

Mariusz Doszyń

## TESTOWANIE EGZOGENICZNOŚCI ZMIENNYCH W MODELACH EKONOMETRYCZNYCH

Od pewnego czasu w literaturze ekonometrycznej pojawiają się głosy krytyczne dotyczące **podejścia Fundacji Cowlesa**, zgodnie z którym podział zmiennych na endogeniczne i egzogeniczne oraz przyczynowa struktura modelu są znane *a priori*. G.S. Maddala wskazuje na trzy główne przyczyny krytyki tych założeń<sup>1</sup>.

W wielu przypadkach podział zmiennych na endogeniczne i egzogeniczne jest arbitralny. Kolejny argument, zwany **krytyką Liu**, dotyczy faktu, że w celu uzyskania identyfikacji często pomija się wiele zmiennych, które powinny być uwzględniane. Trzecia przyczyna dotyczy tzw. **krytyki Lucasa**, który stwierdził, że parametry modeli wielorównaniowych często są zależne od zmian zmiennych egzogenicznych.

Przed oszacowaniem modelu:

$$y_t = \beta_0 + \beta_1 x_t + u_t \quad (1)$$

gdzie  $u_t$  – składnik losowy,

należy rozstrzygnąć, czy zmienna  $x_t$  może być traktowana jako zmienna egzogeniczna. Jeśli zmiennej  $x_t$  nie można traktować jako zmiennej egzogenicznej, należy skonstruować model wielorównaniowy.

---

<sup>1</sup> [3].

## 1. Charakterystyka wybranych koncepcji egzogeniczności

Podstawowe dwie koncepcje dotyczące egzogeniczności to ustalenie z góry oraz ścisła egzogeniczność [3]. Zmienną można traktować jako z góry ustaloną w danym równaniu, jeśli jest niezależna od bieżących i przyszłych wartości składnika losowego rozważanego równania. Jeżeli określona zmienna jest niezależna od przeszłych, bieżących i przyszłych wartości składnika losowego w analizowanym równaniu, to jest to zmienna ściśle egzogeniczna.

G.S. Maddala podaje następujący przykład modelu ze zmiennymi opóźnionymi<sup>2</sup>:

$$y_t = \alpha_1 x_t + \beta_{11} y_{t-1} + \beta_{12} x_{t-1} + u_{1t} \quad (2)$$

$$x_t = \alpha_2 y_t + \beta_{21} y_{t-1} + \beta_{22} x_{t-1} + u_{2t} \quad (3)$$

gdzie  $u_{1t}$ ,  $u_{2t}$  – nie podlegają autokorelacji i są niezależne.

Jeśli  $\alpha_2 = 0$ , to zmienną  $x_t$  można traktować jako zmienną z góry ustaloną w równaniu (2). Nie zależy ona bowiem od bieżących i przyszłych wartości  $u_{1t}$ . Jeśli  $\alpha_2 = 0$  i  $\beta_2 = 0$ , to zmienna  $x_t$  jest zmienną ściśle egzogeniczną, która nie zależy wtedy od bieżących, przyszłych i przeszłych wartości  $u_{1t}$ .

W literaturze można spotkać również takie koncepcje egzogeniczności, jak słaba egzogeniczność, superegzogeniczność i silna egzogeniczność, które zostały zaproponowane przez Engle'a, Hendry'ego i Richarda [1]. Z koncepcją silnej egzogeniczności wiąże się z kolei tak zwana przyczynowość w sensie Grangera.

Zmienna jest określana jako słabo egzogeniczna z punktu widzenia estymacji danego zbioru parametrów  $A$ , jeśli wnioskowanie o zbiorze  $A$  warunkowe względem  $x_t$  nie wiąże się z utratą informacji [3]. Jeśli dwuwymiarowy rozkład zmiennych zapiszemy w postaci:  $f(y_t, x_t) = g(y_t/x_t) h(x_t)$ , gdzie  $g(y_t/x_t)$  zawiera parametry ze zbioru  $A$ , to z warunku słabej egzogeniczności wynika, że parametry te nie występują w rozkładzie brzegowym  $h(x_t)$ . W modelu (1) słaba egzogeniczność miałyby

<sup>2</sup> *Ibidem*, s. 429.

miejsce wtedy, gdyby zbiór parametrów rozkładu zmiennej  $x_t$  był rozłączny względem zbioru parametrów tego modelu.

