\theta), \mathcal{Z})$ — jest minimalną wartością funkcji

rozbieżności w populacji generalnej, której estymatorem jest $[(N-1)F(\mathcal{Z}(\hat{\theta}), \mathcal{Z}) - ss]/(N-1)$.

²¹ $AGFI = 1 - \frac{q(q+1)}{2 \cdot ss}(1 - GFI)$, gdzie dla ML: $GFI = 1 - \frac{tr[(\mathcal{Z}(\theta)^{-1}S - I)^2]}{tr[(\mathcal{Z}(\theta)^{-1}S)^2]}$, dla GLS:

$$GFI = 1 - \frac{tr[(I - \mathcal{Z}(\theta)S^{-1})^2]}{q}.$$

Ponadto znaczne rozbieżności między wynikami uzyskanymi metodą ML i GLS także świadczą na niekorzyść modelu jednoczynnikowego (w przypadku poprawnej specyfikacji modelu obie metody powinny dawać zbieżne wyniki — por. tabl. 3 i 4).

Model trójczynnikowy jest lepiej dopasowany niż jednoczynnikowy (wartość statystyki chi-kwadrat jest znacznie niższa. Implikuje to niższe wartości miar opartych na teorii informacji²²). Jednak nadal nie jest on dobrze dopasowany: średniokwadratowy błąd aproksymacji RMSEA 0,094 wynosi wprawdzie poniżej 0,1, ale nadal powyżej zalecanej wartości 0,05, niektóre miary dopasowania mają poniżej akceptowalnej wartości 0,9, wartości bezwzględne reszt standaryzowanych przekraczają 2, N Hoeltera wynosi poniżej zalecanej wielkości 200.

**TABL. 4. OCENY PARAMETRÓW DLA MODELI UZYSKANE METODĄ ML I GLS
(rozwiązanie standaryzowane)**

Parametr	Model jednoczynnikowy		Parametr	Model trójczynnikowy		Model trójczynnikowy z kbp	
	ML	GS		ML	GLS	ML	GLS
λ_1	0,40	0,45	λ_{11}	0,54	0,55	0,47	0,48
λ_2	0,45	0,55	λ_{41}	0,68	0,70	0,60	0,60
λ_3	0,64	0,70	λ_{61}	0,31	0,38	0,42	0,42
λ_4	0,53	0,60	λ_{81}	0,71	0,75	0,75	0,76
λ_5	0,72	0,72	λ_{91}	0,34	0,40	0,28	0,28
λ_6	0,70	0,75	λ_{32}	0,69	0,72	0,71	0,71
λ_7	0,83	0,85	λ_{52}	0,51	0,56	0,50	0,50
λ_8	0,54	0,63	λ_{72}	0,76	0,81	0,77	0,77
λ_9	0,56	0,61	λ_{23}	0,65	0,75	0,67	0,66
			λ_{53}	0,34	0,27	0,42	0,41
			λ_{63}	0,70	0,66	0,69	0,69
			λ_{73}	0,20	0,14*	0,30	0,30
			λ_{93}	0,46	0,47	0,53	0,53
			ϕ_{12}	0,76	0,75	0,80	0,79
			ϕ_{13}	0,19**	0,10***	0	0
			ϕ_{23}	0,44	0,44	0,20	0,20
			δ_{14}	0	0	0,21	0,20
			δ_{23}	0	0	0,36	0,34
			δ_{45}	0	0	0,15	0,15
			δ_{49}	0	0	0,21	0,21
			δ_{89}	0	0	0,23	0,23

kbp — korelacje błędów pomiaru, δ_{ij} — współczynnik korelacji błędów pomiaru e_i i e_j , * p = 0,004; ** p = 0,001; *** p = 0,041, dla pozostałych ocen parametrów $p < 0,001$, pozostałe oznaczenia jak na wyk. 1.

Źródło: obliczenia przeprowadzono w pakiecie Amos.

²² Liczebność próby i liczba zmiennych obserwowalnych jest taka sama w obu modelach, zaś liczba szacowanych parametrów różni się o 7 i ta różnica jest zbyt mała, by „zrównoważyć” wyższą o 426 wartość chi-kwadrat dla modelu jednoczynnikowego, pominięto więc wartości miar opartych na teorii informacji.

Dopasowanie modelu poprawia się po uwzględnieniu korelacji błędów pomiaru²³ (por. tabl. 6) i usunięciu nieistotnej ($p = 0,055$) korelacji między oceną ważności potrzeby zabezpieczenia własnej starości na tle potrzeb egzystencjalnych oraz na tle potrzeb związanych z rozwojem. Uzasadnienie uwzględnienia korelacji błędów zawiera tabl. 5.

