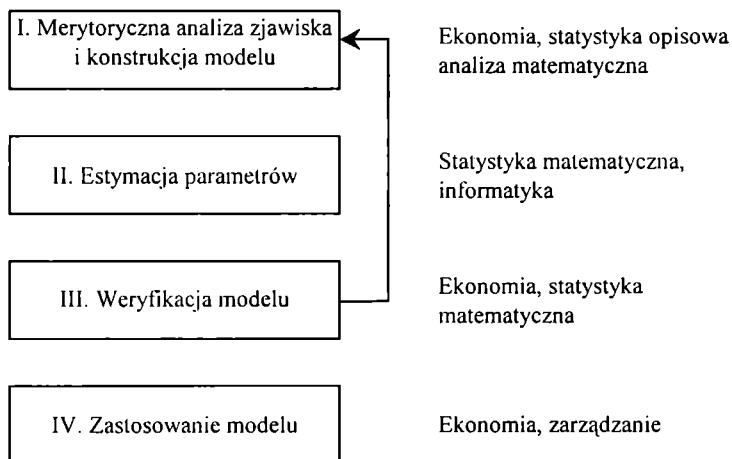


Alicja Grześkowiak

WYKORZYSTANIE REGUŁ DYSKRYMINACYJNYCH DO OKREŚLANIA TYPU HETEROSKEDASTYCZNOŚCI W LINIOWYM MODELU EKONOMETRYCZNYM*

1. Wstęp

Modelowanie ekonometryczne przebiega wieloetapowo, lecz w literaturze ekonometrycznej nie ma zgodności co do liczby etapów, gdyż różni autorzy operują różnym stopniem szczegółowości. Podział modelowania na cztery fazy proponowany przez Gruszczyńskiego i Podgórką [3, s. 13], przedstawia rys. 1.



Rys. 1. Etapy modelowania ekonometrycznego

Źródło: [3, s. 13].

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

Mimo różnego stopnia uszczegółowienia kolejnych faz idea modelowania pozostaje niezmienna. Po sformułowaniu celu badania, określeniu jego przedmiotu i zakresu ustala się, jakie zmienne objaśniane i objaśniające mogą służyć wyjaśnianiu mechanizmu kształtowania się rozpatrywanego zjawiska. Niniejsza praca dotyczy kroku III, czyli weryfikacji, w którym sprawdza się, czy abstrakcyjna konstrukcja, którą stanowi model, jest dostatecznie zgodna z fragmentem rzeczywistości, który opisuje. Pozytywne wnioski z tego etapu stanowią warunek konieczny aplikacji modelu w praktyce. W razie wystąpienia niezgodności z założeniami istnieje możliwość korygowania błędów po ustaleniu przyczyn ich występowania. Jedną z części procedury weryfikacji stanowi badanie założeń dotyczących rozkładu składników losowych. Ponieważ są one wielkościami nieobserwowalnymi, istnieją trudności w formułowaniu wniosków na ich podstawie. W ujęciu klasycznym stosuje się w tym celu testy statystyczne, które nie zawsze prowadzą do rozstrzygnięcia problemu. W niniejszej pracy proponuje się połączenie wnioskowania opartego na testach z zastosowaniem analizy dyskryminacyjnej do rozwiązywania zagadnienia związanego z występowaniem heteroskedastyczności składników losowych. Celem pracy jest zaprezentowanie postulowanej procedury. Sugerowane rozwiązanie zostało zilustrowane wynikami badania empirycznego.

2. Koncepcja zastosowania reguły dyskryminacyjnej do wykrywania rodzaju heteroskedastyczności składników losowych

Założenie klasycznej metody najmniejszych kwadratów dotyczące jednorodności wariancji składników losowych weryfikuje się za pomocą testów statystycznych opartych na resztach modelu. Jeżeli podczas tego etapu modelowania nastąpi wykrycie niesferyczności, pojawi się konieczność zmiany metody estymacji parametrów modelu na uogólnioną metodę najmniejszych kwadratów, która wymaga znajomości postaci macierzy kowariancji składników losowych (bądź jej oszacowania).

W literaturze proponuje się różne testy umożliwiające wykrywanie heteroskedastyczności, a ich przegląd można znaleźć np. w pracach [1; 8]. Istnieją dwa odmienne sposoby przeprowadzania testów weryfikujących hipotezę o równości wariancji. Pierwsza możliwość to zastosowanie testu konstruktywnego, w którym rodzaj heteroskedastyczności jest określony w hipotezie alternatywnej. W razie jej przyjęcia można od razu sformułować wniosek o postaci macierzy kowariancji i metodzie estymacji. Drugi rodzaj testów charakteryzuje się większym stopniem ogólności – pozwala na stwierdzenie, czy heteroskedastyczność istnieje, lecz nie daje żadnych wskazówek co do jej kształtowania się.

