

PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

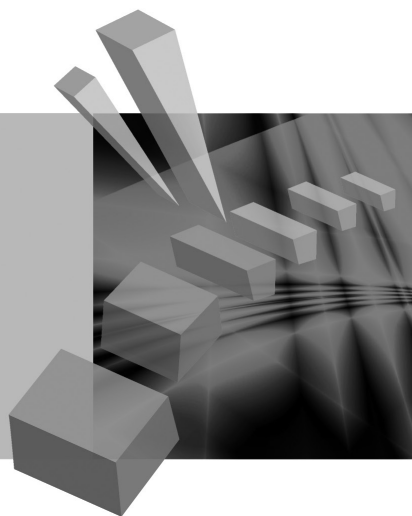
RESEARCH PAPERS

of Wrocław University of Economics

279

Taksonomia 21

Klasyfikacja i analiza danych – teoria i zastosowania



Redaktorzy naukowi

Krzysztof Jajuga

Marek Walesiak



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2013

Redaktor Wydawnictwa: Aleksandra Śliwka

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Małgorzata Czupryńska

Projekt okładki: Beata Dębska

Publikacja jest dostępna w Internecie na stronach:

www.ibuk.pl, www.ebscohost.com,

The Central and Eastern European Online Library www.ceeol.com,

a także w adnotowanej bibliografii zagadnień ekonomicznych BazEkon

http://kangur.uek.krakow.pl/bazy_ae/bazekon/nowy/index.php

Informacje o naborze artykułów i zasadach recenzowania znajdują się

na stronie internetowej Wydawnictwa

www.wydawnictwo.ue.wroc.pl

Tytuł dofinansowany ze środków Narodowego Banku Polskiego

oraz ze środków Sekcji Klasyfikacji i Analizy danych PTS

Kopiowanie i powielanie w jakiegokolwiek formie

wymaga pisemnej zgody Wydawcy

© Copyright by Uniwersytet Ekonomiczny we Wrocławiu

Wrocław 2013

ISSN 1899-3192 (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)

ISSN 1505-9332 (Taksonomia)

Wersja pierwotna: publikacja drukowana

Druk: Drukarnia TOTEM

Spis treści

Wstęp	9
Sabina Denkowska, Kamil Fijorek, Marcin Salamaga, Andrzej Sokolowski: Sejm VI kadencji – maszynka do głosowania	11
Barbara Pawelek, Adam Sagan: Zmienne ukryte w modelach ekonomicznych – respecyfikacja modelu Kleina I	19
Jan Paradysz: Nowe możliwości badania koniunktury na rynku pracy	29
Krzysztof Najman: Samouczące się sieci GNG w grupowaniu dynamicznym zbiorów o wysokim wymiarze	41
Kamila Migdał-Najman: Zastosowanie jednowymiarowej sieci SOM do wyboru cech zmiennych w grupowaniu dynamicznym	48
Aleksandra Matuszewska-Janica, Dorota Witkowska: Zróżnicowanie płac ze względu na płeć: zastosowanie drzew klasyfikacyjnych	58
Iwona Foryś, Ewa Putek-Szeląg: Przestrzenna klasyfikacja gmin ze względu na sprzedaż użytków gruntowych zbywanych przez ANR w województwie zachodniopomorskim	67
Joanna Banaś, Małgorzata Machowska-Szewczyk: Klasyfikacja internetowych rachunków bankowych z uwzględnieniem zmiennych symbolicznych.....	77
Marta Jaročka: Wpływ metody doboru cech diagnostycznych na wynik porządkowania liniowego na przykładzie rankingu polskich uczelni	85
Anna Zamojska: Badanie zgodności rankingów wyznaczonych według różnych wskaźników efektywności zarządzania portfelem na przykładzie funduszy inwestycyjnych.....	95
Dorota Rozmus: Porównanie dokładności taksonomicznej metody propagacji podobieństwa oraz zagregowanych algorytmów taksonomicznych opartych na idei metody <i>bagging</i>	106
Ewa Wędrowska: Wrażliwość miar dywergencji jako mierników niepodobieństwa struktur.....	115
Katarzyna Wójcik, Janusz Tuchowski: Wpływ automatycznego tłumaczenia na wyniki automatycznej identyfikacji charakteru opinii konsumenckich ...	124
Małgorzata Misztal: Ocena wpływu wybranych metod imputacji na wyniki klasyfikacji obiektów w modelach drzew klasyfikacyjnych.....	135
Anna Czapkiewicz, Beata Basiura: Badanie wpływu wyboru współczynnika zależności na grupowanie szeregów czasowych	146
Tomasz Szubert: Czynniki różnicujące poziom zadowolenia z życia oraz wartości życiowe osób sprawnych i niepełnosprawnych w świetle badań „Diagnozy społecznej”	154

