

Maria ROSIENKIEWICZ

Porównanie metod Akaike i Hellwiga w zakresie efektywności konstrukcji modelu regresyjnego

Modelowanie statystyczne znajduje zastosowanie w wielu dziedzinach nauki, m.in. w: ekonomii, mechanice, fizyce, biologii czy medycynie. W ekonomii jest to modelowanie ekonometryczne. Do podstawowych zastosowań modelowania ekonometrycznego należy wyjaśnianie, analizowanie i prognozowanie procesów zachodzących w gospodarce. Różnorakie metody statystyczne i ekonometryczne pozwalają budować m.in. modele inflacji, PKB, eksportu lub importu, poziomu zadłużenia czy kursów walutowych, a następnie na tej podstawie umożliwiają prognozowanie wielu istotnych wskaźników gospodarczych. Jednak metody ekonometryczne mogą być także z powodzeniem wykorzystywane w innych dziedzinach. Zastosowanie ekonometrii w mikroskali może odnosić się do modelowania procesów zachodzących wewnątrz przedsiębiorstw, instytucji oraz małych obszarów geograficznych. Mogą to być modele kosztów, przychodów bądź też obciążenia maszyn lub pracowników. Techniki ekonometryczne w mikroskali można zastosować właściwie do modelowania dowolnych zjawisk, jeśli tylko można je ująć w postaci kwantytatywnej.

Jednym z istotnych problemów jest dobór zmiennych objaśniających do modelu ekonometrycznego. Wybór właściwego zbioru zmiennych objaśniających pozwala na zbudowanie tzw. modelu efektywnego. Jest to tym samym podstawowy, a wręcz krytyczny krok umożliwiający należyte przedstawienie danego zjawiska. Niewłaściwy dobór zmiennych objaśniających powoduje, że model ekonometryczny będzie błędnie odzwierciedlał kształtowanie się analizowanych zależności lub też w ogóle nie będzie ich odzwierciedlał. W konsekwencji uniemożliwia to dokonanie wiarygodnych prognoz. Niezwykle ważne jest zatem wykorzystywanie przy konstrukcji modelu odpowiedniej metody doboru zmiennych objaśniających.

W literaturze światowej można znaleźć bardzo wiele sposobów wiodących do tego celu. Najpopularniejsze z nich to kryteria informacyjne Akaike (*Akaike's information criterion* — AIC) oraz Schwarza (*Schwarz information criterion* — SIC). W literaturze polskiej natomiast najczęściej opisywana jest metoda wskaźników pojemności informacji Hellwiga. Celem naszym było porównanie kryterium Akaike i metody Hellwiga oraz wykazanie, który z tych sposobów jest

bardziej efektywnym narzędziem doboru zmiennych objaśniających do modelu i tym samym uzyskania modelu optymalnego.

Aby umożliwić porównanie opracowano specjalny program w języku R (*The R Project for Statistical Computing, R language*). Program ten oparty jest na prostej symulacji, która polega na wygenerowaniu zestawu danych posiadających rozkład normalny, a następnie zbudowaniu modelu liniowego, w którym zmienna objaśniana jest zależna od danych wcześniej wygenerowanych. W kolejnym etapie zastosowane są analizowane metody wyboru modelu, czyli sposoby Akaike i Hellwiga, a następnie dokonywane jest porównanie, która z tych metod wskazała właściwy model. Przeprowadzenie symulacji według różnych parametrów pozwoliło porównać skuteczność omawianych metod doboru zmiennych objaśniających. Następnie przeprowadzono porównanie tychże metod na podstawie budowy modeli wykorzystujących dane empiryczne. Otrzymane wyniki stały się solidną podstawą porównania kryterium informacyjnego Akaike oraz metody Hellwiga. Pozwoliły one na wskazanie, który z tych sposobów jest właściwym narzędziem doboru zmiennych objaśniających do modelu ekonometrycznego.

KRYTERIUM INFORMACYJNE AKAIKE

W 1971 r. Hirotugu Akaike zaproponował kryterium wyboru modelu (nazwane kryterium informacyjnym Akaike), które przedstawił w pracach *Information theory and an extension of the maximum likelihood principle* z 1973 r. oraz *A new look at the statistical model identification* z 1974 r. Początkowo kryterium to stosowane było tylko dla modeli szeregów czasowych, lecz później zaczęto używać go także w analizie regresji. AIC znajduje zastosowanie nie tylko w przypadku wyboru modeli liniowych, ale także nieliniowych czy też modeli bez stałej. Sam Akaike zauważył, że to kryterium dzięki swej prostocie i stosowności przy wyborze modeli może być używane w wielu dziedzinach, takich jak: inżynieria, geofizyka, hydrologia, ekonometria, psychometria czy medycyna. Kryterium informacyjne Akaike można uznać za uogólnienie kryterium C_p Mallowsa¹. Jednak zamiast średniej sumy kwadratów błędów predykcji do wyboru pomiędzy funkcją gęstości modelu prawdziwego i funkcją gęstości modelu użytego do prognozowania, Akaike stosuje miarę ilości informacji².

