

**Piotr Tarka**

Uniwersytet Ekonomiczny w Poznaniu  
e-mail: piotr.tarka@ue.poznan.pl

---

## **WSKAŹNIKI WRMR I RMSEA ORAZ STATYSTYKA CHI-KWADRAT W OCENIE DOBROCI DOPASOWANIA MODELI SEM DLA DANYCH PORZĄDKOWYCH**

### **THE WRMR AND RMSEA INDICATORS AND CHI-SQUARE STATISTICS IN THE EVALUATION OF GOODNESS-OF-FIT IN SEM MODELS BASED ON ORDINAL DATA**

---

DOI: 10.15611/pn.2017.469.20  
JEL Classification: M31, C3, C90

**Streszczenie:** W niniejszym artykule autor rozważa wybrane problemy dotyczące oceny poziomu dobroci dopasowania modeli równań strukturalnych (SEM) w ujęciu zmiennych porządkowych kategoryalnych. Na ich tle zaprezentowano model strukturalny o tzw. niepełnej informacji, w zakresie którego przyjęto DWLS(M-V) jako metodę estymacji danych, a następnie w ramach analizy porównawczej poziomu dobroci dopasowania modelu SEM, zestawiono statystyki chi-kwadrat z wartościami wskaźników WRMR i RMSEA. Jako przykład ilustrujący strategię postępowania z danymi wyrażonymi na skali porządkowej wykorzystano wyniki badań empirycznych w kontekście postrzeganej przez badaczy i analityków w firmach skuteczności badań marketingowych na tle specyficznych uwarunkowań organizacyjnych.

**Słowa kluczowe:** modele równań strukturalnych (SEM), zmienne porządkowe, estymacja, dobroć dopasowania modeli SEM, wskaźniki WRMR i RMSEA.

**Summary:** In the article, the author discusses the problems of evaluation of goodness-of-fit in structural equation models (SEM) from the perspective of ordinal categorical variables. On their background a SEM model with limited information is considered and the estimation method DWLS(M-V) is diagnosed. Next two indices WRMR and RMSEA are set in comparison with statistic chi-square to observe the efficiency of the model fit. As an example illustrating the strategy for the appropriate dealing with data measured on ordinal scale, we refer to empirical data on researchers and analysts' perception (in companies) of marketing research effectiveness under the influence of specific organizational conditions.

**Keywords:** structural equation models (SEM), ordinal variables, estimation, goodness-of-fit in SEM, WLMR and RMSEA indices.

## 1. Wstęp

Zmienne dyskretne, tak jak skala Likerta, w praktyce badań społecznych i marketingowych pojawiają się nągminnie. Większość badaczy jest zgodna, iż zmienne tego typu, nawet przy najbardziej optymistycznych oczekiwaniach, odzwierciedlają na ogół *quasi* (przybliżone w swym kształcie) rozkłady normalne [Finney, DiStefano 2006; Almeida, Mouchart 2014]. W konsekwencji bezkrytyczna ich aplikacja w modelach równań strukturalnych, w sposób szczególnie uwzględniających klasyczne metody estymacji, jak ML (największej wiarygodności), mogłaby spowodować nie tylko błędną estymację parametrów w modelu, ale skutkowałaby również przyjęciem nieobiektywnych założeń w weryfikacji dobroci dopasowania modelu do zgromadzonych danych. Innymi słowy, ryzyko przyjęcia błędnych wniosków w ramach oceny poprawności postulowanego modelu SEM wzrasta za sprawą nieadekwatnych metod estymacji oraz podejść stosowanych w ocenie dobroci „jakości” jego dopasowania względem danych [Johnson, Creech 1983; Babakus, Ferguson, Jöreskog 1987; Cliff, Kyats 2005].

Wobec powyższego badacze mają do czynienia ze swoistego rodzaju paradoksem. Otóż zmienne porządkowe, pomimo iż wykazują wysoki poziom przydatności w pomiarze zjawisk społecznych, z powodu nieroztropności badaczy, na etapie złej strategii estymacji i oceny modelu, stają się często główną przyczyną jego odrzucenia. Z drugiej strony ten sam model w ramach implementacji odpornych metod estymacji danych i selekcji odpowiednich wskaźników oceny dobroci dopasowania może wykazać zupełną lub przynajmniej częściową adekwatność.

W niniejszym artykule autor, biorąc pod uwagę specyfikę pomiaru zmiennych porządkowych (ujętych w kategoriach odpowiedzi na skali Likerta), omawia na tle teoretycznym i praktycznym wybrane podejścia w analizie tego typu danych.