Zmienna  $x_t$  jest określana jako superegzogeniczna, jeśli występuje słaba egzogeniczność, a parametry funkcji  $g(y_t/x_t)$  są niezmiennicze względem zmian w  $h(x_t)$ . W modelu (1) zmienna  $x_t$  może często nie być superegzogeniczna, ponieważ zmiany parametrów rozkładu  $h(x_t)$  zazwyczaj współwystępują ze zmianami ocen parametrów w modelu (1).

Jeżeli  $x_t$  jest zmienną słabo egzogeniczną i nie poprzedza jej żadna zmienna endogeniczna, to mamy do czynienia ze zmienną silnie egzogeniczną. Przykładem zmiennej silnie egzogenicznej jest zmienna  $x_t$  w następującym modelu:

$$y_t = \beta x_t + u_{1t} \quad (4)$$

$$x_t = \alpha x_{t-1} + u_{2t} \quad (5)$$

gdzie:

$(u_{1t}, u_{2t})$  – składniki losowe niepodlegające autokorelacji i mające dwuwymiarowy rozkład normalny;

$\sigma_{11}, \sigma_{22}$  – wariancje składników losowych  $u_{1t}$  i  $u_{2t}$ ,

$\sigma_{12}$  – kowariancja między  $u_{1t}$  i  $u_{2t}$ .

W przypadku gdy  $\sigma_{12} = 0$ , zmienna  $x_t$  jest słabo egzogeniczna (w rozkładzie brzegowym  $x_t$  nie występują parametry  $\beta$  i  $\sigma_{11}$ ). Zmienna  $x_t$  jest również silnie egzogeniczna, ponieważ w równaniu (5) nie występuje opóźniona zmienna  $y_t$ .

Definicja silnej egzogeniczności wiąże się z tak zwaną przyczynowością w sensie Grangera. Zmienna  $X_t$  nie jest przyczyną zmiennej  $y_t$  w sensie Grangera, jeśli  $\beta_i = 0$  ( $i = 1, 2, \dots, k$ ) w następującym równaniu regresji względem opóźnionych zmiennych [3]:

$$y_t = \sum_{i=1}^k \alpha_i y_{t-i} + \sum_{i=1}^k \beta_i x_{t-i} + u_t \quad (6)$$

Test ten nie jest zbyt użyteczny jako test egzogeniczności. Z braku przyczynowości w sensie Grangera niekoniecznie musi wynikać ścisła egzogeniczność lub fakt, że zmienna  $x_t$  jest z góry ustalona.

W kontekście rozważań dotyczących egzogeniczności interesujące jest podejście zaproponowane prof. J. Hozera, zgodnie z którym w ekonomii mamy do czynienia z trzema rodzajami związków między zdarzeniami [2]. Pomędzy zmiennymi mogą występować związki przyczynowe, celowe i współlistnienia (tabela 1).

Tabela 1. Rodzaje związków między zdarzeniami

Lp.	Rodzaje	Zmienna objaśniana	Zmienna objaśniająca	Rodzaj pytania	Rodzaj odpowiedzi
1.	Przyczynowy	$y_t$	$x_{t+\tau}$	dłaczego $y_t$ ?	dlatego, że $x_{t+\tau}$
2.	Celowy	$y_{t+\tau}$	$x_t$	po co $x_t$ ?	po to, aby $y_{t+\tau}$
3.	Współlistnienia	$y_t$	$x_t$	co z czym?	$y_t$ z $x_t$

Źródło: [2].

W przypadku procesów gospodarczych relacje o charakterze przyczynowym są zazwyczaj trudne do zidentyfikowania. Celowość wiąże się ze świadomą aktywnością człowieka. W ekonomii najczęściej mamy do czynienia z relacjami współlistnienia.

Na podstawie przeprowadzonych rozważań można stwierdzić, że egzogeniczność zmiennych nie zawsze jest oczywista i powinna podlegać testowaniu. Jednym z testów pozwalających na weryfikację hipotezy o egzogeniczności jest test błędu specyfikacji Hausmana.

## 2. Test błędu specyfikacji Hausmana

Hipoteza zerowa ( $H_0$ ) w teście Hausmana oznacza poprawną specyfikację modelu, natomiast hipoteza alternatywna ( $H_1$ ) – błędy specyfikacji. W przypadku modelu:

$$Y = \beta X + u \quad (7)$$

gdzie:

$X$  – macierz obserwacji dokonanych na zmiennych objaśniających (z wyłączeniem wyrazu wolnego),

$u$  – wektor składników losowych.