Marcin Szymkowiak: Konstrukcja estymatorów kalibracyjnych wartości globalnej dla różnych funkcji odległości	164
Wojciech Roszka: Szacowanie łącznych charakterystyk cech nieobserwowanych łącznie	174
Justyna Brzezińska: Metody wizualizacji danych jakościowych w programie R	182
Agata Sielska: Regionalne zróżnicowanie potencjału konkurencyjnego polskich gospodarstw rolnych w województwach po akcesji do Unii Europejskiej	191
Mariusz Kubus: Liniowy model prawdopodobieństwa z regularyzacją jako metoda doboru zmiennych	201
Beata Basiura: Metoda Warda w zastosowaniu klasyfikacji województw Polski z różnymi miarami odległości	209
Katarzyna Wardzińska: Wykorzystanie metody obwiedni danych w procesie klasyfikacji przedsiębiorstw	217
Katarzyna Dębowska: Modelowanie upadłości przedsiębiorstw oparte na próbach niezbilansowanych	226
Danuta Tarka: Wpływ metody doboru cech diagnostycznych na wyniki klasyfikacji obiektów na przykładzie danych dotyczących ochrony środowiska ..	235
Artur Czech: Zastosowanie wybranych metod doboru zmiennych diagnostycznych w badaniach konsumpcji w ujęciu pośrednim	246
Beata Bal-Domańska: Ocena relacji zachodzących między inteligentnym rozwojem a spójnością ekonomiczną w wymiarze regionalnym z wykorzystaniem modeli panelowych	255
Mariola Chrzanowska: <i>Ordinary kriging</i> i <i>inverse distance weighting</i> jako metody szacowania cen nieruchomości na przykładzie warszawskiego rynku	264
Adam Depta: Zastosowanie analizy wariancji w badaniu jakości życia na podstawie kwestionariusza SF-36v2	272
Maciej Beręsewicz, Tomasz Klimanek: Wykorzystanie estymacji pośredniej uwzględniającej korelację przestrzenną w badaniach cen mieszkań	281
Karolina Paradysz: Benchmarkowa analiza estymacji dla małych obszarów na lokalnych rynkach pracy	291
Anna Gryko-Nikitin: Dobór parametrów w równoległych algorytmach genetycznych dla problemu plecakowego	301
Tomasz Ząbkowski, Piotr Jałowiecki: Zastosowanie reguł asocjacyjnych do analizy danych ankietowych w wybranych obszarach logistyki przedsiębiorstw przetwórstwa rolno-spożywczego	311
Agnieszka Przedborska, Małgorzata Misztal: Zastosowanie metod statystyki wielowymiarowej do oceny wydolności stawów kolanowych u pacjentów z chorobą zwyrodnieniową leczonych operacyjnie	321
Dorota Perło: Rozwój zrównoważony w wymiarze gospodarczym, społecznym i środowiskowym – analiza przestrzenna	331

Ewa Putek-Szeląg, Urszula Gieraltowska, Analiza i diagnoza wielkości produkcji energii odnawialnej w Polsce na tle krajów Unii Europejskiej..	342
--	-----

Summaries

Sabina Denkowska, Kamil Fijorek, Marcin Salamaga, Andrzej Sokolowski: VIth-term Sejm – a voting machine	18
Barbara Pawelek, Adam Sagan: Latent variables in econometric models – respecification of Klein I model	28
Jan Paradysz: New possibilities for studying the situation on the labour market	40
Krzysztof Najman: Self-learning neural network of GNG type in the dynamic clustering of high-dimensional data.....	47
Kamila Migdał-Najman: Applying the one-dimensional SOM network to select variables in dynamic clustering	57
Aleksandra Matuszewska-Janica, Dorota Witkowska: Gender wage gap: application of classification trees.....	66
Iwona Foryś, Ewa Putek-Szeląg: Spatial classification of communes by usable land traded by the APA in the Zachodniopomorskie voivodeship...	76
Joanna Banaś, Małgorzata Machowska-Szewczyk: Classification of Internet banking accounts including symbolic variables	84
Marta Jarocka: The impact of the method of the selection of diagnostic variables on the result of linear ordering on the example of ranking of universities in Poland.....	94
Anna Zamojska: Empirical analysis of the consistency of mutual fund ranking for different portfolio performance measures.....	105
Dorota Rozmus: Comparison of accuracy of affinity propagation clustering and cluster ensembles based on bagging idea.....	114
Ewa Wędrowska: Sensitivity of divergence measures as structure dissimilarity measurements	123
Katarzyna Wójcik, Janusz Tuchowski: Machine translation impact on the results of the sentiment analysis	134
Małgorzata Misztal: Assessment of the influence of selected imputation methods on the results of object classification using classification trees ...	145
Anna Czapkiewicz, Beata Basiura: Simulation study of the selection of coefficient depending on the clustering time series.....	153
Tomasz Szubert: Factors differentiating the level of satisfaction with life and the life's values of people with and without disabilities in the light of the "Social Diagnosis" survey	162
Marcin Szymkowiak: Construction of calibration estimators of totals for different distance measures	173

