

Alicja Grześkowiak

Akademia Ekonomiczna we Wrocławiu

ROZPOZNAWANIE RODZAJU HETEROSKEDASTYCZNOŚCI SKŁADNIKÓW LOSOWYCH MODELU EKONOMETRYCZNEGO Z WYKORZYSTANIEM DYSKRYMINACYJNEJ REGUŁY NAJWIĘKSZEJ WIARYGODNOŚCI*

1. Wstęp

Stwierdzenie występowania zjawiska niejednorodności wariancji składników losowych w modelu ekonometrycznym stwarza konieczność zastosowania uogólnionej metody najmniejszych kwadratów do estymacji parametrów. Implementacja uogólnionej metody najmniejszych kwadratów wymaga znajomości elementów diagonalnych macierzy kowariancji składników losowych. W badaniach empirycznych najczęściej zachodzi sytuacja, w której badacz nie ma o nich wiedzy *a priori*, dlatego postuluje się przyjęcie pewnego modelu heteroskedastyczności uwarunkowanego charakterem zmian wariancji składników losowych. Problemem, którego rozwiązanie zaproponowano w niniejszej pracy, jest wybór odpowiedniego w danej sytuacji modelu. Postulowana koncepcja stanowi połączenie podejścia klasycznego, opartego na teście statystycznym, z wykorzystaniem analizy dyskryminacyjnej. W pracy zaprezentowano podstawy teoretyczne sugerowanej koncepcji i zilustrowano ją badaniem empirycznym.

2. Wykrywanie heteroskedastyczności składników losowych modelu ekonometrycznego

Przedmiotem rozważań jest jedno z założeń jednorodności liniowego modelu ekonometrycznego o postaci:

* Praca naukowa finansowana ze środków na naukę w latach 2005/2006 jako projekt badawczy 1 H02B 025 29.

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + \varepsilon, \quad (1)$$

gdzie: Y – zmienna objaśniana,

$X_1, \dots, X_k$ – zmienne objaśniające,

$\beta_0, \dots, \beta_k$ – parametry strukturalne,

ε – składnik losowy.

Szacowanie parametrów modelu (1) przeprowadza się metodą najmniejszych kwadratów, której estymator ma pożądane własności, o ile spełnione są określone założenia (zob. np. [Greene 1997; Tomaszewicz 1985; Welfe 1998]). Z tego względu niezbędne jest weryfikowanie przyjętych założeń dotyczących rozkładu składników losowych, m. in. stałości wariancji, czyli tzw. homoskedastyczności.

Składniki losowe modelu ekonometrycznego są wielkościami nieobserwowalnymi, stąd też wynikają trudności w formułowaniu stwierdzeń dotyczących ich postaci. Wnioskowanie o składnikach losowych przeprowadza się na podstawie reszt modelu, które stanowią pewne przybliżenie składników losowych, lecz nie są ich realizacjami. Przegląd testów jednorodności wariancji można znaleźć np. w pracach [Greene 1997; Tomaszewicz 1985]. Weryfikowanie hipotezy o równości wariancji może mieć dwojaki aspekt (zob. [Tomaszewicz 1985, s. 136]):

- konstruktywny, gdy hipoteza alternatywna wskazuje na określony model heteroskedastyczności i jej przyjęcie może być podstawą wyboru metody estymacji,

- niekonstruktywny, gdy hipoteza alternatywna jest ogólna – stwierdza występowanie heteroskedastyczności, nie definiując jej postaci, dostarczając co najwyżej informacji dotyczących relacji większości między wariancjami składników losowych.

Każde z wymienionych podejść ma zarówno wady, jak i zalety. Zastosowanie testu konstruktywnego daje badaczowi wskazówki dotyczące ewentualnego dalszego postępowania, gdyż odrzucenie hipotezy zerowej prowadzi do przyjęcia alternatywy, w której zdefiniowany jest typ heteroskedastyczności. Istotny problem pojawia się w momencie braku podstaw do odrzucenia hipotezy zerowej, gdyż jedyną informacją jest wówczas niewystępowanie heteroskedastyczności danego rodzaju, co nie jest równoznaczne z jednorodnością wariancji.