Zastosowanie testu konstruktywnego jest z jednej strony bardzo dogodnym postępowaniem – przeprowadzając test, od razu otrzymuje się wskazówki dotyczące ewentualnego dalszego postępowania. Z drugiej strony stanowi bardzo znaczne

zawężenie problemu – nieodrzućcenie hipotezy zerowej pozwala wnioskować jedynie, że schemat niejednorodności wariancji określony w tym teście nie występuje, natomiast nie uprawnia do sformułowania sądu o jednorodności wariancji.

Z kolei testy niekonstruktywne pozwalają na stwierdzenie występowania zjawiska heteroskedastyczności składników losowych (gdy przyjmuje się hipotezę alternatywną) lub wnioskowania o homoskedastyczności (gdy nie odrzuca się hipotezy zerowej). Niedogodnością tego rodzaju postępowania jest natomiast brak informacji o kształtowaniu się wariancji w przypadku jej niejednorodności, co uniemożliwia zastosowanie uogólnionej metody najmniejszych kwadratów.

Wskazane byłoby zastosowanie narzędzia opartego na obiektywnym kryterium, pomocnego w dokonaniu wyboru rodzaju heteroskedastyczności, po uprzednim stwierdzeniu jej występowania za pomocą testu konstruktywnego. Do rozwiązania tego problemu może zostać zastosowana analiza dyskryminacyjna. Jeśli hipoteza o homoskedastyczności zostanie odrzucona, należy zaproponować możliwe modele opisujące kształtowanie się wariancji i oszacować ich parametry. Posiadając oszacowania elementów diagonalnych macierzy kowariancji składników losowych, wykonuje się drugą część procedury, polegającą na zastosowaniu reguły dyskryminacyjnej, w celu wyboru „najlepszego” wariantu.

Odrzućcenie hipotezy o sferyczności składników losowych w standardowym modelu liniowym oznacza, że macierz kowariancji składników losowych jest nieskalarna, postaci $\sigma^2 \mathbf{V}$. Przy założeniu normalności rozkładu funkcja gęstości wyraża się wzorem:

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

Funkcja (1) będzie miała różną postać, w zależności od przyjętego modelu heteroskedastyczności, pociągającego za sobą określoną konstrukcję macierzy $\sigma^2 \mathbf{V}$. Modele heteroskedastyczności proponowane w literaturze zestawiono w tab. 1.

Dla każdego proponowanego modelu można określić postać macierzy $\mathbf{V}$. W zależności od przyjętego schematu można ją wyznaczyć bezpośrednio lub oszacować niezbędne elementy na podstawie modelu wyjaśniającego kształtowanie się wariancji, wykorzystując do tego wartości bezwzględne lub kwadraty reszt uzyskanych przy szacowaniu parametrów modelu metodą najmniejszych kwadratów. Znając postaci macierzy $\mathbf{V}$ dla g rozpatrywanych możliwości, można określić g funkcji gęstości postaci odpowiadających poszczególnym rodzajom heteroskedastyczności składników losowych.

Zgodnie z dyskryminacyjną regułą największej wiarygodności (zob. [5]) wektor $\boldsymbol{\varepsilon}$ należy przydzielić do rozkładu o numerze j , dla którego wiarygodność L jest maksymalna:

Tabela 1. Modele heteroskedastyczności składników losowych modelu ekonometrycznego