Odpowiednie hipotezy są formułowane w następujący sposób:

$H_0$  :  $X$  i  $u$  są niezależne,

$H_1$  :  $X$  i  $u$  są zależne.

W teście Hausmana konstruuje się dwa estymatory wektora parametrów  $\beta$ . Estymator  $\hat{\beta}_0$  jest zgodny i efektywny, jeśli  $H_0$  jest prawdziwa, lecz nie jest zgodny przy prawdziwości  $H_1$ . Estymator  $\hat{\beta}_1$  jest zgodny, jeśli  $H_0$  lub  $H_1$  jest prawdziwa, ale nie jest efektywny przy prawdziwości hipotezy zerowej<sup>3</sup>.

Jeżeli prawdziwa jest hipoteza  $H_1$ , estymator  $\hat{\beta}_0$  nie jest zgodny. W związku z tym dla uzyskania zgodnego estymatora  $\hat{\beta}_1$  stosuje się metodę zmiennych instrumentalnych:

$$\hat{\beta}_1 = (Z^T X)^{-1} Z^T Y \quad (8)$$

gdzie  $Z$  – macierz zmiennych instrumentalnych.

Statystyka testu Hausmana jest następująca<sup>4</sup>:

$$m = \hat{q}' [\hat{V}(\hat{q})]^{-1} \hat{q} \quad (9)$$

gdzie:

$$\hat{q} = \hat{B}_1 - \hat{B}_0; \quad \hat{V}(\hat{q}) = V_1 - V_0,$$

$V_0, V_1$  – macierze wariancji i kowariancji estymatorów parametrów  $\hat{\beta}_0$  i  $\hat{\beta}_1$ .

Przy wyznaczaniu  $V_1$  estymatorem  $\sigma^2$  może być estymator wariancji składnika losowego w modelu oszacowanym KMNK. Statystyka (9) ma rozkład  $\chi^2$  z  $k$  stopniami swobody.

Jeśli zmienne objaśniające w równaniu (7) nie są skorelowane ze składnikiem losowym, to można je traktować jako zmienne egzogeniczne. Właśnie to stwierdza  $H_0$ . Estymator KMNK jest w takim przypadku zgodny. Jeśli prawdziwa jest  $H_1$ , należy przyjąć, że zmienne objaśniające są endogeniczne (są powiązane ze składnikiem losowym). Estymator KMNK nie jest wtedy estymatorem zgodnym. Równanie (7) powinno być zatem traktowane jako jedno z równań modelu wielorównaniowego.

<sup>3</sup> *Ibidem*, s. 558.

<sup>4</sup> *Ibidem*.

Model (7) można również traktować jako model z błędami w zmiennych. Skorelowanie zmiennych objaśniających ze składnikiem losowym świadczy o tym, że zmienne zawierają w sobie błędy, co oznacza, iż prawdziwa jest  $H_1$ . Błędy w zmiennych nie występują, gdy  $H_0$  jest prawdziwa.

### 3. Przykład empiryczny

Oszacowano następujący model (w nawiasach pod ocenami parametrów podane są wartości statystyk  $t$ -Studenta):

$$\hat{y}_t = 123536 + 0,512x_t + 11965,4t \quad (10)$$

(4,874)            (7,913)            (8,140)

W modelu (10) zmienną objaśnianą jest spożycie indywidualne z dochodów osobistych w mln zł. Zmienne objaśniające to dochody do dyspozycji brutto w mln zł ( $x_t$ ) oraz zmienna czasowa ( $t$ ). Dane wykorzystane w przykładzie odnoszą się do sytuacji w Polsce w latach 1993–2005 i zostały zaczerpnięte z Rocznika Statystycznego GUS. Zmienne zostały wyrażone w cenach z 2005 roku. Do wyeliminowania wpływu zmian cen zastosowano wskaźnik cen spożycia indywidualnego z dochodów osobistych.

Wartości teoretyczne są bardzo dobrze dopasowane do wartości empirycznych (skorygowany współczynnik determinacji  $\bar{R}^2 = 0,997$ ). Wszystkie oceny parametrów modelu różnią się w sposób statystycznie istotny od zera, o czym świadczą podane wartości statystyk  $t$ -Studenta. Błąd standardowy reszt wyniósł  $S_e = 4950,42$  zł. Po przeprowadzeniu testu RESET, Jarque'a-Bery, Breuscha-Godfrey'a oraz testu White'a można stwierdzić, że w modelu (10) nie ma podstaw do odrzucenia hipotez zerowych, w których stwierdza się poprawną specyfikację modelu, normalność rozkładu reszt, brak autokorelacji reszt pierwszego i drugiego rzędu oraz homoskedastyczność reszt (poziom istotności  $\alpha = 0,05$ ).