## 2. Problemy ze zmiennymi porządkowymi w modelach SEM

Zmienne porządkowe kategoryjne powodują w praktyce badań społecznych wiele ograniczeń w ramach przeprowadzanych analiz SEM. Jednym z nich są zaniżone wartości współczynników korelacji zmiennych obserwowalnych, które badacz wylicza na podstawie podejścia Pearsona. Jak wyjaśnia Konarski [2009, s. 484]: „osoba badana, która wybiera określoną kategorię odpowiedzi na skali porządkowej, może posiadać wyższy poziom cechy niż osoba, która wybiera niższą kategorię. Jednakże skala porządkowa nie niesie informacji o wielkości różnicy w natężeniu cechy między osobami, ponieważ zmienne porządkowe nie mają znaczących jednostek miary”. Tym samym w wielu przypadkach korelacje tego typu nie dostarczają statystyk wyczerpujących, przez co modele SEM generują niedoszacowane szacunki relacji w zakresie rozpatrywanych zmiennych. W sytuacji gdy badacz analizuje macierz korelacji za pośrednictwem np. estymatora największej wiarygodności (ML), otrzymuje on nieznacznie ujemnie obciążone szacunki parametrów modelu.

Z kolei błędy standardowe i wartości testu opartego na statystyce chi-kwadrat, które wykorzystuje się w ocenie dobroci dopasowania modelu SEM, są zwykle przeszacowane. Zatem traktowanie zmiennych porządkowych kategoryalnych w modelach SEM i przyjęcie wobec nich rozwiązań analitycznych zarezerwowanych najczęściej dla danych ciągłych, bez uprzedniego sprawdzenia własności samych zmiennych, prowadzi na ogół do wielu błędów, w tym złej interpretacji wyników modelu SEM [Lubke, Muthén 2004]<sup>1</sup>.

### 3. Metody estymacji zmiennych porządkowych w modelach SEM

Zmienne ukryte, które nie spełniają założeń normalności w swych rozkładach, zmuszają badacza do wdrożenia alternatywnych, tj. odpornych metod estymacji [Finney, DiStefano 2006], które nakładają korekty na wartości statystyki chi-kwadrat oraz pozwalają skorygować wielkości błędów standardowych w modelu SEM. Podstawy teoretyczne dotyczące konstrukcji tego typu metod w zakresie modeli ze zmiennymi ukrytymi (rozpatrywanymi w kontekście zmiennych dyskretnych), odnajdujemy w pracy Muthéna [1983] i Jöreskoga [1990, 1994]. Według nich modele SEM (bazujące na danych porządkowych), należy przede wszystkim analizować w kontekście ograniczonych zasobów informacyjnych, gdyż wyniki generowane są na podstawie tzw. parametrów progów, obliczanych osobno dla każdej zmiennej obserwowalnej, stanowiącej określony wkład informacyjny w zmienną ukrytą. Rolą badacza jest odnalezienie informacji zawartych w parametrach progów, które stanowią punkty na kontinuum zmiennej ukrytej i które dzielą dyskretne poziomy skali porządkowej zmiennej obserwowalnej.

Badacz, dysponując  $r$  i  $s$  kategoriami określonych par zmiennych obserwowalnych, w tabeli kontyngencji, po wyznaczeniu odpowiedniej liczby progów, wyznacza następnie poziom prawdopodobieństwa  $\pi_{ij}$ , iż określona obserwacja znajdzie się w danej komórce tabeli kontyngencji. W ten sposób otrzymuje logarytm funkcji wiarygodności [Olsson 1979]:

$$\log L(n_{ij} | \pi_{ij}) = c \sum_{i=1}^s \sum_{j=1}^r n_{ij} \log \pi_{ij}, \quad (1.0)$$

gdzie:  $n_{ij}$  – to liczba obserwacji w komórce  $ij$  danej tabeli kontyngencji.

Maksymalizacja wyrażenia (1.0) prowadzi do uzyskania estymatora ML w oparciu o korelacje polichoryczne.

---

<sup>1</sup> Analiza taka jest konieczna w sytuacji, gdy: 1) kategorie odpowiedzi zmiennych obserwowalnych wykorzystują jedynie skalę 4 pkt, 2) związek pomiędzy zmienną obserwowalną i zmienną ukrytą jest nieliniowy, 3) zmienna ukryta odznacza się mocną asymetrią, wobec czego kluczowego znaczenia nabiera weryfikacja rozkładów jednowymiarowych (zmiennych obserwowalnych) oraz analiza rozkładu wielowymiarowego zmiennej ukrytej.