TABL. 5. INDEKSY MODYFIKACJI I OCZEKIWANE ZMIANY WARTOŚCI PARAMETRÓW ZWIĄZANE Z UWZGLĘDNIENIEM KORELACJI BŁĘDÓW POMIARU

Skorelowane błędy pomiaru	MI	Per change	Uzasadnienie
e2 ↔ e3	50,613	1,448	korelacje te występowały również w roku 2004, wskazując na występowanie ukrytych czynników nieuwjętych w modelu
e1 ↔ e4	17,260	0,592	
e4 ↔ e5	15,845	0,516	
e4 ↔ e9	8,151	0,424	uprawianie sportu bywa postrzegane jako rozrywka, a jednocześnie pozwala utrzymać dobrą kondycję fizyczną, stąd potrzeby te mogą być postrzegane podobnie
e8 ↔ e9	5,276	0,363	potrzeby zapewnienia sobie wygodnych warunków życia oraz zapewnienia rozrywki i przyjemności wiążą się z hedonizmem, a przez to mogą być postrzegane podobnie

MI — indeks modyfikacji, mierzący poprawę modelu oczekiwanym spadkiem wartości statystyki chi-kwadrat, Per change — szacunek wartości parametru dodawanego do modelu.

Źródło: opracowanie własne na podstawie wyników uzyskanych w pakiecie Amos.

Zmodyfikowany model jest dobrze dopasowany²⁴ z punktu widzenia ogólnych miar dopasowania. Wskazują na to: średniokwadratowy błąd aproksymacji RMSEA poniżej wartości 0,05, N Hoeltera dla 0,05 równe 555 oraz miary dopasowania modelu powyżej 0,95. Analiza reszt wskazuje, że 4 kowariancje są niedoszacowane przez model.

Współczynniki determinacji dla poszczególnych stwierdzeń są akceptowalne i przyjmują wartości od 0,22 (w przypadku oceny ważności potrzeby zabezpieczenia własnej starości w porównaniu z potrzebą zdobywania pieniędzy i osiągnięcia wyższych dochodów) do 0,77 (w przypadku ważności potrzeby zabezpieczenia własnej starości na tle potrzeby poprawienia wyglądu zewnętrznego). Dodatkowo znaki wszystkich ładunków ścieżkowych są zgodne z oczekiwanymi. Warto dodać, że dodatkowo i wysokie ładunki ścieżkowe są warunkiem występowania wewnętrznej homogeniczności skal (por. Sagan, 2003).

²³ Korelacje błędów pomiaru dla poszczególnych par stwierdzeń mogą wynikać bądź z sąsiedztwa pytań w kwestionariuszu, bądź z faktu, że respondenci postrzegają je z jakichś powodów jako podobne, por. Bollen (1989). W tym ostatnim przypadku będą one świadczyły o istnieniu ukrytych czynników nieuwzględnionych w modelu (wykluczając oczywiście podobne brzmienie stwierdzeń).

²⁴ Należy zaznaczyć, że fakt, iż model jest dobrze dopasowany nie przesądza o jego prawdziwości, istnieje wiele alternatywnych modeli, które mogą mieć takie same dopasowanie.

Wnioski

W artykule przedstawiono niektóre metody estymacji parametrów modelu stosowane w modelowaniu równań strukturalnych oraz ich własności. Z uwagi na rozmaite własności metod estymacji, w różnych sytuacjach powinno wybierać się różne metody. Metody GLS i ML dostarczają asymptotycznie zbieżnych oszacowań (a w przypadku rozkładu normalnego — zbieżnych wyników również do ADF). Tak więc może się wydawać, że nie ma większego znaczenia, która z nich zostanie wybrana.

Należy mieć jednak świadomość, że w pewnych sytuacjach wspomniane metody będą prowadziły do odmiennych wyników (np. w przypadku błędnej specyfikacji modelu). Dla rozpatrywanego w przykładzie modelu jednoczynnikowego o znacznym stopniu uproszczenia, wyniki uzyskane metodą ML i GLS różniły się znacznie, nie tylko pod kątem oceny modelu, ale różne były też oceny parametrów. Rozbieżności między metodami mogą wskazywać na błędy specyfikacji. Z kolei dla modelu trójczynnikowego uwzględniającego korelacje błędów, który był najlepiej dopasowany, różnice między ML i GLS były niewielkie.