Jak już wspomniano, bardzo ważnym problemem w modelowaniu jest wskazanie na właściwy wektor zmiennych, który prowadziłby do zbudowania modelu optymalnego. Częstym podejściem jest wybranie kilku wektorów o różnych wymiarach, czyli kilku podzbiorów zawierających różne kombinacje zmiennych objaśniających, następnie zbudowanie modeli odpowiadających tym wektorom. W dalszej kolejności oblicza się jakieś kryterium wyboru modelu dla każdego z modeli kandydujących i w końcu wskazuje na wektor optymalnego modelu na

¹ Peracchi (2001), s. 331.

² Chow (1995), s. 354.

podstawie wartości zadanego kryterium. I właśnie za najszerzej znane i najczęściej używane do tego celu uznawane jest kryterium informacyjne Akaike. AIC może być interpretowane jako miara odległości pomiędzy modelem dopasowanym (teoretycznym) do zebranych, niekompletnych danych statystycznych a modelem efektywnym, który wygenerował te dane. Dobrze znaną miarą odległości pomiędzy dwoma modelami jest nieznormalizowana informacja Kullbacka-Leiblera³, określana także jako odległość Kullbacka-Leiblera (KL)⁴. Odległość KL można zinterpretować jako traconą informację, kiedy rzeczywistość (model efektywny) zostaje aproksymowana danym modelem teoretycznym. Z tego względu sama ta odległość nie może być używana jako kryterium selekcji modeli, ponieważ zależy ona od nieznanego rozkładu. Dlatego zachodzi konieczność zastosowania estymatora informacji KL, którym jest właśnie funkcja AIC. Dla odległości KL zachodzi zależność polegająca na tym, że im mniejsza jest rozbieżność pomiędzy modelem efektywnym i teoretycznym, tym lepszy jest model teoretyczny i tym lepiej aproksymuje on model prawdziwy, a zatem model optymalny, który można zbudować minimalizując utratę informacji.

Akaike proponował estymację⁵ informacji Kullbacka-Leiblera za pomocą metody największej wiarygodności. W jej wyniku wyprowadził kryterium informacyjne AIC, o postaci:

$$AIC(\hat{\theta}) = -2 \ln \hat{L} + 2p$$

gdzie:

- p — liczba parametrów modelu (liczba zmiennych objaśniających wraz z wyrazem wolnym),
- $\hat{L} = L(\hat{\theta}) = \max \{L(\theta_i, D)\}$ — maksimum funkcji wiarygodności dla estymowanego modelu.

Kryterium informacyjne Akaike jest to zatem różnica podwojonej liczby parametrów modelu i podwojonego logarytmu naturalnego z maksimum funkcji wiarygodności. Spośród różnych modeli za optymalny uznany jest ten, dla którego AIC osiąga najmniejszą wartość⁶.

Kiedy analiza statystyczna oparta jest na metodzie najmniejszych kwadratów (MNK), wzór na AIC można zapisać również następująco⁷:

³ Cavanaugh, Shumway (1998), s. 45—65.

⁴ Kukier i in. (2008), s. 8: http://www.kul.lublin.pl/files/57/working_papers/kukier_i2008.pdf.

⁵ AIC jest więc estymatorem całości informacji KL, czyli funkcją pozwalającą oszacować, aproksymować wartość odległości KL.

⁶ Peracchi (2001), s. 331.

⁷ <http://uczelnia.warszawska.pl/upl/1181568659.pdf?PHPSESSID=43887cbc2ca6003603a28048ff7ce921>
z 20 lipca 2008 r.

$$AIC = n \ln \left(\sum_{i=1}^n e_i^2 \right) + 2K$$

gdzie:

$\sum_{i=1}^n e_i^2$ — suma kwadratów reszt,
 K — liczba parametrów modelu,
 n — wielkość próby.

Oba wzory na AIC są w tym przypadku równoważne.