Ostatecznie po wygenerowaniu korelacji i ich asymptotycznie oszacowanych kowariancji badacz przystępuje do implementacji funkcji WLS [Browne 1984]:

$$F_{\text{WLS}_c} = [\mathbf{r} - \boldsymbol{\rho}(\boldsymbol{\theta})]^T \mathbf{W}^{-1} [\mathbf{r} - \boldsymbol{\rho}(\boldsymbol{\theta})], \quad (1.1)$$

gdzie:  $\mathbf{r}$  to wektor kolumnowy elementów macierzy korelacji polichorycznych,  $\boldsymbol{\rho}(\boldsymbol{\theta})$  – wektor elementów w implikowanej macierzy kowariancji,  $\mathbf{W}^{-1}$  – macierz wag jako odwrotność oszacowanej asymptotycznej macierzy kowariancji wektora  $\mathbf{r}$ .

Zwróćmy uwagę na fakt, iż w miarę jak powiększa się w modelu liczba zmiennych obserwowalnych, automatycznie rośnie też rozmiar macierzy wag i liczba nieredundantnych elementów. W rezultacie rośnie też wymóg posiadania dużych prób w ramach przeprowadzanych obliczeń. Dlatego też Muthén, DuToit i Spisic [1997] zaproponowali zmodyfikowaną wersję estymatora (1.1) pod nazwą diagonalnie ważonych najmniejszych kwadratów (DWLS<sub>c</sub>):

$$F_{\text{DWLS}_c} = [\mathbf{r} - \boldsymbol{\rho}(\boldsymbol{\theta})]^T \mathbf{W}_D^{-1} [\mathbf{r} - \boldsymbol{\rho}(\boldsymbol{\theta})], \quad (1.2)$$

gdzie: poszczególne składniki (1.2) są tak samo definiowane jak w rozwiązaniu (1.1), z tym wyjątkiem, że macierz wag  $\mathbf{W}_D^{-1}$  zawiera jedynie odwrotności diagonalnych elementów pełnej macierzy w równaniu (1.1). W rezultacie unika się problemów z wyliczaniem odwrotności macierzy  $\mathbf{W}$ , dzięki czemu maleje liczba wolnych parametrów w macierzy  $\mathbf{W}_D^{-1}$ .

Niestety, podejście (1.2) prowadzi do redukcji statystycznej efektywności estymatora, która z kolei prowadzi do obciążonych oszacowań błędów standardowych i obciążonych statystyk testowych. Dlatego też Muthén zaproponował pewne rozwiązanie zastępcze, wprowadzając do pakietu komputerowego Mplus [Muthén, Muthén 2014] korekty błędów standardowych oraz odporną statystykę chi-kwadrat dopasowania modelu w postaci skorygowanej średniej  $T_M$  oraz wersji alternatywnej [Satorra, Bentler 1994] – statystyki skorygowanej średnią i wariancją  $T_{MV}$ .

#### 4. Strategie oceny dobroci dopasowania modeli SEM ze zmiennymi porządkowymi

W literaturze przedmiotu liczne ograniczenia statystyki testowej chi-kwadrat doprowadziły do powstania różnych miar opisowych dopasowania modelu SEM, wśród których wyróżnia się miary akcentujące: prostotę modelu, możliwości porównywania modelu zerowego z modelem postulowanym, oraz prawdopodobieństwo walidacji modelu w niezależnych próbach danych. Ta różnorodność doprowadziła do powstania czterech klas wskaźników z rodziny miar: absolutnych (GFI, AGFI, CN, RM, SRMR, WLMR i RMSEA), porównawczych (NFI, NNFI, RFI, IFI, CFI), walidacji krzyżowej (ECVI, AIC, BIC, CAIC) i oszczędnych (PNFI, PGFI,