Postać modelu	Charakterystyka modelu	Literatura
$\sigma_i^2 = \sigma^2 X_{im} $ lub $\sigma_i^2 = \sigma^2 X_{im}^2 $	wariancja proporcjonalna do wartości bezwzględnych wybranej zmiennej objaśniającej lub jej kwadratów	Johnston Judge <i>et al.</i> Greene
$\sigma_i^2 = \sigma^2 E(Y)$ lub $\sigma_i^2 = \sigma^2 E^2(Y)$	wariancja proporcjonalna do wartości oczekiwanej zmiennej objaśnianej lub jej kwadratów	Prais i Houthakker Amemiya Theil
$\sigma_i = \alpha_0 + \alpha_1 E(Y)$	odchylenie standardowe zależy liniowo od wartości oczekiwanych zmiennej objaśnianej	Geary
$\sigma_i^2 = \alpha^T z_i$	addytywny model heteroskedastyczności – wariancja zależy liniowo od zmiennych objaśniających modelu heteroskedastyczności	Amemiya Jobson i Fuller Buse Amemiya
$\sigma_i^2 = \sigma^2 (\alpha^T z_i)^2$	odchylenie standardowe zależy liniowo od zmiennych objaśniających modelu heteroskedastyczności	Rutemiller i Bowers
$\sigma_i^2 = \alpha_0 + \alpha_1 Z_i + \alpha_2 Z_i^2 + \dots + \alpha_h Z_i^h$	wariancja opisana za pomocą wielomianu zmiennej Z będącej funkcją zmiennej lub zmiennych objaśniających	Glejser
$\sigma_i^2 = \sigma^2 Z_i^\alpha$	multiplikatywny model heteroskedastyczności	Fisher Park
$\sigma_i^2 = \sigma^2 \exp(\alpha^T z_i)$	multiplikatywny model heteroskedastyczności (wykładniczy)	Rutemiller i Bowers Harvey

Źródło: opracowanie własne.

$$L_j(\epsilon) = \max_i L_i(\epsilon) \text{ dla } i = 1, 2, \dots, g \quad (2)$$

lub równoważnie:

$$f_j(\epsilon) = \max_i f_i(\epsilon) \text{ dla } i = 1, 2, \dots, g. \quad (3)$$

Jeżeli m funkcji wiarygodności ($2 \leq m \leq g$) osiąga maksymalną wartość przy ustalonym ϵ , to można wybrać którykolwiek z m rozkładów o największej wiarygodności.

W przypadku porównywania modeli charakteryzujących się normalnością rozkładu składników losowych zagadnienie sprowadza się do wyboru tego wariantu, dla którego funkcja określona wzorem (1) osiąga największą wartość wśród wszystkich g zaproponowanych możliwości.

Ze względu na własność monotoniczności funkcji logarytmicznej można przedstawić postępowanie równoważne do przedstawionego powyżej. Postępując zgodnie z dyskryminacyjną regułą największej wiarygodności, wektor ϵ przydziela się do rozkładu o numerze j , dla którego jest spełniony warunek (4) lub (5):

$$\ln L_j(\epsilon) = \max_i [\ln L_i(\epsilon)] \text{ dla } i = 1, 2, \dots, g, \quad (4)$$

$$\ln f_j(\epsilon) = \max_i [\ln f_i(\epsilon)] \text{ dla } i = 1, 2, \dots, g. \quad (5)$$

Przedstawiona idea może zostać zaadaptowana do wyboru najlepszego spośród zaproponowanych modeli heteroskedastyczności. Modyfikacja koncepcji polega na porównywaniu wiarygodności poszczególnych modeli z wiarygodnością wektora ϵ , otrzymaną z zastosowaniem estymatora macierzy kowariancji składników losowych proponowanego przez Domowitza i White'a. W pracy poświęconej m.in. poszukiwaniu metody szacowania macierzy kowariancji estymatora metody najmniejszych kwadratów (zob. [2]) przedstawiono konstrukcję estymatora zgodnego, do budowy którego wykorzystuje się oszacowanie macierzy kowariancji składników losowych, które może również znaleźć zastosowanie w konstrukcji reguł dyskryminacyjnych. Zasadniczym problemem jest to, iż elementy macierzy kowariancji nie są znane. Jakubczyc [4, s. 143] zauważa: „Wyłania się zatem potrzeba wyznaczenia estymatora tej macierzy. Może się wydawać, że z rozwiązaniem tego zagadnienia nie powinno być większych problemów. Macierz utworzoną następująco:

$$ee^T = \begin{bmatrix} e_1^2 & e_1e_2 & \dots & e_1e_n \\ e_2e_1 & e_2^2 & \dots & e_2e_n \\ \vdots & \vdots & & \vdots \\ e_ne_1 & e_ne_2 & \dots & e_n^2 \end{bmatrix} \quad (6)$$

można bowiem potraktować jako estymator macierzy V . Główna trudność tkwi jednak w tym, że rząd macierzy ee^T jest równy jedności, a więc nie istnieje macierz odwrotna $(ee^T)^{-1}$, wobec czego nie ma żadnych szans na to, aby w tak prosty sposób można było sformułować estymator macierzy”.