METODA WSKAŹNIKÓW POJEMNOŚCI INFORMACJI HELLWIGA

Metodę wskaźników pojemności informacji przedstawił Zdzisław Hellwig w roku 1969 na łamach „Przeglądu Statystycznego”⁸. Hellwig wyróżnił pojęcie czystego oraz zanieczyszczonego nośnika informacji. Zaznaczył, że dana zmienna X_j jest tym lepszym nośnikiem informacji o danej zmiennej Y , im wartość bezwzględna współczynnika korelacji r_j tych dwóch zmiennych jest bliższa jedności. Uznał ponadto, że dana zmienna X_j jest czystym nośnikiem informacji jedynie w przypadku, gdy nie jest ona skorelowana z innymi predyktantami. Natomiast, jeśli zmienna X_j jest z nimi skorelowana, to wówczas informacja, jaką zmienna ta niesie o zmiennej Y , zawiera w sobie informacje przekazywane o niej za pośrednictwem zmiennej X_j przez pozostałe predyktanty skorelowane z X_j . Wtedy zmienna X_j jest określana mianem zanieczyszczonego nośnika informacji o zmiennej Y .

W metodzie Hellwiga procedura wyboru optymalnego zbioru zmiennych objaśniających, który pozwoli zbudować optymalny model, polega na tym, że dla każdej zmiennej z rozpatrywanej kombinacji wyznaczana jest indywidualna pojemność nośników informacji⁹:

$$h_{kj} = \frac{r_j^2}{\sum_{i \in Z_k} |r_{ij}|} \quad (i \in Z_k)$$

gdzie:

k — numer kombinacji ($k = 1, 2, \dots, 2^m - 1$),
 Z_k — zbiór numerów zmiennych w rozpatrywanej kombinacji,

⁸ Hellwig (1969), s. 225—236.

⁹ Dziechciarz (2002), s. 33.

j — numer zmiennej w danej kombinacji,
 r_j — współczynnik korelacji potencjalnej zmiennej objaśniającej o numerze j ze zmienną objaśnianą,
 r_{ij} — współczynnik korelacji między i -tą i j -tą potencjalną zmienną objaśniającą.

Wskaźnik h_{kj} mierzy wielkość informacji, jaką zmienna X_j wnosi o zmiennej objaśnianej Y w k -tej kombinacji. Po wyznaczeniu wartości h_{kj} dla wszystkich zmiennych oblicza się pojemność integralną kombinacji nośników informacji. Wyznacza się ją dla każdej kombinacji według wzoru:

$$H_k = \sum_{j \in Z_k} h_{kj}$$

Jest ona sumą indywidualnych pojemności nośników informacji, które wchodzi w skład danej kombinacji. Kryterium wyboru odpowiedniej kombinacji zmiennych objaśniających stanowi pojemność informacji. Według tego kryterium należy wybrać kombinację zmiennych, która wnosi najwięcej informacji, czyli taką, dla której H_k osiąga najwyższą wartość.

Zaletę metody Hellwiga stanowi fakt, że wykorzystywany w niej parametr H jest wielkością unormowaną, zawartą w przedziale $<0,1>$. Ta własność parametru H umożliwia określenie, jaka jest relatywna pojemność informacji danej kombinacji zmiennych objaśniających. Jeśli pojemność ta jest bliska jedności, oznacza to, że zmienne objaśniające, które wchodzi w skład danej kombinacji, dostarczają niemal pełnego zasobu informacji o zmiennej objaśnianej Y i poszerzanie danego zbioru o dalsze zmienne objaśniające jest bezcelowe. Zaletą tej metody jest też to, że pozwala ona ocenić siłę, z jaką poszczególne zmienne objaśniające, wchodzące w skład optymalnej kombinacji, wpływają na zmienną objaśnianą Y . Poza tym metoda ta jest interesująca z tego względu, że nie zakłada się w niej znajomości wartości ocen parametrów modeli, które odpowiadają różnym kombinacjom zmiennych. Podstawowa wada tej metody polega jednak na tym, że zakres jej stosowania jest bardzo wąski. Przeznaczona jest ona bowiem tylko do doboru zmiennych do modeli liniowych. Nie znajduje zatem ona zastosowania np. przy budowie modeli nieliniowych.

Prowadzone badania empiryczne wskazują, że wybrana tą metodą optymalna kombinacja zmiennych objaśniających prowadzi często do modelu, który nie spełnia wielu podstawowych warunków, w szczególności warunku istotności ocen parametrów stojących przy wybranych zmiennych oraz warunku koincydencji. Należy również zauważyć, że istotnym minusem omawianej metody jest to, iż wymaga ona bardzo dużej liczby rachunków. Przy dużej liczbie zmiennych objaśniających (ponad 20) zachodzi konieczność przeanalizowania kilku, kilkunastu i więcej milionów kombinacji możliwych podzbiorów zmiennych objaśniających.