Na podstawie modelu (10) można stwierdzić, że wzrost realnych dochodów do dyspozycji o 100,00 zł współwystępował ze wzrostem spożycia przeciętnie o 51,20 zł. Na skutek występowania inercji spożycie rosło z roku na rok przeciętnie o 11 965,40 zł. Ocenę wyrazu wolnego (123 536,00 zł) można traktować jako wydatki autonomiczne.

W kontekście rozważań o egzogeniczności pojawia się pytanie: czy zmienną  $x_t$  można traktować jako zmienną egzogeniczną? Hipoteza ta zostanie zweryfikowana na podstawie przedstawionego wcześniej testu Hausmana.

Rozważany model (10) oszacowano również metodą zmiennych instrumentalnych (MZI) po przyjęciu za zmienną instrumentalną wydatków inwestycyjnych ogółem  $i_t$  (w nawiasach podane są wartości statystyk t-Studenta):

$$\hat{y}_t = 142228 + 0,464 x_t + 13022 t \quad (11)$$

(2,930)                      (3,726)                      (4,710)

Wydatki inwestycyjne ogółem zostały „oczyszczone” z wpływu zmian cen za pomocą wskaźnika cen nakładów inwestycyjnych. Współczynnik korelacji między nakładami inwestycyjnymi oraz dochodami do dyspozycji brutto wynosi 0,783 i jest statystycznie istotny przy poziomie istotności 0,05.

Wszystkie oceny parametrów modelu (11) istotnie różnią się od zera. Błąd standardowy reszt wyniósł  $S_e = 5084,98$  zł. W modelu nie występuje autokorelacja reszt pierwszego rzędu (statystyka Durбина-Watsona  $d = 1,85$ ). Na podstawie testu Jarque’a-Bery stwierdzamy, iż nie ma podstaw do odrzucenia hipotezy zerowej zakładającej normalność rozkładu reszt modelu (poziom istotności  $\alpha = 0,05$ ).

W przypadku testowania egzogeniczności jednej zmiennej statystykę testu Hausmana (9) można zapisać następująco [3]:

$$m = \frac{\hat{q}^2 r^2}{(1 - r^2) \hat{V}_0} \quad (12)$$

gdzie:

- $\hat{q}$  – różnica między oceną parametru przy zmiennej  $x_t$  uzyskaną MZI oraz KMNK,
- $r^2$  – kwadrat współczynnika korelacji między zmienną  $x_t$  i zmienną instrumentalną  $x_p$ ,
- $\hat{V}_0$  – wariancja estymatora parametru przy zmiennej  $x_t$  oszacowanego KMNK.

Wartość statystyki  $m = 0,875$ . Wartość krytyczna odczytana z rozkładu  $\chi^2$  dla pierwszego stopnia swobody i poziomu istotności  $\alpha = 0,05$  wynosi 3,84. Nie ma zatem

podstaw do odrzucenia hipotezy zerowej, w której zakłada się egzogeniczność zmiennej  $x_t$ .

### Literatura

1. Greene W., *Econometric analysis*, Fifth Edition, Prentice Hall 2003.
2. Hozer J., Doszyń M., *Ekonometria skłonności*, PWE, Warszawa 2004.
3. Maddala G.S., *Ekonometria*, Wydawnictwo Naukowe PWN, Warszawa 2006.

## TESTING EXOGENEITY IN ECONOMETRIC MODELS

### Summary

In the article necessity of testing exogeneity of variables in econometric models was emphasized. Most popular in econometric literature conceptions of exogeneity such as pre-determination, strict exogeneity, weak exogeneity, superexogeneity, strong exogeneity, Granger causality were presented. Hausman test useful while testing exogeneity (but also when verifying hypothesis that states errors in variables) was described. Types of relationships between variables were presented. Hausman test is based on comparison of estimator of ordinary least squares method (OLS) and instrumental variable estimator (IV).

In empirical example exogeneity of disposable incomes in econometric model with individual consumption as a dependent variable was tested. Time variable was also involved as a independent variable.

*Translated by Mariusz Doszyń*