Wojciech Roszka: Joint characteristics' estimation of variables not jointly observed.....	181
Justyna Brzezińska: Visualizing categorical data in \mathbf{R}	190
Agata Sielska: Regional diversity of competitiveness potential of Polish farms after the accession to the European Union	200
Mariusz Kubus: Regularized linear probability model as a filter	208
Beata Basiura: The Ward method in the application for classification of Polish voivodeships with different distances.....	216
Katarzyna Wardzińska: Application of Data Envelopment Analysis in company classification process.....	225
Katarzyna Dębowska: Modeling corporate bankruptcy based on unbalanced samples	234
Danuta Tarka: Influence of the features selection method on the results of objects classification using environmental data.....	245
Artur Czech: Application of chosen methods for the selection of diagnostic variables in indirect consumption research.....	254
Beata Bal-Domańska: Assessment of relations occurring between smart growth and economic cohesion in regional dimension using panel models	263
Mariola Chrzanowska: Ordinary kriging and inverse distance weighting as methods of estimating prices based on Warsaw real estate market	271
Adam Depta: Application of analysis of variance in the study of the quality of life based on questionnaire SF-36v2	280
Maciej Beręsewicz, Tomasz Klimanek: Using indirect estimation with spatial autocorrelation in dwelling price surveys.....	290
Karolina Paradysz: Benchmark analysis of small area estimation on local labor markets	300
Anna Gryko-Nikitin: Selection of various parameters of parallel evolutionary algorithm for knapsack problems	310
Tomasz Ząbkowski, Piotr Jałowiecki: Application of association rules for the survey of data analysis in the selected areas of logistics in food processing companies	320
Agnieszka Przedborska, Małgorzata Misztal: Using multivariate statistical methods to assess the capacity of the knee joint among the patients treated surgically for osteoarthritis	330
Dorota Perło: Sustainable development in the economic, social and environmental dimensions – spatial analysis.....	341
Ewa Putek-Szeląg, Urszula Gieraltowska: Analysis and diagnosis of the volume of renewable energy production in Poland compared to EU countries	352

Barbara Pawelek, Adam Sagan

Uniwersytet Ekonomiczny w Krakowie

ZMIENNE UKRYTE W MODELACH EKONOMICZNYCH – RESPECYFIKACJA MODELU KLEINA I

Streszczenie: Modele strukturalne ze zmiennymi ukrytymi stanowią jedno z dominujących podejść analitycznych w naukach społecznych. Stanowią one połączenie dwóch typów modeli: konfirmacyjnej analizy czynnikowej oraz analizy regresji. Celem pracy jest respecyfikacja modelu Kleina I uwzględniająca ekonometryczną i psychometryczną perspektywę w budowie modelu strukturalnego ze zmiennymi ukrytymi. Polega ona na wprowadzeniu do modelu Kleina I zmiennych ukrytych i uwzględnieniu w procesie budowy modelu błędów pomiaru konstruktów ekonomicznych. W pracy wykorzystano podejście SEM do szacowania modelu ekonometrycznego z błędami pomiaru, tożsamościami i ograniczeniami nałożonymi na parametry modelu.

Słowa kluczowe: model Kleina I, zmienna ukryta, błąd pomiaru.

1. Modele ze zmiennymi ukrytymi – między psychometrią a ekonometrią

Modele strukturalne ze zmiennymi ukrytymi (SEM) stanowią jedno z dominujących podejść analitycznych w naukach społecznych. Są one połączeniem dwóch typów modeli: konfirmacyjnej analizy czynnikowej (część pomiarowa) oraz analizy regresji (część strukturalna). Wyrażają one powiązanie tradycji psychometrycznej z jednej strony i ekonometrycznej z drugiej. Wykorzystanie modeli strukturalnych w psychometrii i badaniach społecznych dotyczy głównie problemów oceny przypadkowych i systematycznych błędów pomiaru konstruktów teoretycznych i związanej z nimi rzetelności i trafności skal. Konfirmacyjna analiza czynnikowa pozwala na budowę modelu oceny rzetelności i wykorzystanie licznych wskaźników tzw. największej dolnej granicy rzetelności w analizie wiarygodności pomiaru. Część strukturalna modelu związana z analizą zależności między konstruktami teoretycznymi czerpie najsilniej z dorobku ekonometrii [Heckman 2000]. Dotyczy ona problemów teoretycznej specyfikacji modelu [Spanos 1986], wynikających z błędu substancjalnego (aproksymacji) lub statystycznego (estymacji) modelu oraz zrozumienia roli relacji między modelem teoretycznym wyjaśniającym dane a stochastycznym proce-

sem je generującym i założeniami z nim związanymi (dotyczącymi typu rozkładu, niezależności i heterogeniczności) [Johansen 2007]. Identyfikacja zależności przyczynowych w badaniach nieeksperymentalnych jest dokonywana w tych modelach na podstawie zasady wspólnej przyczyny i przyczynowego warunku Markowa [Pearl 2000] przez testowanie efektów mediacyjnych i warunkowej niezależności zmiennych oraz wprowadzanie zmiennych instrumentalnych do modelu [Angrist, Imbens, Rubin 1996]. Analiza efektów potencjalnych (kontrafaktycznych) w podejściach eksperymentalnych jest dokonywana w SEM najczęściej z wykorzystaniem modeli ukrytych różnic [Steyer 2005].