Dużo większym stopniem ogólności charakteryzują się testy niekonstruktywne pozwalające na stwierdzenie występowania zjawiska heteroskedastyczności składników losowych (w przypadku odrzucenia hipotezy zerowej i przyjęcia alternatywnej) lub wnioskowanie o homoskedastyczności (w sytuacji przeciwnej). Wadą tego rodzaju postępowania jest natomiast niewiedza w zakresie kształtowania się wariancji w przypadku jej niejednorodności, co uniemożliwia dalsze postępowanie.

Przedstawione w dalszej części pracy rozwiązanie wykorzystuje zalety testu niekonstruktywnego oraz przezwycięża dalsze trudności przez zastosowanie analizy dyskryminacyjnej do określenia modelu heteroskedastyczności.

3. Propozycja wykrywania rodzaju heteroskedastyczności za pomocą dyskryminacyjnej reguły największej wiarygodności

W literaturze (por. [Welfe 1998, s. 117]) można spotkać pogląd, że wybór modelu heteroskedastyczności powinien opierać się na „wiedzy dotyczącej modelowanych zjawisk”. Podejście tego rodzaju bazuje na subiektywnej ocenie badacza, o którym zakłada się, że dysponuje niezbędną wiedzą umożliwiającą dokonanie prawidłowej selekcji na podstawie przesłanek merytorycznych. Wydaje się, że w wielu sytuacjach spełnienie tego rodzaju warunków nie jest możliwe. Doświadczenie badacza oraz wnikliwa znajomość badanych zjawisk są zawsze bardzo pomocne i pożądane, lecz podejmowanie decyzji na podstawie subiektywnych wrażeń może być obciążone dużym błędem.

Sugerowane autorskie rozwiązanie stanowi połączenie podejścia klasycznego oraz zastosowania analizy dyskryminacyjnej. Procedura rozpoznawania rodzaju heteroskedastyczności składników losowych przebiega wieloetapowo. W pierwszym etapie należy wykonać test niekonstruktywny. Jeśli hipoteza o homoskedastyczności zostanie odrzucona, to proponuje się możliwe modele opisujące kształtowanie się wariancji i szacuje ich parametry. Na podstawie wyestymowanych modeli wyznacza się wartości elementów diagonalnych macierzy kowariancji składników losowych. Następnie, posługując się dyskryminacyjną regułą największej wiarygodności, dokonuje się wyboru optymalnego wariantu.

Odrzucenie hipotezy o homoskedastyczności składników losowych w standardowym modelu liniowym oznacza, że macierz kowariancji składników losowych ma postać $\sigma^2 \mathbf{V}$, przy czym $\mathbf{V} \neq \mathbf{I}$. Przy założeniu normalności rozkładu funkcja gęstości dla składników losowych ma postać:

$$f(\boldsymbol{\varepsilon}) = \frac{1}{(2\pi)^n |\sigma^2 \mathbf{V}|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} \boldsymbol{\varepsilon}^T (\sigma^2 \mathbf{V})^{-1} \boldsymbol{\varepsilon} \right\}. \quad (2)$$

Funkcja (2) będzie miała różną postać, w zależności od przyjętego modelu heteroskedastyczności, pociągającego za sobą określoną konstrukcję macierzy $\sigma^2 \mathbf{V}$.

Problem decyzyjny, który pojawia się po stwierdzeniu istnienia niejednorodności wariancji za pomocą testu niekonstruktywnego, może zostać rozwiązany dzięki analizie dyskryminacyjnej. Otóż, wiadomo wówczas, że funkcja gęstości ma postać (2), ale nieznany jest model heteroskedastyczności, co prowadzi do nieznajomości macierzy kowariancji składników losowych. Jak zaznaczono wcześniej, wybór schematu heteroskedastyczności na podstawie wiedzy badacza o analizowanych zjawiskach jest metodą skrajnie subiektywną. Alternatywnym postępowaniem do podejścia intuicyjnego może być zaproponowanie g modeli opisujących zachowanie się wariancji, a następnie dokonanie wyboru najlepszego z nich.