PORÓWNANIE METODY AKAIKE I HELLWIGA NA PODSTAWIE SYMULACJI KOMPUTEROWEJ

W celu porównania omówionych metod wyboru optymalnego modelu opracowano program uzależniony od dwóch parametrów. Pierwszym z nich jest „ n ”, czyli wielkość próby, natomiast drugi parametr to „ i ”, który odpowiada za ilość powtórzeń procedury. Zarówno „ n ”, jak i „ i ” należą do zbioru liczb naturalnych. Symulację można przeprowadzać zadając dowolną wartość wielkości próby „ n ” oraz dowolną wartość powtórzeń procedury „ i ” w zakresie od 1 do nieskończoności. Należy zaznaczyć, że nie zastosowano tolerancji odchylenia od ideału. W symulacji porównano, czy badana metoda (odpowiednio Akaike lub Hellwiga) wskazuje na taką samą liczbę zmiennych, jaką zawiera model prawdziwy, a następnie, czy zawiera identyczne zmienne, jak te występujące w modelu prawdziwym. Jeśli w obu przypadkach odpowiedź jest twierdząca, wówczas można wnioskować, że wybrany został optymalny zbiór zmiennych objaśniających, który pozwala na skonstruowanie modelu optymalnego. Model prawdziwy wykorzystany w symulacji ma postać:

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_4 X_4 + \varepsilon$$

natomiast model oszacowany według metody najmniejszych kwadratów ma postać:

$$\hat{y} = \hat{a}_0 + \hat{a}_1 x_1 + \hat{a}_2 x_2 + \hat{a}_3 x_3 + \hat{a}_4 x_4$$

Przeprowadzono szereg symulacji dla różnych wartości zmiennych parametrów „ n ” oraz „ i ”. Ostatnia z nich wykazała, że dla próby równej 10 tys., na 1 tys. powtórzeń metoda Akaike wskazała model optymalny 839 razy, czyli sprawdziła się w prawie 84%, zaś metoda Hellwiga wskazała model optymalny w 450 przypadkach na 1 tys., czyli sprawdziła się w 45%. Zwiększenie wielkości próby „ n ” z 1 tys. do 10 tys., w porównaniu do poprzedniej symulacji, znacznie poprawiło otrzymany wynik w przypadku metody Hellwiga (z 32% do 45%), natomiast w metodzie Akaike pozostał on praktycznie na tym samym poziomie. Na podstawie tych wyników opracowano wykresy.

Wykres 2 zawiera zestawienie wyników sześciu symulacji przeprowadzonych dla różnych wartości parametrów „ n ” oraz „ i ”. Wykonane badania pozwalają zauważyć, że każda z symulacji doprowadziła do wyniku, w którym metoda Akaike zdecydowanie więcej razy wskazywała model optymalny niż metoda wskaźników pojemności informacji Hellwiga. Przewaga kryterium informacyjnego AIC była widoczna niezależnie od rozmiaru próby i liczby powtórzeń procedury obliczeń.

Pierwsze trzy symulacje przeprowadzone zostały na stosunkowo niewielkiej próbie, zatem ich wyniki nie mogą zostać uznane za wiążące. Jednak pozostałe symulacje, w których próby były równe $n = 1$ tys., $n = 10$ tys. oraz $n = 100$ tys., wskazują na tę samą tendencję, w której metoda Akaike sprawdza się w ponad 80%. Oznacza to, że z prawdopodobieństwem ponad 80% metoda Akaike pozwala wybrać optymalny model ekonometryczny. Z kolei metoda Hellwiga wskazuje na model optymalny z prawdopodobieństwem nie większym niż 50%.

Porównując otrzymane wyniki można zaryzykować stwierdzenie, że przy wyborze właściwego modelu ekonometrycznego kryterium informacyjne Akaike jest prawie dwa razy bardziej skuteczne od metody wskaźników pojemności informacji Hellwiga. Przy uwzględnieniu faktu, że zakres stosowania metody Hellwiga jest bardzo ograniczony — ze względu na to, że przeznaczona jest ona jedynie do modeli liniowych — można stwierdzić, że kryterium informacyjne Akaike jest lepszym, bardziej skutecznym i bardziej wiarygodnym narzędziem wyboru prawdziwego modelu ekonometrycznego i tym samym optymalnego zbioru zmiennych objaśniających.

ZASTOSOWANIE METODY AKAIKE I HELLWIGA DO DANYCH EMPIRYCZNYCH — PORÓWNANIE

W następnym kroku wykonano porównanie metody AIC i metody Hellwiga na podstawie analiz empirycznych. Na podstawie kwartalnych danych empirycznych z lat 1947—2008 w Stanach Zjednoczonych (od I kwartału 1947 r. aż do II kwartału 2008 r.) zbudowano cztery modele ekonometryczne. Zmienne do pierwszego modelu wyselekcjonowano metodą Akaike, a do drugiego modelu metodą Hellwiga. Następnie zbudowano analogicznie kolejne dwa modele, ale tym razem dla danych w ujęciu przyrostowym. We wszystkich przedstawianych modelach przyjęto tolerancję precyzji na poziomie 95%. Do zbioru potencjalnych zmiennych objaśniających wybrano 6 zmiennych (zestawienie), a wielkość próby wynosi $n = 246$.