Celem artykułu jest respecyfikacja modelu Kleina I uwzględniająca ekonometryczną i psychometryczną perspektywę w budowie modelu strukturalnego ze zmiennymi ukrytymi. Wykorzystanie dorobku psychometrii w budowie modeli ekonometrycznych jest najczęściej związane z wprowadzeniem refleksywnych wskaźników do modelu pomiarowego zmiennych ukrytych, występujących jako egzo- lub endogeniczne zmienne w części strukturalnej modelu. W przypadku modelu Kleina I jego respecyfikacja polega na 1) wprowadzeniu zmiennych ukrytych reprezentujących odpowiednie konstrukty ekonomiczne występujące w klasycznym podejściu Kleina jako zmienne jawne, 2) wykorzystaniu jednowskaźnikowych zmiennych obserwowalnych w modelach pomiarowych zmiennych ukrytych oraz 3) uwzględnieniu w oszacowanym modelu błędu pomiaru konstruktów ekonomicznych¹. Dodatkowym celem jest wykorzystanie możliwości modelowania SEM w szacowaniu modeli ekonometrycznych z tożsamościami i ograniczeniami nałożonymi na parametry modelu.

2. Estymacja modelu Kleina I – porównanie metod

Próbę zastosowania metodologii SEM w badaniach ekonomicznych zaczęto od rozważenia wybranych klasycznych modeli ekonometrycznych dla zjawisk makroekonomicznych. W pierwszej kolejności wzięto pod uwagę model gospodarki USA, który został skonstruowany przez L.R. Kleina w latach czterdziestych (tzw. model Kleina I). Model ten składa się (w oryginalnej wersji opublikowanej w 1950 r.) z sześciu równań. Model Kleina I jest bardzo często podawany w pracach z zakresu ekonometrii jako przykład układu równań współzależnych. Zapisy prezentowane w pracach różnią się między sobą (por. np. [Welfe 2003, s. 245-246; Greene 2003, s. 381]).

¹ W podejściu psychometrycznym w celu oszacowania błędu pomiaru i tym samym oceny rzetelności skali pomiarowej wymagane jest zastosowanie metody powtórzonych pomiarów lub stosowanie skal wielopozycyjnych (z co najmniej trzema wskaźnikami przypadającymi na jedną zmienną ukrytą). W respecyfikowanym modelu Kleina I podjęto próbę oszacowania błędów pomiarowych dla wskaźników jednopozycyjnych bez powtarzanych pomiarów, co nie jest całkowicie zgodne z klasycznym podejściem psychometrycznym.

W analizie prezentowanej w niniejszym artykule przyjęto następującą postać modelu Kleina I (na podstawie [Ekonometria... 2007, s. 16]):

$$C = \alpha_1 + \alpha_2 P + \alpha_3 P_{-1} + \alpha_4 W + \varepsilon_1, \quad (1)$$

$$I = \alpha_5 + \alpha_6 P + \alpha_7 P_{-1} + \alpha_8 K_{-1} + \varepsilon_2, \quad (2)$$

$$Wp = \alpha_9 + \alpha_{10} E + \alpha_{11} E_{-1} + \alpha_{12} t + \varepsilon_3, \quad (3)$$

$$Y = C + I + G - T, \quad (4)$$

$$P = Y - W, \quad (5)$$

$$K = K_{-1} + I, \quad (6)$$

$$W = Wp + Wg, \quad (7)$$

$$E = Y + T - Wg, \quad (8)$$

gdzie: C – konsumpcja,
 I – inwestycje,
 W – płace ogółem,
 Wp – płace w sektorze prywatnym,
 Wg – płace w sektorze państwowym,
 P – zysk,
 K – kapitał,
 E – produkcja sektora prywatnego,
 Y – dochód,
 T – podatki pośrednie,
 t – zmienna czasowa,
 G – wydatki rządowe,
 $\varepsilon_1, \varepsilon_2, \varepsilon_3$ – składniki losowe, natomiast P_{-1} , K_{-1} i E_{-1} oznaczają wartości zmiennej, odpowiednio, P , K i E z okresu poprzedniego.

Na tak zapisany model (1)-(8) składają się trzy równania stochastyczne (1)-(3) i pięć równań tożsamościowych (4)-(8). Wśród zmiennych można wyróżnić osiem zmiennych łącznie współzależnych (C , I , Wp , Y , P , K , W , E) i siedem zmiennych z góry ustalonych (Wg , G , T , t , P_{-1} , K_{-1} , E_{-1}).

L.R. Klein [1950, s. 135-141] do oszacowania wartości zmiennych C , I , G , P , Wp i Wg wykorzystał 31 zmiennych makroekonomicznych. Na przykład wzór na konsumpcję wygląda następująco [Klein 1950, s. 135-136]:

$$C = \frac{C_1 + C_2}{C_3}, \quad (9)$$

gdzie: C_1 – wydatki na konsumpcję dóbr i usług (w miliardach dolarów z 1934 r.),

- C_2 – indywidualne wydatki netto na wynajem (w miliardach dolarów z 1934 r.),
 C_3 – wskaźnik wydatków konsumenckich.