PCFI). Niestety, większość z nich opiera się na wartościach statystyki chi-kwadrat, której głównym mankamentem (nawet przy występowaniu danych ciągłych) jest wysoki poziom wrażliwości na [Monroe, Cai 2015]: 1) zbyt duże lub małe liczebności prób badawczych, 2) błędną specyfikację konfiguracji zmiennych w modelu, 3) zbyt dużą liczbę przyjętych zmiennych obserwowalnych przypadających na określone zmienne ukryte oraz 4) naruszenie założeń normalności rozkładów wielowymiarowych zmiennych ukrytych [Cudeck, Henly 1991]. Co ciekawe, nawet dobre dopasowanie modelu w oparciu o statystykę chi-kwadrat nie dowodzi jeszcze jego prawdziwości, lecz pozwala jedynie na jego nieodrzućenie. Dodajmy, że kierowanie się w interpretacji modelu wyłącznie samą statystyką, nawet po nałożeniu na nią korekt ( $T_M$  i  $T_{MV}$ ), także może skutkować dalszymi błędami, ponieważ inne miary mogą zakwestionować jej wynik. Głównym problemem różnorodnych miar dopasowania jest to, że są one, po pierwsze, opisowe, zaś ich rozkłady teoretyczne nie są znane, a po drugie miary nie mają porównywalnych skal i w efekcie model postulowany jako adekwatny przez jedną miarę może być odrzućony w sytuacji aplikacji innej klasy miar dopasowania [Konarski 2009].

Zdaniem autora, przewyżczanie powyższych ograniczeń może jednak dokonywać się w drodze aplikacji wskaźników z klasy miar absolutnych. Jednym z nich jest indeks RMSEA (pierwiastek średniego kwadratu błędu aproksymacji) dopuszczający punktową i jednocześnie przedziałową estymację wyniku w zakresie rozbieżności spowodowanej aproksymacją modelu lub błędem modelu [Steiger 1990]. Ponieważ indeks RMSEA ma wbudowaną korektę na złożoność modelu SEM, stąd jest on niezwykle przydatny na etapie selekcji właściwego modelu spośród różnych wersji alternatywnych. Co więcej, pozwala on przeprowadzić test w zakresie bardziej realistycznej hipotezy przybliżonego dopasowania modelu SEM.

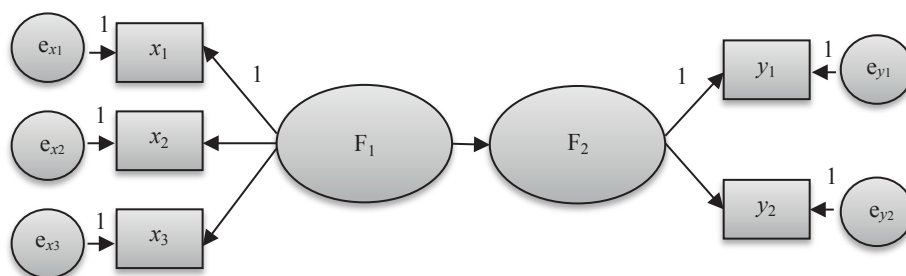
Wartość wskaźnika RMSEA jako funkcji populacyjnej macierzy kowariancji nie jest wprawdzie znana, lecz może być oszacowana zgodnie z formułą zaproponowaną przez Browne'a i Cudecka [1993]. Autorzy ci podali jednocześnie formułę na 90-procentowy przedział ufności dla RMSEA, który pozwala określić poziom precyzji szacunku punktowego. Tym samym przy interpretacji indeksu, nawet jeśli odrzuća się postulowany model, to ważna staje się wiedza w zakresie przedziałów (tj. dolnego i górnego limitu) tego odrzućenia. W sumie dobre dopasowanie modelu wyznacza wartość  $RMSEA \leq 0,05$ .

Kolejny wskaźnik z rodziny miar absolutnych to wystandaryzowany średni pierwiastek kwadratowy reszt – SRMR, który zakłada ocenę ogólnego poziomu dopasowania modelu SEM (od 0 do 1) pod kątem powstałych reszt w wyniku zestawienia macierzy empirycznej do odtworzonej. Ponieważ sam indeks odzwierciedla tzw. poziom „mierności dopasowania” modelu, dlatego też przy jego interpretacji przyjmuje się, iż wartości zerowe świadczą o idealnym dopasowaniu, a wyższe wskazują na gorsze dopasowanie, przy czym jak słusznie zauważyli Hu i Bentler [1999], w praktyce badań społecznych poziom 0,08 dowodzi już adekwatności rozpatrywanego modelu. Warto też zaznaczyć, że istotną przesłanką

stosowania wskaźnika SRMR jest jego prostota w interpretacji wyniku, który wyraża bezwzględną wielkość średniej reszty w jednostkach współczynnika korelacji. Jednakże zdaniem B. Muthéna i L. Muthén [2014] znacznie lepszym odpowiednikiem SRMR, z uwagi na dyskretny charakter zmiennych, jest indeks średnich ważonych reszt kwadratowych (WRMR), sprawdzający się w modelach, wobec których zachodzi podejrzenie, iż statystyki zmiennych obserwowalnych (ujętych w modelu) będą odznaczać się nieporównywalnymi poziomami wariancji bądź też będą generować rozkłady skrajnie odstające od założeń normalności. Jak podają Yu i Muthén [2002], dobre dopasowanie modeli SEM opartych na danych porządkowych uzyskuje się poniżej poziomu 1,0.