Jednakże okazuje się, że ograniczenie wynikające z zależności liniowych kolumn (i wierszy) macierzy (6) można stosunkowo łatwo zniwelować. We wspomnianej wyżej pracy Domowitza i White'a zaproponowano rozwiązanie wywodzące się z macierzy postaci (6), eliminujące problem jej osobliwości. Sugerowana idea opiera się na swoistym „ucinianiu” macierzy (6) doprowadzającym do postaci $\Sigma(\epsilon)_{DW}$ o elementach równych:

$$s_{tk} = \begin{cases} e_t e_k & \text{dla } |t - k| \leq 1, \\ 0 & \text{dla } |t - k| > 1. \end{cases}$$

Wybór wartości l stanowi dla twórców omawianej koncepcji „interesującą otwartą kwestię”. Niemniej jednak autorzy wskazują, że l powinno się zwiększać w miarę wzrostu liczby obserwacji, lecz wolniej niż $n^{1/3}$.

Model heteroskedastyczności, którego funkcja wiarygodności różni się najmniej od funkcji wiarygodności wektora ε otrzymanej z użyciem macierzy $\Sigma(\varepsilon)_{DW}$, jest najlepszy wśród wszystkich zaproponowanych przez badacza rozwiązań. Powinien zostać wybrany jako najlepszy spośród schematów określonych przez prowadzącego badanie i wykorzystany do estymacji parametrów modelu uogólnioną metodą najmniejszych kwadratów lub uogólnioną metodą najmniejszych kwadratów z estymacją.

3. Rezultaty badania empirycznego

Przeprowadzone badanie dotyczyło modeli opisujących zależności wydatków na różne towary i usługi od dochodów. Analizą objęto grupę państw europejskich należących do strefy euro, dla których były dostępne dane dotyczące przeciętnych dochodów i wydatków w bazie *Euromonitor*. Zestawienie krajów i lat, dla których możliwe było uzyskanie danych statystycznych, prezentuje tab. 2.

Tabela 2. Państwa, dla których przeprowadzono analizę zależności dochodów i wydatków

Państwo	Lata, dla których były dostępne dane
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	1990, 1995, 2000, 2005
Irlandia	1990, 1995, 2000, 2005
Niemcy	1995, 2000, 2005
Portugalia	1990, 1995, 2000, 2005
Włochy	1995, 2000, 2005

Źródło: opracowanie własne.

Wybór krajów ze strefy euro gwarantuje porównywalność danych – zarówno dochody, jak i wydatki podane są we wspólnej walucie. Stopień szczegółowości dostępnych informacji odnośnie do wydatków obejmuje dwanaście grup towarów i

usług zgodnych z klasyfikacją przychodów i rozchodów stosowaną przez Europejski Urząd Statystyczny EUROSTAT, a opartą na *COICOP/HBS (Classification of Individual Consumption by Purpose/Households Budgets Survey)* – Klasyfikacji Spożycia Indywidualnego według Cech dla Badań Budżetów Gospodarstw Domych. Na dwanaście podstawowych grup towarów i usług, według których prezentowane są informacje o wydatkach, składają się następujące kategorie:

- 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.

W pierwszym etapie badania metodą najmniejszych kwadratów oszacowano linowe zależności wydatków ponoszonych na różne kategorie dóbr i usług od dochodów osiąganych przez gospodarstwa domowe. W rozpatrywanych modelach można spodziewać się wystąpienia zjawiska heteroskedastyczności składników losowych. Modele budowane są na bazie danych przekrojowo-czasowych, co może stanowić źródło niejednorodności wariancji. Przeprowadzenie niekonstruktywnego testu heteroskedastyczności Harrisona-McCabbe'a ujawniło występowanie niejednorodności wariancji w modelach 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 (poziom istotności 0,1).

Do opisu kształtowania się wariancji składników losowych zaproponowano rozważenie następujących modeli:

$$1. \sigma_i = \alpha_0 + \alpha_1 X,$$

$$2. \sigma_i^2 = \alpha_0 + \alpha_1 X,$$

$$3. \sigma_i^2 = \sigma_0^2 X^\alpha,$$

$$4. \sigma_i^2 = \sigma_0^2 \exp(\alpha X),$$

$$5. \sigma_i = \alpha_0 + \alpha_1 X + \alpha_2 X^2,$$

$$6. \sigma_i^2 = \alpha_0 + \alpha_1 X + \alpha_2 X^2.$$

Parametry modeli (1-6) zostały oszacowane metodą najmniejszych kwadratów, a ich wartości teoretyczne posłużyły do budowy głównej przekątnej macierzy V.