Dla każdego proponowanego modelu znana jest postać macierzy $\mathbf{V}$, której elementy diagonalne można oszacować na podstawie modelu wyjaśniającego kształt-

townie się wariancji, wykorzystując do tego wartości bezwzględne lub kwadraty reszt uzyskanych przy szacowaniu parametrów modelu metodą najmniejszych kwadratów.

Posiadając g postaci macierzy V dla g rozpatrywanych możliwości, można określić g funkcji gęstości o postaci (2), odpowiadających poszczególnym modelom heteroskedastyczności składników losowych.

Zagadnienie wyboru schematu opisującego kształtowanie się wariancji spośród g proponowanych możliwości można rozwiązać, korzystając z metod analizy dyskryminacyjnej. Rozpatrywanych jest g modeli ($g \geq 2$), w których liczba g jest skończona i z góry znana. Dla każdej z możliwości istnieje n -wymiarowy rozkład wektora składników losowych ϵ . Dla każdego wariantu ($i=1, \dots, g$) określona jest funkcja gęstości $f_i(\epsilon)$ charakteryzująca ten rozkład.

Należy skonstruować regułę dyskryminacyjną, na podstawie której wielowymiarowy wektor ϵ zostanie przydzielony do jednej i tylko jednej z g populacji.

Tabela 1. Etapy postępowania mającego na celu wykrycie typu heteroskedastyczności

Etap	Cel	Postępowanie
1	Wykrycie, czy w modelu istnieje heteroskedastyczność składników losowych	Zastosowanie jednego z niekonstruktywnych testów homoskedastyczności
2	Określenie zbioru możliwych modeli heteroskedastyczności dla danego modelu	Zaproponowanie rozwiązań opisanych w literaturze bądź własnych. Elementami wspomagającymi mogą być: wiedza merytoryczna o badanych zjawiskach, analiza reszt otrzymanych na podstawie modelu oszacowanego metodą najmniejszych kwadratów
3	Wyznaczenie elementów macierzy kowariancji składników losowych dla zaproponowanych modeli heteroskedastyczności	W zależności od postaci zaproponowanych modeli heteroskedastyczności: 1. Bezpośrednie wykorzystanie wybranej zmiennej objaśniającej lub jej funkcji 2. Estymacja parametrów modeli heteroskedastyczności metodą najmniejszych kwadratów i wyznaczenie na ich podstawie wartości teoretycznych (oszacowań wariancji)
4	Wybór najlepszego z zaproponowanych modeli	Użycie analizy dyskryminacyjnej – obliczenie wartości funkcji wiarygodności dla poszczególnych modeli heteroskedastyczności i znalezienie wartości maksymalnej
5	Estymacja parametrów modelu	Zastosowanie uogólnionej metody najmniejszych kwadratów lub uogólnionej metody najmniejszych kwadratów z estymacją z wykorzystaniem macierzy kowariancji zdefiniowanej w modelu heteroskedastyczności charakteryzującym się największą wartością funkcji wiarygodności

Źródło: opracowanie własne.

Wśród reguł dyskryminacyjnych opisanych w literaturze (zob. np. [Anderson 2003; Jajuga 1990; Marida, Kent, Bibby 1979; McLachlan 1992]) na uwagę zasługuje dyskryminacyjna reguła największej wiarygodności niewymagająca apriorycznej wiedzy o populacjach, a stosowana w sytuacjach, gdy nie są znane koszty błędnej alokacji oraz prawdopodobieństwa *a priori* przynależności do klas i nie ma możliwości ich określenia. W przypadku badania założeń modelu ekonometrycznego z tego rodzaju sytuacją badacz ma do czynienia najczęściej. Zazwyczaj nie jest on w stanie określić prawdopodobieństw *a priori* wystąpienia poszczególnych modeli heteroskedastyczności, nie zna też kosztów błędnej alokacji. Oczywiście w momencie posiadania takiej wiedzy istnieje możliwość jej wykorzystania przez konstrukcję bayesowskich reguł dyskryminacyjnych. W pozostałych okolicznościach, występujących w praktyce bardzo często, przyjmuje się występowanie równych strat i równych prawdopodobieństw *a priori* zgodnie z postulatem Bayesa zakładającym, że w przypadku braku informacji przeczącej wszystkie prawdopodobieństwa *a priori* są jednakowe. Możliwa wówczas do stosowania dyskryminacyjna reguła największej wiarygodności przydziela wektor ϵ do tej populacji, dla której funkcja wiarygodności $L_i(\epsilon)$ ($i=1, \dots, g$) osiąga maksimum, czyli dla której funkcja określona wzorem (2) osiąga największą wartość wśród wszystkich g zaproponowanych możliwości. Ze względu na monotoniczność funkcji logarytmicznej wnioski można formułować również na podstawie wartości $\ln L_i(\epsilon)$.