Dane statystyczne wykorzystane przez L.R. Kleina do obliczenia poziomu konsumpcji (jednej ze zmiennych endogenicznych w modelu Kleina I) w USA w latach 1920-1941 zaczerpnięto z prac źródłowych [Painter 1945, s. 873; Kuznets, Epstein, Jenks 1946, s. 735; Kuznets, Epstein, Jenks 1941, s. 145]. Po wykonaniu obliczeń na danych pierwotnych uzyskano dokładniejsze oceny konsumpcji niż te, które są zamieszczone w pracy L.R. Kleina z 1950 r. Wyniki obliczeń wykonanych na podstawie danych oryginalnych różnią się $\pm 0,1$ dla lat 1927, 1929, 1932, 1933 i 1935 w stosunku do danych zamieszczonych w pracy L.R. Kleina [1950, s. 135] i wykorzystywanych przez wielu autorów w późniejszych pracach (zob. np. [Goldberger, Nagar, Odeh 1961, s. 563-564; Greene 2003, s. 950]). Po przekształceniu oryginalnych danych – zgodnie z przyjętymi przez L.R. Kleina zaokrągleniami – uzyskuje się takie same wartości dla zmiennej C jak te opublikowane przez Kleina w 1950 r.

Zadano pytanie o to, czy zachowując konwencję prezentacji wyników estymacji parametrów modelu Kleina I przyjętą w pracy [Klein 1950], tzn. prezentacji ocen parametrów z dokładnością do drugiego miejsca po przecinku, będzie można zobaczyć wpływ przyjęcia określonego zbioru wartości konsumpcji (C lub C').

Tabela 1. Wyniki estymacji równania dla zmiennej C w modelu Kleina I na podstawie różnych wartości konsumpcji z wykorzystaniem metody największej wiarygodności z pełną informacją

Zmienna	Ocena parametru	Ocena błędu standardowego	Wartość p -value
Zmienna zależna: C – dane z pracy [Klein 1950, s. 135] z dokładnością do 1 miejsca po przecinku			
$const$	18,3433	2,48502	1,56e-013 ***
P	-0,232387	0,311955	0,4563
P_{-1}	0,385672	0,217357	0,0760*
\bar{W}	0,801844	0,035893	1,52e-110 ***
Zmienna zależna: C' – dane na podstawie prac źródłowych z dokładnością do 6 miejsc po przecinku			
$const$	18,2960	2,43607	5,89e-014 ***
P	-0,213747	0,298655	0,4742
P_{-1}	0,375730	0,208096	0,0710*
\bar{W}	0,799509	0,035622	1,45e-111 ***

Źródło: obliczenia własne z wykorzystaniem programu Gretl 1.9.7.

W tabeli 1 zamieszczono wyniki estymacji równania dla zmiennej C w modelu Kleina I metodą największej wiarygodności z pełną informacją. Analizując wyniki zamieszczone w tab. 1, można zauważyć, że różnice w ocenach parametrów spowodowane zwiększeniem dokładności oszacowań konsumpcji (przy niezmiennych wartościach pozostałych zmiennych w stosunku do wartości zamieszczonych w pracy [Klein 1950, s. 135]) są widoczne – w wielu przypadkach – już na drugim miejscu

po przecinku. Podobne rozważania można przeprowadzić dla pozostałych dwóch równań stochastycznych w modelu Kleina I.

3. Model Kleina I z jednowskaźnikowymi zmiennymi ukrytymi

W modelach strukturalnych ze zmiennymi ukrytymi stosowanymi w badaniach ekonomicznych wykorzystuje się dwie klasy wskaźników: refleksyjne oraz formatywne. Pierwszy typ wskaźników jest związany z tradycją psychometryczną (klasyczną teorią testu i teorią reakcji na pozycję). Wskaźniki formatywne są podstawowym rodzajem wskaźników wykorzystywanych w badaniach społecznych i ekonomicznych. Wyróżnić można dwa rodzaje wskaźników formatywnych. Pierwszy z nich związany z formatywną zmienną ukrytą jest mierzony z określonym błędem pomiarowym. Drugi typ zmiennej formatywnej jest najczęściej definiowany w podobny sposób jak główna składowa, która stanowi w całości liniową kombinację wskaźników (bez błędu pomiaru). Wykorzystanie w budowie modelu wskaźników zarówno formatywnych, jak i refleksyjnych jest charakterystyczne dla modeli z wielorakimi przyczynami i wielorakimi wskaźnikami (MIMIC), w których zmienna ukryta jest mierzona za pomocą obu typów wskaźników.

Szczególnym typem modeli ze zmiennymi ukrytymi są modele jednowskaźnikowe (z refleksyjnym wskaźnikiem). Ten rodzaj modeli jest rzadziej spotykany w badaniach psychologicznych, które powszechnie wykorzystują skale wielopozycyjne, ale mają szerokie zastosowania w badaniach marketingowych i ekonomicznych. Wprowadzenie jednowskaźnikowych zmiennych ukrytych może wynikać z następujących przesłanek: 1) założenie braku błędu pomiarowego, 2) założenie, że wskaźnik formatywny w pełni kształtuje mierzone zjawisko lub je odzwierciedla bez błędu pomiaru, 3) łączenie wskaźników, 4) wykorzystanie skali sumowanych ocen, 5) uzyskanie prostej struktury czynnikowej Thurstone'a.