Do konstrukcji funkcji dyskryminacyjnych zastosowano macierze $\Sigma(\epsilon)_{DW}$ i V oraz wektor reszt, traktując go jako dobre przybliżenie wektora składników losowych. Rezultaty otrzymane na podstawie takiego postępowania prezentuje tab. 3. Wyłuszczoneym drukiem zaznaczono wartości minimalne, wskazujące na odpowiedność danego modelu do opisu zjawiska heteroskedastyczności w konkretnym przypadku.

Tabela 3. Wartości bezwzględne różnic pomiędzy wartością logarytmu przybliżonej funkcji gęstości wektora ϵ a wartościami funkcji wiarygodności dla modeli 1-6

Rodzaj modelu, w którym wystąpiła heteroskedastyczność	Moduł różnicy logarytmów					
	model 1	model 2	model 3	model 4	model 5	model 6
Ogólny	91,377	91,003	163,741	162,887	91,296	89,908
Żywność i napoje bezalkoholowe	82,584	70,835	90,562	90,366	82,586	69,988
Napoje alkoholowe i wyroby tytoniowe	92,084	74,867	108,894	108,616	92,027	74,922
Wyposażenie mieszkania i prowadzenie gospodarstwa domowego	91,673	74,289	89,234	94,341	92,678	74,949
Łączność	97,820	69,456	77,693	75,387	97,920	66,595
Rekreacja i kultura	86,365	71,159	89,125	90,346	86,234	71,679
Edukacja	111,299	69,789	84,447	83,040	110,693	71,351
Inne towary i usługi	85,465	73,979	100,260	102,579	85,466	75,123

Źródło: obliczenia własne.

Aplikacja rozpatrywanej metody ma na celu selekcję najlepszego modelu opisującego kształtowanie się wariancji składników losowych z zestawu proponowanych możliwości. Po zastosowaniu macierzy $\Sigma(\epsilon)_{DW}$ i implementacji sugerowanej procedury można wnioskować, że do opisu zmian wariancji składników losowych rozpatrywanych relacji dochodów i wydatków znajdują zastosowanie modele nr 2 lub nr 6.

4. Podsumowanie

Prezentowana koncepcja stanowi uzupełnienie podejścia klasycznego, w którym o rodzaju heteroskedastyczności wnioskuje się na podstawie testów statystycznych. Przewaga sugerowanego podejścia polega na umożliwieniu badaczowi korzystania z jak najbardziej ogólnego testu wykrywającego niejednorodność wariancji bez utraty możliwości określenia jej charakteru. Wybór odpowiedniego modelu następuje dzięki porównaniu wiarygodności dla poszczególnych modeli z wiarygodnością wektora składników losowych, możliwą do wyznaczenia dzięki zastosowaniu estymatora macierzy kowariancji proponowanego przez Domowitza i White'a.

Literatura

- [1] Adijolosoo S.B.-S.K., *Estimation of Parameters of Heteroscedastic Error Models Under Various Hypothesized Error Structures*, „The Statistician” 1993 vol. 42, s. 123-133.
- [2] Domowitz I., White H., *Nonlinear Regression with Dependent Observations*, „Econometrica” 1984 vol. 52, s. 143-161.
- [3] *Ekonometria*, red. M. Gruszczyński, M. Podgórska. Oficyna Wydawnicza Szkoły Głównej Handlowej, Warszawa 2000.
- [4] Jakubczyc J., *Jednorównaniowe modele ekonometryczne*, PWE, Warszawa 1982.
- [5] Mardia K.V., Kent J.T., Bibby J.M., *Multivariate Analysis*, Academic Press, London 1979.
- [6] McLachlan G.J., *Discriminant Analysis and Statistical Pattern Recognition*, Wiley, New York 1992.
- [7] Theil H., *Zasady ekonometrii*, PWN, Warszawa 1979.
- [8] Tomaszewicz A., *Jednorównaniowe modele ekonometryczne przy nieklasycznych założeniach*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź 1985.

THE APPLICATION OF DISCRIMINANT RULES TO THE DETERMINATION OF HETEROSCEDASTICITY TYPE IN THE LINEAR ECONOMETRIC MODEL

Summary

This paper considers the utility of the application of discriminant functions to determine the best possible heteroscedasticity model of disturbances. Likelihood functions of certain models were proposed to determine the solution which was gained by conducting a comparison with the likelihood function of the vector ϵ obtained with the usage of covariance matrix estimator suggested by Domowitz and White.

Alicja Grześkowiak – dr. asystentka w Katedrze Ekonometrii Akademii Ekonomicznej we Wrocławiu.