## 5. Porównanie wskaźników WRMR i RMSEA ze statystyką chi-kwadrat

Biorąc pod uwagę powyższe założenia, zweryfikujemy zachowanie wskaźników WRMR i RMSEA, a następnie zestawimy ich wyniki z wartościami statystyki dopasowania modelu chi-kwadrat skorygowanymi: 1) średnią oraz 2) średnią i wariancją. W ramach prezentacji podejścia obrazującego poziom dobroci dopasowania modelu SEM wykorzystamy przykład z obszaru badań marketingowych. Sam model, który poddano weryfikacji, przedstawia rys. 1. W jego zakresie uwzględniono dwie zmienne ukryte oraz pięć zmiennych obserwowalnych (tab. 1 i 2). Zadaniem modelu było określenie wpływu specyfiki kulturowej oraz postaw wyrażanych przez kadre



Legenda: Elipsa  $F_1$  definiuje zmienną ukrytą niezależną „uogólnione podejście firm do badań marketingowych”; elipsa  $F_2$  oznacza zmienną ukrytą zależną „podejście firm do wykorzystania informacji i podejmowania decyzji rynkowych”; kwadraty  $x_1, x_2, x_3$  to zmienne obserwowalne zmiennej ukrytej niezależnej  $F_1$ ; zaś:  $y_1, y_2$  to wskaźniki zmiennej niezależnej  $F_2$ ; wariancje błędów pomiaru  $e_{x1}, e_{x2}, e_{x3}$ , oraz  $e_{y1}, e_{y2}$  zmiennych obserwowalnych; ładunki czynnikowe znajdują się na strzałkach określających relacje między zmiennymi ukrytymi ( $F_1$  i  $F_2$ ) a zmiennymi obserwowalnymi.

**Rys. 1.** Model strukturalny (SEM) diagnozujący związek przyczynowo-skutkowy

Źródło: opracowanie na podstawie badań empirycznych [Tarka 2017].

w firmach wobec badań marketingowych z uwagi na dalsze podejście do wykorzystania informacji w podejmowanych decyzjach i planowanych operacjach rynkowych. Dane pochodzą z badania empirycznego, które przeprowadzono od marca do sierpnia 2014 r. na próbie ekspertów zatrudnionych w firmach na stanowisku „badacza marketingowego” i „analityka” [Tarka 2017].

**Tabela 1.** Opis wskaźników przyjętych w części pomiarowej modelu SEM

Wpływ $F_1$ i $F_2$ na $x_1, x_2, x_3, y_1, y_2$		Pytanie. <i>Od czego zależy Pana/i zdaniem organizacyjny poziom skuteczności badań marketingowych w firmie?</i>	
$F_1$	→	$x_1$	przyjętych w firmie właściwych postaw wobec badań i informacji marketingowych
	→	$x_2$	istnienia odpowiedniej struktury i kultury organizacyjnej w firmie
	→	$x_3$	przyjęcia określonej orientacji rynkowej przez firmę
$F_1$	→	$y_1$	sprawnego stylu podejmowania decyzji przez kadrę w firmie
	→	$y_2$	właściwego wykorzystania informacji przez kadrę w firmie do działań rynkowych

Legenda: Zmienne obserwowalne mierzono na 7-punktowej skali Likerta, gdzie krańcowe wartości wynosiły odpowiednio: 1 – „zdecydowanie się nie zgadzam”, 7 – „zdecydowanie się zgadzam”.

Źródło: opracowanie na podstawie badań empirycznych [Tarka 2017].