Podstawowym problemem w stosowaniu jednowskaźnikowych modeli pomiarowych jest kwestia oszacowania wariancji błędu pomiaru. Przy założeniu braku błędu pomiarowego model pomiaru z jednym wskaźnikiem jest tożsamy z modelem, w którym występują jedynie zmienne obserwowalne.

Ocena wariancji błędu dla jednowskaźnikowych zmiennych ukrytych wynika z podstawowego równania modelu pomiarowego:

$$\sigma_x = \lambda^2 \cdot \theta + \delta_x, \quad (10)$$

gdzie: σ_x – wariancja wskaźnika,
 λ – ładunek czynnikowy,
 θ – wariancja zmiennej ukrytej,
 δ_x – wariancja błędu.

Korzystając z równania (10), można pokazać, że wariancja błędu jest funkcją wariancji wskaźnika i rzetelności pomiaru:

$$\delta_x = (1 - \alpha) \cdot \sigma_x, \quad (11)$$

gdzie współczynnik rzetelności pomiaru zmiennej ukrytej jest postaci:

$$\alpha = \frac{\lambda^2 \cdot \theta}{\sigma_x}. \quad (12)$$

Pomiar błędu zależy również od roli zmiennych ukrytych w modelu strukturalnym i rodzaju danych wejściowych. Dla niezależnych zmiennych ukrytych i danych wyrażonych w postaci macierzy korelacji wariancje wskaźnika oraz zmiennej ukrytej są równe 1. Dla zależnych zmiennych ukrytych skalowanie tych zmiennych zakłada ustalenie ładunku czynnikowego na poziomie 1. Zmienne ukryte reprezentujące błąd pomiaru są również nazywane błędem w zmiennych, zaś zmienna stanowiąca zmienną ukrytą reprezentującą zakłócenie związane z niewyjaśnioną wariancją zmiennej zależnej jest także nazywana błędem w równaniach. Brak uwzględnienia błędu pomiaru w modelach prowadzi do uzyskiwania obciążonych parametrów regresji, w których „prawdziwa” wartość współczynnika regresji lub korelacji między zmiennymi ukrytymi jest „tłumiona” nierzetelnością ich wskaźników. Obciążenie parametrów regresji wynika ze skorelowania składnika losowego ze zmienną niezależną. Dodatkowo brak uwzględnienia błędu pomiarowego powoduje niedoszacowanie współczynnika determinacji i obniżenie mocy predykcyjnej modelu [Wansbeck i Meijer 2000, s. 17].

4. Respecyfikacja modelu Kleina I – model z ograniczonymi parametrami

Specyfikacja modelu Kleina I w obszarze modelowania strukturalnego wiąże się z wykorzystaniem metody największej wiarygodności lub uogólnionej najmniejszej wariancji resztowej. Zastosowanie równań tożsamościowych w modelu powoduje brak identyfikacji modelu spowodowany niespełnieniem warunku rzędu dla empirycznej i odtworzonej macierzy kowariancji. W celu identyfikacji modelu strukturalnego obserwowalne zmienne endogeniczne w modelu są traktowane jako zmienne ukryte, z których tylko ukryte zmienne zależne związane z konsumpcją, inwestycjami i płacami są mierzone za pomocą jednostkowych ładunków czynnikowych. Dodatkowo, w związku z większą liczbą identyfikowanych zależnych zmiennych ukrytych niż zmiennych obserwowalnych, ustalone zostały startowe wartości parametrów dla wariancji zakłóceń. Ustalenie wartości początkowych powoduje, że odtworzona macierz kowariancji jest określona dodatnio podczas pierwszych kroków iteracji w procedurze najszybszego spadku Newtona-Raphsona. Jej zastosowanie jest jednak związane z otrzymywaniem obciążonych wartości statystyki χ^2 oraz błędów standardowych (założenie normalności rozkładu wskaźników i niewielka liczebność próby) oraz złamane jest założenie niezależności obserwacji w modelu Kleina I.

Oszacowanie błędu pomiarowego w modelu Kleina I z jednopozycyjnymi wskaźnikami zmiennych ukrytych zostało przeprowadzone na podstawie trzech metod: 1) indeksów modyfikacyjnych i wartości oczekiwanej zmiany parametrów, 2) metody dwuetapowej i 3) poszukiwania specyfikacji maksymalizującej kryteria informacyjne AIC i BIC. Metoda indeksów modyfikacyjnych nie wskazała na konieczność uwzględnienia wariancji lub kowariancji błędów pomiaru. W podejściu dwuetapowym przyjęto następujące rozwiązanie: w pierwszym etapie został oszacowany model z zerową wariancją błędu dla wskaźników (jest on tożsamy z klasycznym modelem strukturalnym bez zmiennych ukrytych), a w drugim etapie dla ustalonych parametrów modelu strukturalnego zostały oszacowane wariancje błędu (przy ograniczeniach wynikających z modeli przedstawionych na rys. 1). Trzecie podejście polegało na eksploracyjnym poszukiwaniu wartości wariancji błędu dla wskaźników minimalizujących kryteria informacyjne AIC i BIC.