Analiza rzetelności nie pozwoliła wybrać między skalą jedno- a trójczynnikową, gdyż obie skale okazały się rzetelne. Z kolei weryfikacja modeli pomiarowych za pomocą procedury SEM dała podstawy wyboru modelu trójczynnikowego jako lepiej dopasowanego do zaobserwowanej macierzy kowariancji.

mgr Agnieszka Pleśniak — Szkoła Główna Handlowa

LITERATURA

- Bollen K. A. (1989), *Structural Equations with Latent Variables*, Wiley, New York
- Bollen K. A., Long J. S. (1993), *Testing Structural Equation Models*, Sage Publications, Inc, Newbury Park
- Boomsma A., Hoogland J. F. (2001), *The Robustness of LISREL Modeling Revisited*, www.ppsw.rug.nl/~boomsma/ssi.pdf
- Browne M. W. (1973), *Generalized Least Squares Estimators in the Analysis of Covariance Structures*, National Institute for Personnel Research (South Africa) and Educational Testing Service
- Curran P. J., Finch J. F., West S. G. (1996), *The Robustness of Test Statistics to Nonnormality and Specification Error in Confirmatory Factor Analysis*, Psychological Methods
- Górniak J. (2000), *My i nasze pieniądze*, Aureus, Kraków
- Grynberg G., Gilliam M. (1993), *Should We Take Don't Know for an Answer?*, Public Opinion Quarterly, vol. 57
- Hoyle R. H. (1995), *Structural Equation Modeling. Concepts, Issues, and Applications*, Sage Publications, Inc, Thousand Oaks
- Jöreskog K. G. (2002), *Structural Equation Modeling with Ordinal Variables Using LISREL*, <http://www.ssicentral.com/lisrel/techdocs/ordinal.pdf>
- Kaplan D. (2000), *Structural Equation Modeling. Foundations and Extensions*, Sage Publications, Inc, Thousand Oaks

- King G., Honaker J., Joseph A., Scheve K. (2001), *Analyzing Incomplete Political Science Data, An Alternative Algorithm for Multiple Imputation*, „American Political Science Review”, vol. 95
- Labovitz S. (1967), *Some Observations on Measurement and Statistics*, „Social Forces”, vol. 46
- Muthén B. (1984), *A general Structural Equation Model with Dichotomous, Ordered Categorical, and Continuous Latent Variable Indicators*, Psychometrika
- Nowak S. (2007), *Metodologia badań społecznych*, PWN, Warszawa
- Olsson U. (1979a), *Maximum Likelihood Estimation of The Polychoric Correlation Coefficient*, Psychometrika
- Olsson U. (1979b), *On the Robustness of Factor Analysis against Crude Classification of the Observations*, „Multivariate Behavioral Research”, vol. 14
- Olsson U. H., Troye S. V., Howell R. D. (1999), *Theoretic Fit and Empirical Fit, The Performance of Maximum Likelihood versus Generalized Least Squares Estimation in Structural Equation Models*, „Multivariate Behavioral Research”, vol. 34
- Olsson U. H., Foss T., Troye S. V., Howell R. D. (2000), *The Performance of ML, GLS, and WLS Estimation in Structural Equation Modeling Under Conditions of Misspecification and Non-normality*, „Structural Equation Modeling”, vol. 7
- Olsson U. H., Foss T., Troye S. V. (2003), *Does the ADF fit function decrease when the kurtosis increase?*, „British Journal of Mathematical and Statistical Psychology”, vol. 56
- Pleśniak A. (2003), *Modelowanie równań strukturalnych*, maszynopis, Instytut Statystyki i Demografii, SGH
- Pleśniak A. (2005), *Identyfikacja zależności korelacyjnej w procedurze SEM*, „Zeszyty Naukowe Akademii Ekonomicznej w Poznaniu”, nr 71
- Rudnicki L. (2000), *Zachowanie konsumentów na rynku*, PWE, Warszawa
- Sagan A. (2003), *Symbolika produktu w systemie komunikacji marketingowej*, Wydawnictwo Akademii Ekonomicznej w Krakowie, Kraków
- Yuan K. H., Bentler P. M. (1997), *Mean and Covariance Structure Analysis, Theoretical and Practical Improvements*, „Journal of the American Statistical Association”, vol. 92

SUMMARY

Some estimation models in the structural equation modeling as well as their characteristics are discussed in this article. An attempt of the scale construction to measure securing in households' old age with the use of structural equation modeling is presented too. Characteristics of each method decided on the estimation choice. Simulation survey results concerning robustness degree of each method on the deviation from the normal distribution, sample size, defective model specification etc. were taken into account too. Two measuring models were discussed: taking one common factor and three common factors. The factor confirmative analysis was conducted to verify the proposed models. Estimations made by the maximum likelihood (ML) method and by generalized least square estimator were compared, which shall give symptomatic correlated results assuming that the model specification is correct.