Model heteroskedastyczności, któremu odpowiada funkcja wiarygodności o maksymalnej wartości, jest najlepszy wśród wszystkich zaproponowanych przez badacza rozwiązań. Powinien zostać wybrany jako najlepszy ze schematów określonych przez prowadzącego badanie i wykorzystany do estymacji parametrów modelu uogólnioną metodą najmniejszych kwadratów lub uogólnioną metodą najmniejszych kwadratów z estymacją.

Wszystkie etapy proponowanej procedury przedstawia tab. 1.

4. Rezultaty badania empirycznego

Postulowana metoda została wykorzystana do określania rodzaju heteroskedastyczności w jednoczynnikowych modelach ekonometrycznych, w których zmienną objaśnianą stanowiły wydatki *per capita* na różne cele, a zmienną objaśniającą były dochody *per capita* mieszkańców państw strefy euro. Badanie zostało zawężone do państw i lat, dla których były dostępne niezbędne dane w bazie Euromonitor, takich jak: Austria (1990, 1995, 2000, 2005), Belgia (1995, 2000, 2005), Finlandia (1990, 1995, 2000, 2005), Francja (1990, 1995, 2000, 2005), Grecja (1995, 2000), Hiszpania (2000, 2005), Holandia (1995, 2000, 2005), Irlandia (1995, 2000, 2005), Niemcy (1995, 2000, 2005), Portugalia (1990, 1995, 2000, 2005), Włochy (1995, 2000, 2005). W ramach wydatków rozpatrzono dwanaście grup towarów i usług zgodnych z klasyfikacją przychodów i rozchodów

stosowaną przez EUROSTAT, tzw. klasyfikacją spożycia indywidualnego według cech dla badań budżetów gospodarstw domowych, która wyróżnia takie kategorie, jak: żywność i napoje bezalkoholowe, napoje alkoholowe i wyroby tytoniowe, odzież i obuwie, użytkowanie mieszkania i nośniki energii, wyposażenie mieszkania i prowadzenie gospodarstwa domowego, zdrowie, transport, łączność, rekreacja i kultura, edukacja, restauracje i hotele, inne towary i usługi.

Korzystając z dostępnych danych, oszacowano metodą najmniejszych kwadratów parametry modeli opisujących zależność wydatków konsumpcyjnych poniesionych na określone towary i usługi od dochodów. Po wyznaczeniu wektora reszt zastosowano niekonstruktywny test Harrisona-McCabe'a (zob. [Harrison, McCabe 1979]) w celu sprawdzenia, czy występuje zjawisko heteroskedastyczności składników losowych. Na poziomie istotności 0,1 stwierdzono niejednorodność wariancji dla następujących kategorii: żywność i napoje bezalkoholowe, napoje alkoholowe i wyroby tytoniowe, wyposażenie mieszkania i prowadzenie gospodarstwa domowego, łączność, edukacja, rekreacja i kultura, inne towary i usługi. W kolejnym kroku zaproponowano sześć możliwych modeli opisujących zależność wariancji do wartości zmiennej objaśniającej (tj. dochodów):

- 1) $\sigma_i = \alpha_0 + \alpha_1 X$,
- 2) $\sigma_i^2 = \alpha_0 + \alpha_1 X$,
- 3) $\sigma_i^2 = \sigma_0^2 X^\alpha$,
- 4) $\sigma_i^2 = \sigma_0^2 \exp(\alpha X)$,
- 5) $\sigma_i = \alpha_0 + \alpha_1 X + \alpha_2 X^2$,
- 6) $\sigma_i^2 = \alpha_0 + \alpha_1 X + \alpha_2 X^2$.