Tabela 2 przedstawia oszacowania wariancji błędów pomiaru dla trzech symulacji. Indeksy modyfikacyjne wskazywały na uwzględnienie kowariancji między błędami pomiaru zmiennych ukrytych LWp i LT , natomiast rezultaty podejścia dwuetapowego i poszukiwania specyfikacji okazały się podobne.

Tabela 2. Oszacowania błędów pomiaru konstruktów ekonomicznych

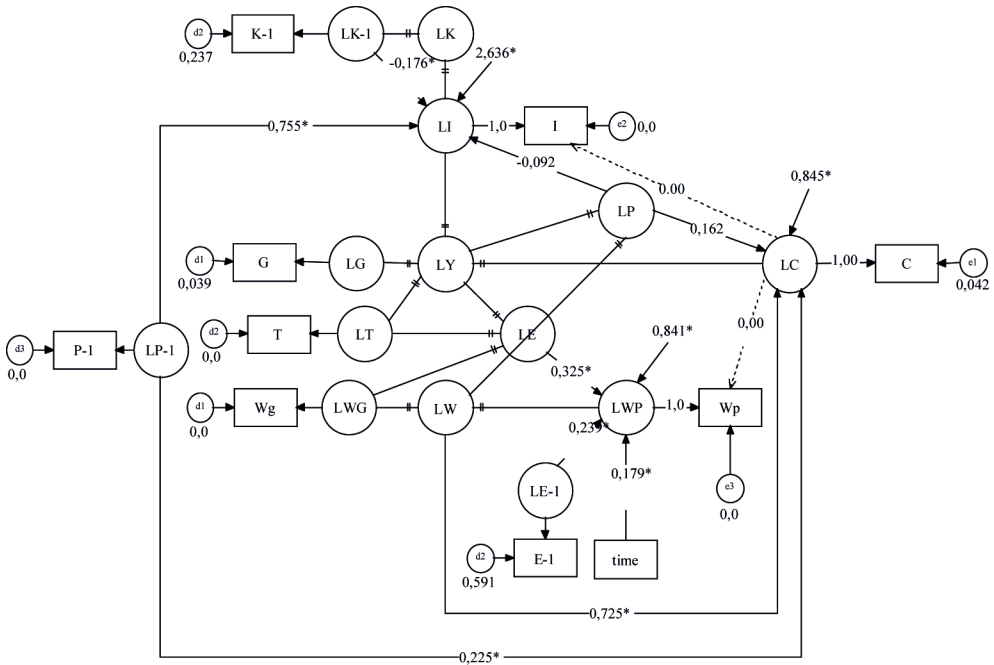
Zmienna	Indeksy modyfikacyjne	Podejście dwuetapowe	Poszukiwanie specyfikacji
C	–	0,042	0,00
I	–	0,000	0,00
Wp	-0,539	0,000	0,00
P 1	–	0,000	0,00
K 1	–	0,237	0,20
E 1	–	0,591	0,60
Wg	–	0,000	0,00
T	-0,539	0,000	0,00
G	–	0,039	0,00

Źródło: obliczenia własne z wykorzystaniem programu Mplus.

Porównując logarytmy wiarygodności, należy podkreślić, że wprowadzenie błędów pomiaru dla endogenicznych zmiennych opóźnionych LK _1 i LE _1 (one okazały się statystycznie istotne) pozwoliło na (jedynie niewielką) poprawę dopasowania modelu Kleina I (w porównaniu do modelu z zerowymi wariancjami błędu) o 1% (statystyka χ^2 dla modelu „bezbłędnego” wynosiła 215,785, a dla modelu z błędami pomiaru minimalizującego AIC była na poziomie 214,523)². Na rysunku 1 przedstawiony jest model Kleina I oszacowany za pomocą metody największej wiarygod-

² Podkreślić należy, że oszacowania błędów pomiarowych wynikały jedynie z analiz symulacyjnych. Autorzy dotarli jedynie do wielowskaźnikowych danych związanych z pomiarem konsumpcji (zmienną LC) w modelu Kleina I. Wykorzystanie wielu wskaźników w pomiarze konstruktów pozwoliłoby na bardziej realistyczną ocenę rzetelności pomiaru.

ności z jednowskaźnikowymi zmiennymi ukrytymi z uwzględnieniem błędów pomiaru.



Rys. 1. Model Kleina I z jednowskaźnikowymi zmiennymi ukrytymi

Źródło: opracowanie własne.

Na rysunku 1 każda zmienna ukryta (z wyjątkiem zmiennej czasowej *time*) jest mierzona za pomocą właściwego dla niej wskaźnika z uwzględnieniem błędu pomiaru (d). Ładunki czynnikowe i współczynniki regresji są oznaczone jako (\uparrow), a równania tożsamościowe jako ($\#$).