ZMIENNE OBJAŚNIAJĄCE WYSTĘPUJĄCE W MODELACH

Numer zmiennej	Nazwa zmiennej	Skrót
1	wydatki na konsumpcję	<i>PCE</i>
2	oszczędności	<i>GS</i>
3	stopa inflacji	<i>IR</i>
4	inwestycje	<i>GPDI</i>
5	eksport	<i>EX</i>
6	import	<i>IM</i>

Źródło: opracowanie własne.

Przeprowadzone obliczenia wykazały, że do pierwszego modelu ekonometrycznego PKB według kryterium AIC powinny wejść wszystkie zmienne ze

zbioru potencjalnych zmiennych objaśniających. Postać modelu po oszacowaniu parametrów jest następująca:

$$\hat{y}=118,574+1,370x_1-0,194x_2-410,884x_3+1,071x_4+0,905x_5-1,393x_6$$

Wyniki (MNK) zastosowanej do oszacowania parametrów i oceny ich precyzji oraz dokładności dopasowania przedstawiono w tabl. 1.

TABL. 1. WYNIKI MNK DLA MODELU STRUKTURALNEGO Z SZEŚCIOMA ZMIENNYMI OBJAŚNIAJĄCYMI

Wyszczególnienie	Błąd standardowy	<i>t</i> (45)	Poziom <i>p</i>
Wyraz wolny	13,88302	8,5410	0,000000
Zmienne: <i>PCE</i>	0,00977	140,2352	0,000000
<i>GS</i>	0,04456	-4,3523	0,000020
<i>IR</i>	98,00175	-4,1926	0,000039
<i>GDPI</i>	0,05770	18,5656	0,000000
<i>EX</i>	0,07176	12,6175	0,000000
<i>IM</i>	0,04395	-31,6938	0,000000
Błąd standardowy estymacji	47,79311		
Skorygowany <i>R</i> ²	0,99974		

Źródło: jak przy wykr. 1.

Można sądzić, że wszystkie parametry modelu są istotne, o czym świadczą wartości poziomu istotności *p*. Standardowe błędy szacunku nie są bardzo wysokie, co oznacza, że oszacowane wartości parametrów nie różnią się znacznie od ich rzeczywistej wartości. Standardowy błąd estymacji świadczy o tym, że oszacowana wartość PKB różni się średnio o 47,79 miliarda dolarów od rzeczywistej wartości.

Następny ekonometryczny model PKB zbudowano na zmiennych objaśniających wytypowanych przez metodę wskaźników pojemności informacji Hellwiga, która wskazała na jednoelementowy zbiór zmiennych objaśniających. Zatem liniowy ekonometryczny model PKB będzie miał tylko zmienną objaśniającą. Wyniki metody MNK prezentuje tabl. 2.

TABL. 2. WYNIKI MNK DLA MODELU STRUKTURALNEGO Z JEDNĄ ZMIENNĄ OBJAŚNIAJĄCĄ

Wyszczególnienie	Błąd standardowy	<i>t</i> (45)	Poziom <i>p</i>
Wyraz wolny	15,91951	21,7568	0,000000
Zmienna <i>PCE</i>	0,00379	367,2855	0,000000
Błąd standardowy estymacji	125,94093		
Skorygowany <i>R</i> ²	0,99819		

Źródło: jak przy wykr. 1.

Otrzymane wyniki wskazują na to, że wszystkie parametry modelu są istotne. Biorąc pod uwagę fakt, że model oparty na zmiennych wyselekcjonowanych przez metodę Akaike miał 6 zmiennych, a model oparty na zmiennych wyselekcjonowanych przez metodę Hellwiga tylko 1, nie można porównywać wartości współczynników determinacji R^2 tych modeli. Porównanie dopasowania modeli do danych empirycznych można natomiast oprzeć na skorygowanym współczynniku determinacji.

W przypadku modelu z 6 zmiennymi skorygowany R^2 wyniósł 0,99973892, natomiast dla modelu z 1 zmienną był równy 0,99818710. Oznacza to, że model zbudowany na podstawie kryterium informacyjnego Akaike ma wyższą wartość skorygowanego współczynnika determinacji, co pozwala sądzić, że model ten jest lepiej dopasowany do danych empirycznych niż model zbudowany metodą wskaźników pojemności informacji Hellwiga.