Tabela 2. Wartości logarytmów naturalnych funkcji wiarygodności dla sześciu modeli heteroskedastyczności

Rodzaj modelu, w którym wystąpiła heteroskedastyczność	Wartości logarytmu funkcji wiarygodności					
	model 1	model 2	model 3	model 4	model 5	model 6
Ogólny	-29,770	-29,396	-102,134	-101,279	-29,688	-28,300
Żywność i napoje bezalkoholowe	2,839	14,587	-5,140	-4,944	2,836	15,434
Napoje alkoholowe i wyroby tytoniowe	10,444	27,661	-6,365	-6,088	10,502	27,607
Wyposażenie mieszkania i prowadzenie gospodarstwa domowego	10,697	28,081	13,136	8,030	9,693	27,422
Łączność	24,386	52,750	44,513	46,820	24,286	55,611
Rekreacja i kultura	4,178	19,384	1,417	0,197	4,309	18,864
Edukacja	37,869	79,379	64,720	66,128	38,475	77,816
Inne towary i usługi	0,243	11,729	-14,552	-16,871	0,242	10,586

Źródło: obliczenia własne.

Parametry modeli (1-6) zostały oszacowane metodą najmniejszych kwadratów, a ich wartości teoretyczne posłużyły do konstrukcji macierzy V. Oszacowanie macierzy kowariancji pozwoliło na zastosowanie dyskryminacyjnej reguły największej wiarygodności. Dla modeli (1-6) wyznaczono wartości funkcji wiarygodności

o postaci (2), korzystając z przybliżenia wektora składników losowych modelu za pomocą wektora reszt. Najlepszy schemat heteroskedastyczności dla konkretnego modelu wybrano, szukając wariantu o maksymalnej wartości funkcji wiarygodności (lub jej logarytmu). Otrzymane rezultaty prezentuje tab. 2.

Zastosowanie dyskryminacyjnej reguły największej wiarygodności prowadzi do wyboru najlepszego schematu heteroskedastyczności spośród zaproponowanych (w tab. 2 wyłuszczonego optymalny wariant). W analizowanym zagadnieniu do opisu zmian wariancji znajdują zastosowanie modele nr 2 lub nr 6. Wybór adekwatnego w danej sytuacji modelu heteroskedastyczności umożliwia zastosowanie uogólnionej metody najmniejszych kwadratów do szacowania parametrów.

Literatura

- Anderson T.W. (2003), *An Introduction to Multivariate Statistical Analysis*, Wiley, New York.
- Greene W. H. (1997), *Econometric Analysis*, Prentice Hall, Upper Saddle River, New Jersey.
- Harrison M.J., McCabe B.P.M. (1979), *A Test for Heteroscedasticity Based on Ordinary Least Squares Residuals*, „Journal of the American Statistical Association” vol. 74, s. 494-499.
- Jajuga K. (1990), *Statystyczna teoria rozpoznawania obrazów*, PWN, Warszawa.
- Mardia K.V., Kent J.T., Bibby J.M. (1979), *Multivariate Analysis*, Academic Press, London.
- McLachlan G.J. (1992), *Discriminant Analysis and Statistical Pattern Recognition*, Wiley, New York.
- Tomaszewicz A. (1985), *Jednorównaniowe modele ekonometryczne przy nieklasycznych założeniach*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź.
- Welfe W. (1998), *Ekonometria*, PWE, Warszawa.

RECOGNITION OF THE HETEROSCEDASTICITY TYPE OF DISTURBANCES OF THE ECONOMETRIC MODEL BY APPLYING THE MAXIMUM LIKELIHOOD DISCRIMINANT RULE

Summary

This paper presents a method of determination of the heteroscedasticity type in the linear econometric model. The postulated concept combines advantages of the non-constructive test with the usage of the maximum likelihood discriminant rule. The suggested solution is illustrated by models describing the dependence of expenses and income.