**Tabela 2.** Wstępna weryfikacja zmiennych obserwowalnych na podstawie wybranych statystyk opisowych

Statystyki wskaźników	Kategorie odpowiedzi na przyjętej skali							Skośność /kurtoza
	1	2	3	4	5	6	7	
$x_1$ – liczebność	14	14	67	119	126	(–)	(–)	–0,952/0,546
Proporcje	0,041	0,041	0,197	0,350	0,371	(–)	(–)	
$x_2$ – liczebność	7	14	32	64	118	105	(–)	–0,906/0,443
Proporcje	0,021	0,041	0,094	0,188	0,347	0,309	(–)	
$x_3$ – liczebność	7	53	83	98	99	(–)	(–)	–0,398/0,887
Proporcje	0,021	0,156	0,244	0,288	0,291	(–)	(–)	
$y_1$ – liczebność	7	14	39	56	133	91	(–)	–0,980/0,500
Proporcje	0,021	0,041	0,115	0,165	0,391	0,268	(–)	
$y_2$ – liczebność	7	53	83	98	99	(–)	(–)	–1,430/2,442
Proporcje	0,021	0,156	0,244	0,288	0,291	(–)	(–)	

Legenda: oznaczenie kategorii odpowiedzi jak w tab. 2; (–) wartość nie występuje ze względu na braki odpowiedzi.

Źródło: opracowanie własne w programie Mplus (wersja 7.4).

W pierwszej kolejności poddano weryfikacji rozkłady zmiennych obserwowalnych (tab. 2). Na ich podstawie można stwierdzić, iż kategoria z nr 7 była spora-



dycznie typowana przez respondentów w badaniu ankietowym. Natomiast statystyki opisowe zmiennych (por. wartości według miar skośności i kurtozy) sugerują asymetrię lewostronną. Wartości te odbiegają od założeń rozkładów normalnych. Tym samym wstępna obserwacja wyników uprawnia nas do przyjęcia strategii estymacji modelu SEM w oparciu o formułę (1.2), którą uzupełnimy dwoma rodzajami korekt:  $T_M$  i  $T_{MV}$ . Analizę przeprowadzono w programie Mplus (wersja 7.4 obejmująca moduł bazowy). W konsekwencji do modelu SEM wprowadzono korekty w zakresie błędów standardowych i statystyki chi-kwadrat z uwzględnieniem dwóch alternatywnych rozwiązań  $T_M$  i  $T_{MV}$ . Wyniki estymacji przedstawiają tabele 3-4.

Porównując obie wersje statystyk chi-kwadrat, zauważamy, iż pierwsza z nich,  $T_M = 40,805$  (w świetle korekty obejmującej wyłącznie średnią), uzyskała większą wartość w porównaniu z wartością statystyki  $T_{MV} = 36,588$ , na którą nałożono korektę w postaci średniej i wariancji. Jednakże sam zabieg nie doprowadził do radykalnej poprawy jakości dopasowania modelu. Model nadal nie uzyskał wystarczającego poziomu istotności dobroci dopasowania, co *de facto* potwierdza także inny wskaźnik RMSEA, który w pierwszym podejściu w ramach opcji  $T_M$ , wyniósł 0,165 (z 90-proc. przedziałem ufności 0,121-0,212) oraz 0,155 (z przedziałem 0,111-0,203) – w ramach opcji  $T_{MV}$ . W tym wypadku przedział ufności dla wyniku RMSEA potwierdza, iż wartość błędu aproksymacji modelu nie mieści się w granicach przyjętych dla bliskiego dopasowania.

**Tabela 3.** Statystyki dopasowania modelu równań strukturalnych (SEM)

Statystyki	Metody estymacji	
	$T_M$	$T_{MV}$
Model SEM postulowany: $T(df); p$	40,805 (4); 0,000	36,588 (4); 0,000
Model zerowy: $T(df); p$	2824,087 (10); 0,000	1966,421 (4); 0,000
RMSEA	0,165 przy 90% przedziale: (0,121-0,212)	0,155 przy 90% przedziale: (0,111-0,203)
WRMR	0,700	0,700

Legenda: Statystyka  $T_M$  korygowana wyłącznie średnią;  $T_{MV}$  korygowana średnią i wariancją według podejścia estymacji DWLS<sub>c</sub>.

Źródło: opracowanie własne w programie Mplus (wersja 7.4).

Ponieważ wartości RMSEA w obu przypadkach ( $T_M$  i  $T_{MV}$ ) przekraczają poziom  $\geq 0,05$ , który obrazuje poziom akceptacji (0,05) hipotezy bliskiego dopasowania modelu z poziomem istotności testu hipotezy bliskiego dopasowania na poziomie  $p > 0,5$ ), stąd moglibyśmy założyć, iż postulowany model SEM zupełnie odbiega od postulowanych założeń teoretycznych. Innymi słowy, taki model należałoby odrzucić. Jednakże, gdy przyjrzymy się kolejnemu wskaźnikowi WRMR, to zauważymy, iż wartości tego indeksu znajdują się poniżej poziomu 1,0, co oznacza, że model