Standardowe błędy szacunku nie są bardzo wysokie, zatem oszacowana wartość parametrów nie różni się znacznie od ich rzeczywistej wartości. Z kolei oszacowana wartość PKB różni się średnio o 125,94 miliarda dolarów od rzeczywistej wartości. A zatem łatwo zauważyć, że standardowy błąd estymacji modelu z 6 zmiennymi jest znacznie niższy od standardowego błędu estymacji modelu z 1 zmienną (odpowiednio 47,79 i 125,94 miliarda dolarów). Pozwala to wnioskować, że model z 6 zmiennymi jest lepiej dopasowany do danych empirycznych niż model z 1 zmienną. Lepsze dopasowanie do danych empirycznych modelu z 6 zmiennymi potwierdza także wyk. 3.

W kolejnym kroku opracowano modele na podstawie danych w ujęciu przyrostowym. Model wyselekcjonowany przez kryterium informacyjne AIC zbudowano z 5 zmiennych objaśniających. Postać modelu po oszacowaniu parametrów to:

$$\hat{y} = 5,3672 + 1,038x_1 - 0,0695x_2 + 1,1455x_4 + 1,1693x_5 - 1,1258x_6$$

Wyniki metody MNK przedstawiono w tabl. 3.

TABL. 3. WYNIKI MNK DLA MODELU STRUKTURALNEGO Z PIĘCIOMA ZMIENNYMI OBJAŚNIAJĄCYMI

Wyszczególnienie	Błąd standardowy	$t(45)$	Poziom p
Wyraz wolny	1,18252	4,5387	0,000009
Zmienna: 1 (przyrostowo) <i>deltaPCE</i>	0,03252	31,9223	0,000000
2 (przyrostowo) <i>deltaGS</i>	0,02880	-2,4148	0,016496
4 (przyrostowo) <i>deltaGDPI</i>	0,03231	35,4542	0,000000
5 (przyrostowo) <i>deltaEX</i>	0,06889	16,9719	0,000000
6 (przyrostowo) <i>deltaIM</i>	0,06821	-16,5034	0,000000
Błąd standardowy estymacji			12,27210
Skorygowany R^2			0,92681

Źródło: jak przy wyk. 1.

Otrzymane wyniki wskazują, że wszystkie parametry modelu są istotne. Ponownie standardowe błędy szacunku nie są bardzo wysokie, co oznacza, że oszacowane wartości parametrów nie różnią się znacznie od ich rzeczywistej wartości. Oszacowana wartość zmiany PKB różni się średnio o 12,27 miliarda dolarów od rzeczywistej wartości.

Ostatni z analizowanych modeli został skonstruowany na podstawie trzelementowego zbioru zmiennych objaśniających, który wskazany został przez metodę Hellwiga. Postać modelu po oszacowaniu parametrów była następująca:

$$\hat{y} = 7,2390 + 0,8301x_1 - 0,8843x_4 + 0,6470x_5$$

Wyniki metody MNK przedstawiono w tabl. 4.

TABL. 4. WYNIKI MNK DLA MODELU STRUKTURALNEGO Z TRZEMA ZMIENNYMI OBJAŚNIAJĄCYMI

Wyszczególnienie	Błąd standardowy	<i>t</i> (45)	Poziom <i>p</i>
Wyraz wolny	1,72816	4,1888	0,000039
Zmienna: 1 (przyrostowo) <i>deltaPCE</i>	0,04398	18,8773	0,000000
4 (przyrostowo) <i>deltaGDPI</i>	0,03849	22,9771	0,000000
5 (przyrostowo) <i>deltaEX</i>	0,09017	7,1756	0,000000
Błąd standardowy estymacji	18,06700		
Skorygowany <i>R</i> ²	0,84137		

Źródło: jak przy wykr. 1.

Można sądzić, że wszystkie parametry modelu są istotne. W przypadku modelu z 5 zmiennymi skorygowany *R*² wyniósł 0,92681149, natomiast dla modelu z 3 zmiennymi był równy 0,84137305. Oznacza to, że model zbudowany według kryterium informacyjnego Akaike miał zdecydowanie wyższą wartość skorygowanego współczynnika determinacji, bo o prawie 9% (92,68%–84,14%=8,54%). Można zatem sądzić, że model ten jest znacznie lepiej dopasowany do danych empirycznych niż model zbudowany na podstawie metody wskaźników pojemności informacji Hellwiga. Oszacowana wartość zmiany PKB różni się średnio o 18,066995641 miliarda dolarów od rzeczywistej wartości. Można więc ponownie zauważyć, że standardowy błąd estymacji modelu z 5 zmiennymi jest niższy od standardowego błędu estymacji modelu z 3 zmiennymi (odpowiednio 12,27 i 18,07 miliarda dolarów). Pozwala to stwierdzić, że model z 5 zmiennymi jest lepiej dopasowany do danych empirycznych niż model z 3 zmiennymi.