5. Uwagi końcowe

Integracja psychometrycznego i ekonometrycznego nurtu w modelach ze zmiennymi ukrytymi jest obiecującym kierunkiem rozwoju tego typu podejścia. Pierwszy z nich kładzie szczególny nacisk na interpretację konstruktów ekonomicznych w kategoriach zmiennych ukrytych, mierzonych za pomocą jedno- lub wielopozycyjnych wskaźników z *explicitie* modelowanym błędem pomiarowym. Drugi nurt jest związany z metodologicznie poprawną specyfikacją przyczynowych zależności strukturalnych między konstruktami. W zaprezentowanym modelu Kleina I parametry modelu zostały oszacowane z wykorzystaniem standardowego programu do modelowania zmiennych ukrytych Mplus 6. Ocena wartości oszacowanych błędów pomiarowych

wskazuje, że wprowadzenie błędów pomiarowych w zasadzie nie zmienia interpretacji parametrów modelu Kleina I oszacowanych za pomocą metody dwuetapowej najmniejszych kwadratów. Wprowadzenie zmiennych ukrytych z błędami pomiaru również nie poprawia uzyskiwanych oszacowań z wyjątkiem istotnego błędu pomiaru zmiennych E_1 i K_1 . Oznacza to, że produkcja sektora prywatnego w modelu Kleina I stanowiąca zróżnicowany agregat ekonomiczny oraz (w mniejszym stopniu) kapitał zawierają w sobie istotny błąd pomiarowy.

Uwzględnienie w modelach ekonometrycznych zmiennych ukrytych pozwala na oszacowanie błędów pomiaru konstruktów i kategorii ekonomicznych i tym samym na istotne obniżenie błędu specyfikacji modelu ekonometrycznego.

Literatura

- Angrist J.D., Imbens G.W., Rubin D.B., *Identification of causal effects using instrumental variables (with discussion)*, "Journal of the American Statistical Association" 1996, 91, s. 444-472.
- Ekonometria współczesna*, red. M. Osińska, Wydawnictwo „Dom Organizatora”, Toruń 2007.
- Goldberger A.S., Nagar A.L., Odeh H.S., *The covariance matrices of reduced-form coefficients and of forecasts for a structural econometric model*, "Econometrica" 1961, vol. 29, no. 4.
- Greene W.H., *Econometric Analysis*, fifth edition, Prentice Hall, Pearson Education, Inc., Upper Saddle River, New Jersey 2003.
- Heckman J.J., *Causal parameters and policy analysis in economics: a twentieth century retrospective*, "Quarterly Journal of Economics" 2000, 115(1), s. 45-97.
- Johansen S., *Confronting the Economic Model with the Data*, [in:] D. Colander, *Post Walrasian Macroeconomics: Beyond the DSGE Model*, Cambridge University Press, Cambridge 2007.
- Klein L.R., *Economic Fluctuations in the United States 1921-1941*, Cowles Commission for Research in Economics, Monograph No. 11, John Wiley & Sons, Inc., New York, Chapman & Hall, Limited, London 1950.
- Kuznets S., Epstein L., Jenks E., *National Income, Aggregate Payments, and Consumers' Outlay*, [in:] *National Income and Its Composition, 1919-1938*, volume I, ed. S. Kuznets, assisted by L. Epstein and E. Jenks, National Bureau of Economic Research NBER, 1941, Chapter URL: <http://www.nber.org/chapters/c5540>.
- Kuznets S., Epstein L., Jenks E., *Basic Data, Sources and Methods: Finance*, [in:] *National Income and Its Composition, 1919-1938*, Volume II, ed. S. Kuznets, assisted by L. Epstein and E. Jenks, National Bureau of Economic Research NBER, 1946, Chapter URL: <http://www.nber.org/chapters/c5557>.
- Painter M.S., *Estimates of Gross National Product, 1919-1928 (On the basis of the Department of Commerce concept)*, Board of Governors of the Federal Reserve System (U.S.) Washington, Federal Reserve Bulletin, September 1945, vol. 31, no. 9, accessed Feb 6, 2012 from FRASER, <http://fraser.stlouisfed.org/publication/?pid=62>.
- Pearl J., *Causality*, Cambridge University Press, Cambridge 2000.
- Spanos A., *Statistical Foundations of Econometric Modelling*, Cambridge University Press, Cambridge 1986.
- Steyer R., *Analyzing individual and average causal effects via structural equation models*, "Methodology" 2005, 1(1), s. 39-54.
- Welfe A., *Ekonometria*, Polskie Wydawnictwo Ekonomiczne, Warszawa 2003.
- Wansbeck T., Meijer E., *Measurement Error and Latent Variables in Econometrics*, Nord Holland, 2000.

LATENT VARIABLES IN ECONOMETRIC MODELS – RESPECIFICATION OF KLEIN I MODEL

Summary: Structural equation models with latent variables becomes one of dominant analytical approaches in social sciences. Usually, these models have FASEM form, that is combining confirmatory factor analysis (FA) and regression path modeling (SEM). The aim of the paper is the respecification of Klein I model taking into account both econometric and psychometric perspectives in the model development, introducing latent variables with measurement errors of economic constructs. The SEM-based models of Klein I with measurement errors, identities and parameter constrains are developed and compared.

Keywords: Klein model, latent variable, measurement error.