**Tabela 4.** Estymowane wartości parametrów w modelu równań strukturalnych (SEM)

Kierunek wpływu $F_1$ i $F_2$ na $x_1, x_2, x_3, y_1, y_2$		Szacunek (E)	Błąd Standardowy (SE)	E/SE	Poziom istotności
$F_1$	→ $x_1$	1,000 (0,742)	(-); (0,029)	(-); (25,877)	(-)
	→ $x_2$	1,110 (0,824)	0,056 (0,024)	19,925 (34,848)	0,000
	→ $x_3$	0,894 (0,664)	0,046 (0,029)	19,443 (22,819)	0,000
$F_2$	→ $y_1$	1,000 (0,813)	(-); (0,030)	(-); (26,899)	(-)
	→ $y_2$	0,998 (0,812)	0,066 (0,030)	15,219 (25,517)	0,000
$F_1$	→ $F_2$	0,866 (0,790)	0,048 (0,030)	18,017 (25,988)	0,000
$F_1^*$		0,551 (1,000)	0,043 (0,000)	12,938, (-)	0,000
$F_2^\circ$		0,249 (0,376)	0,038 (0,048)	6,481 (7,826)	0,000

Legenda: W nawiasach podano wartości standaryzowane. Oznaczenie (-) opisuje wartość stałą parametru na poziomie 1 w ramach identyfikacji zmiennej ukrytej.  $F_1^*$  – poziom zmienności;  $F_2^\circ$  – poziom zmienności, tzw. błąd strukturalny.

Źródło: opracowanie własne w programie Mplus (wersja 7.4).

można zaakceptować, zwłaszcza jeśli weźmiemy pod uwagę specyfikę rozkładów danych. Postulowany model nie jest wprawdzie najlepszy, ale też nie jest najgorszy. Na jego korzyść przemawiają dodatkowo estymowane wartości parametrów (tab. 4), które uzyskały na wszystkich ścieżkach w modelu wysoki poziom istotności.

## 6. Wnioski

Wobec powyższych faktów stwierdzamy, iż kierowanie się w ocenie modeli SEM wyłącznie statystyką chi-kwadrat może prowadzić badacza do odrzucenia założeń teoretycznych w postulowanym modelu, który w rzeczywistości, za sprawą innego rodzaju wskaźników, może okazać się poprawny. Przykładowo, wskaźnik WRMR sprawdza się najlepiej w tych modelach SEM, w których statystyki wykazują niestabilny poziom zmienności rozpatrywanych zmiennych bądź gdy zmienne te przyjmują charakter porządkowy. Wskaźnik ten wypada również dobrze w małych próbach badawczych z uwzględnieniem zmiennych wykazujących skrajne odchylenia od rozkładów normalnych. W przeciwieństwie do niego wskaźnik RMSEA jest mniej odporny na zaburzenia normalności, zaś wielkości jego błędów uwarunkowane są liczebnościami prób badawczych ( $N$ ) i liczbą stopni swobody ( $df$ ). W konsekwencji modele o stosunkowo niskim poziomie  $df$  i  $N$  sztucznie zawyżają wartości RMSEA. Warto o tym pamiętać, zanim jeszcze podejmie się decyzję o odrzuceniu modelu SEM bądź przystąpi do jego radykalnej modyfikacji (np. w ramach nakładania ograniczeń na parametry). Nie ulega też najmniejszej wątpliwości, iż w przypadku danych porządkowych kategoryalnych rozwiązaniem znacznie bezpieczniejszym jest wykorzystanie w obliczeniach metod opartych na odpornych procedurach estymacji

(DWLS<sub>c</sub>). Bezskrytyczne bowiem przyjęcie przez badacza estymatora największej wiarygodności (ML) mogłoby doprowadzić nie tylko do zniekształcenia struktury czynnikowej modelu pomiarowego (CFA), ale w ostateczności do zafalszowania wyników otrzymanych na podstawie całego modelu strukturalnego (SEM). Wiemy już, że efektem naruszeń przez zmienne założeń normalności w swych rozkładach jest zniekształcenie wartości statystyki testowej chi-kwadrat, na podstawie której wnioskuje się o odrzuceniu bądź przyjęciu modelu. W rezultacie badacz może popełnić błąd polegający na odrzuceniu modelu SEM, który w rzeczywistości nie jest zły, bądź może on popełnić błąd, w świetle którego nie odrzuca modelu SEM wykazującego błędne podstawy założeń teoretycznych.