Przeprowadzone analizy pozwalają ponownie wnioskować, że zastosowanie AIC jako kryterium wyboru optymalnego zbioru zmiennych objaśniających umożliwia zbudowanie modelu, który charakteryzuje zdecydowanie lepsze dopasowanie do danych empirycznych w porównaniu z modelem uzyskanym przy zmiennych objaśniających wyselekcjonowanych metodą wskaźników pojemności informacji Hellwiga. Tezę tę wydają się potwierdzać wartości skorygowa-

nych współczynników determinacji, standardowych błędów estymacji oraz zamieszczone wykresy.

Zakończenie

W artykule przedstawiono dwie analizy pozwalające porównać metody Akaike i Hellwiga. W obu przypadkach przeprowadzone badania wykazały, że biorąc za kryterium oceny zgodność dopasowania modelu do danych empirycznych, zdecydowanie lepsze rezultaty pozwala osiągnąć metoda Akaike. Modele skonstruowane na podstawie zbioru zmiennych objaśniających wyselekcjonowanych tą metodą charakteryzują się mniejszymi standardowymi błędami estymacji, wyższymi współczynnikami determinacji oraz lepszym dopasowaniem.

Przeprowadzone analizy danych empirycznych oraz otrzymane w nich wyniki zdają się potwierdzać wyniki otrzymane w wyniku symulacji, które wykazywały przewagę metody AIC niezależnie od rozmiaru próby i liczby powtórzeń procedury obliczeń. Analiza empiryczna z kolei udowodniła, że metoda Akaike wskazuje zbiór zmiennych objaśniających, na podstawie którego można skonstruować model pozwalający na lepsze dopasowanie i mniejsze błędy szacunku w porównaniu z modelem opartym na zmiennych wyselekcjonowanych przez metodę Hellwiga.

Podsumowując otrzymane wyniki warto odnieść się do pracy Dobromiła Serwy¹⁰. Celem przedstawionej tam analizy było rozstrzygnięcie, na ile metoda Hellwiga jest użyteczna w odniesieniu do konstruowania modeli szeregów czasowych i w jakim zakresie jest ona konkurencyjna wobec innych metod, opartych np. na kryterium informacyjnym Akaike czy Schwarza. Przeprowadzone tam badania wykazały, że metoda Hellwiga w pewnych, często w praktyce ekonometrycznej występujących przypadkach nie prowadzi do wyboru modelu oryginalnego (efektywnego, rzeczywistego), czyli takiego, który opisuje faktyczną relację między zmiennymi ekonomicznymi, fizycznymi lub innymi.

Według niektórych badaczy wyniki, które w badaniach empirycznych uzyskuje się dobierając układ zmiennych objaśniających o największej pojemności informacyjnej są w wielu przypadkach bardzo dyskusyjne. Z przeprowadzonego we wspomnianej pracy badania wynika również, iż przy analizie szeregów czasowych metoda Hellwiga nie zawsze pozwala na automatyczny wybór postulowanego modelu „idealnego” lub też model idealny nie zawsze jest równoważny modelowi oryginalnemu. Dokonano też tam analizy metody Hellwiga jako kryterium doboru zmiennych do modeli autoregresji. Zauważono, że przy spełnieniu pewnych prostych warunków metoda Hellwiga sugeruje wybór modelu autoregresji nieodpowiedniego rzędu. Wynik ten pozwala na uogólnienie do innych liniowych modeli szeregów czasowych, zwłaszcza modeli dynamicznych. W modelach tych często występują wzajemnie skorelowane zmienne objaśniające.

¹⁰ Serwa (2004), s. 5—17.

W wyniku zastosowania metody optymalnego doboru zmiennych objaśniających część z nich może zostać usunięta z grupy kandydatek, pomimo tego, że faktycznie wpływa ona istotnie na zmienną objaśnianą. Pominięcie ważnych zmiennych objaśniających w modelu może spowodować, że model zostanie odrzucony przy weryfikacji z powodu: niestabilności postaci strukturalnej, obciążonych ocen parametrów strukturalnych, autokorelacji składnika losowego bądź występowania dużych błędów prognozy. Z tego powodu w praktycznym zastosowaniu należy proponować wykorzystanie kryteriów informacyjnych, takich jak kryterium informacyjne Akaike. Tezę tę potwierdzają badania *stricte* analityczne, które przedstawił Tadeusz Bednarski i Filip Borowicz (2009).