## Literatura

- Almeida C., Mouchart M., 2014, *Testing normality of latent variables in the polychoric correlation*, Statistica, LXXIV, nr 1, s. 1-25.
- Babakus E., Ferguson C.E., Jöreskog K.G., 1987, *The sensitivity of confirmatory maximum likelihood factor analysis to violations of measurement scale and distributional assumptions*, Journal of Marketing Research, vol. 24, s. 222-229.
- Browne M.W., 1984, *Asymptotically distribution-free methods for the analysis of covariance structures*, British Journal of Mathematical and Statistical Psychology, vol. 37, s. 62-83.
- Browne M.W., Cudeck R., 1993, *Alternative ways of assessing model fit*, [w:] Bollen K.A., Long J.S. (red.), *Testing Structural Equation Models*, Sage, Newbury Park, s. 136-162.
- Cliff N., Kyats J.A., 2005, *Ordinal measurement in the behavioural sciences*, Lawrence Erlbaum, New Jersey.
- Cudeck R., Henly S.J., 1991, *Model selection in covariance structures analysis and the "problem" of sample size: a clarification*, Psychological Bulletin, vol. 109, s. 512-519.
- Finney S.J., DiStefano C., 2006, *Non-normal and categorical data in structural equation modelling*, [w:] Hancock G.R., Mueller R.O. (red.), *Structural Equation Modelling: a second course*, Information Age Publishing, Greenwich, s. 269-314.
- Hu L.-T., Bentler P.M., 1999, *Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives*, Structural Equation Modeling, vol. 6, s. 1-55.
- Johnson D.R., Creech J.C., 1983, *Ordinal measures in multiple indicator models: A simulation study of categorization error*, American Sociological Review, vol. 48, s. 398-407.
- Jöreskog K.G., 1990, *New developments in LISREL: analysis of ordinal variables using polychoric correlations and weighted least squares*, Quality & Quantity, vol. 24, s. 387-404.
- Jöreskog K.G., 1994, *On the estimation of polychoric correlations and their asymptotic covariance matrix*, Psychometrika, vol. 59, s. 381-390.
- Konarski R., 2009, *Modele równań strukturalnych*, PWN, Warszawa.
- Lubke G.H., Muthén B.O., 2004, *Applying multi-group confirmatory factor models for continuous outcomes to Likert scale data complicates meaningful group comparisons*, Structural Equation Modeling, vol. 11, s. 514-534.
- Monroe S., Cai L., 2015, *Evaluating structural equation models for categorical outcomes: a new test statistic and practical challenge of interpretation*, Multivariate Behavioral Research, vol. 50, s. 569-583.
- Muthén B., 1983, *Latent variable structural equation modeling with categorical data*, Journal of Econometrics, vol. 22, s. 43-65.

- Muthén B., 1993, *Goodness of fit with categorical and other non normal variables*, [w:] Bollen K.A., Long J.S. (red.), *Testing structural equation models*, Sage, Newbury Park, s. 205-234.
- Muthén B.O., DuToit S.H., Spisic D., 1997, *Robust inference using weighted least squares and quadratic estimation equations in latent variable modeling with categorical and continuous outcomes*, <http://www.statmodel.com/bmuthen/categorical.htm>.
- Muthén L.K., Muthén B.O., 2014, *Mplus user's guide*, Muthén and Muthén, Los Angeles.
- Olsson U., 1979, *Maximum likelihood estimation of the polychoric correlation coefficient*, *Psychometrika*, vol. 44, s. 443-460.
- Satorra A., Bentler P.M., 1994, *Corrections to test statistics and standard errors in covariance structure analysis*, [w:] von. Eye A., Clogg C.C. (red.), *Latent Variables Analysis: Applications for Development Research*, Sage Publications, Thousand Oaks, s. 399-419.
- Steiger J.H., 1990, *Structural model evaluation and model modification: an interval estimation approach*, *Multivariate Behavioral Research*, vol. 25, s. 173-180.
- Tarka P., 2017, *Uwarunkowania skuteczności badań marketingowych*, Wydawnictwo Uniwersytetu Ekonomicznego w Poznaniu.
- Yu C.-Y., Muthén B., 2002, *Evaluation of model fit indices for latent variable models with categorical and continuous outcomes*, report, <https://www.statmodel.com/download/Yudissertation.pdf>.
- Yuan K.-H., Bentler P.M., 1998, *Normal theory based test statistics in structural equation modeling*, *British Journal of Mathematical and Statistical Psychology*, vol. 51, s. 289-309.