Po przeprowadzeniu badań na danych empirycznych nie można zaprzeczyć przytoczonemu wcześniej stwierdzeniu, że przy wyborze właściwego modelu ekonometrycznego kryterium informacyjne Akaike jest prawie dwa razy bardziej skuteczne od preferowanej w polskich podręcznikach ekonometrycznych metody wskaźników pojemności informacji Hellwiga. Należy także przypomnieć, że ze względu na ograniczony zakres stosowania metody Hellwiga, gdyż jest przeznaczona tylko do modeli liniowych, można zaryzykować stwierdzenie, iż kryterium informacyjne Akaike jest zdecydowanie lepszym, bardziej skutecznym i wiarygodnym narzędziem wyboru efektywnego modelu ekonometrycznego i tym samym optymalnego zbioru zmiennych objaśniających. Wnioski te wydają się potwierdzać badania przeprowadzone w pracy T. Bednarskiego i F. Borowicza¹¹.

Podsumowując zatem można stwierdzić, że analizy, zarówno przeprowadzone na danych wygenerowanych sztucznie w wyniku symulacji jak i wykorzystujące dane empiryczne dowodzą, że aby skutecznie rozwiązać problem doboru optymalnego zbioru zmiennych objaśniających, który pozwoli skonstruować model prawdziwy, należy posługiwać się kryterium informacyjnym Akaike, gdyż jest to metoda zdecydowanie bardziej skuteczna i poprawna niż metoda wskaźników pojemności informacji Hellwiga.

mgr inż. Maria Rosienkiewicz — *Politechnika Wrocławska*

LITERATURA

- Bednarski T., Borowicz F. (2009), *On inconsistency of Hellwig's variable choice method in regression models*, „Discussiones Mathematicae Probability and Statistics”, No. 29
- Cavanaugh J. E., Shumway R. H. (1998), *An Akaike information criterion for model selection in the presence of incomplete data*, „Journal of Statistical Planning and Inference”, vol. 67, Issue 1, 16 March

¹¹ Bednarski, Borowicz (2009), s. 41—51.

- Chow G. C. (1995), *Ekonometria*, Wydawnictwo Naukowe PWN, Warszawa
- Dziechciarz J. (2002), *Ekonometria. Metody, przykłady, zadania*, Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu
- Hellwig Z. (1969), *Problem optymalnego wyboru predykant*, „Przegląd Statystyczny”, R. XVI, Zeszyt 3—4
- Kukier Ł., Szydłowski M., Tambor P. (2008), *Kryterium Akaike: prostota w języku statystyki*, KUL, Lublin
- Peracchi F. (2001), *Econometrics*, John Wiley & Sons Ltd, Chichester, West Sussex
- Serwa D. (2004), *Metoda Hellwiga jako kryterium doboru zmiennych do modeli szeregów czasowych*, SGH, Kolegium Analiz Ekonomicznych, Instytut Ekonometrii

SUMMARY

This paper presents a comparison of the Akaike Information Criterion (AIC) and the Hellwig Method as methods that select explanatory variables to a model and thus enable a choice of the true model. This comparison was made in two ways. At first, the simulations were constructed in the R software (The R Project for Statistical Computing). The purpose of these simulations was to identify which of the analyzed methods more times indicates the true model. Secondly, a comparison of both methods was made on the basis of empirical data. From the set of potential explanatory variables, a set of variables was selected according to the Akaike method and the other set according to the Hellwig method. On the basis of each of the selected sets, a model was developed. Subsequently, both models were compared in terms of their adjusted coefficients of determination, standard errors of estimate and the goodness-of-fit. Results given by the simulations and results coming from the empirical models analysis indicated that the Akaike Information Criterion is a better, more efficient and more reliable tool for selecting the optimal set of explanatory variables and the true econometric model.

РЕЗЮМЕ

В статье был сопоставлен информационный критерий Акайке (Akaike Information Criterion — AIC) и методы показателей информационного потенциала Гельвига в качестве методов выбора объяснительных переменных для модели и методов выбора эффективной модели. Это сопоставление проводилось двойным способом. В первую очередь были проведены симуляции в программе разработанной на языке R (The R Project for Statistical Computing). Целью симуляций было обследование, который из анализированных методов более раз укажет на истинную модель. Следующий анализ проводился на эмпирических данных. Из набора потенциальных объяснительных переменных был выбран методом Акайке один набор переменных, а второй методом Гельвига. На основе каждого из наборов была построена модель, а в дальнейшем методы были сопоставлены с учетом величины скорректированных коэффициентов

детерминации, стандартных ошибок оценки, а также соответствия подбора. Результаты полученные после проведения симуляции, а также результаты анализа эмпирических моделей ясно показывают, что информационный критерий Акайке является лучшим, более эффективным и надежным инструментом оптимального выбора переменных объясняющих и таким образом истинной эконометрической моделью.