

Sabina Denkowska

Akademia Ekonomiczna w Krakowie

MODYFIKACJA METODY ANALIZY GRAFÓW

1. Wstęp

Dobór zmiennych objaśniających do modeli ekonometrycznych jest jednym z najistotniejszych zagadnień współczesnej ekonometrii. Od prawidłowego wyboru zmiennych objaśniających zależy zgodność modelu z opisywaną przez niego rzeczywistością gospodarczą, a tym samym – jego użyteczność.

Metoda analizy grafów jest powszechnie znaną metodą doboru zmiennych objaśniających do liniowego modelu regresji opartą na badaniu istotności współczynników korelacji w macierzy korelacji. W metodzie tej ignorowany jest fakt testowania wielokrotnego w macierzy korelacji i wnioski wysnuwane są po zweryfikowaniu wielu hipotez statystycznych, każdej na poziomie istotności α . Lekceważenie efektu testowania wielokrotnego powoduje, iż wykrywamy zbyt dużo istotnych współczynników korelacji¹, co skutkuje zbytnią redukcją zmiennych. Tym samym w ostatecznym zbiorze zmiennych może zabraknąć ważnych zmiennych objaśniających². Celem artykułu jest zaproponowanie modyfikacji metody analizy grafów, która zapewni kontrolę efektu testowania wielokrotnego w macierzy korelacji.

¹ Zob.: Badania Monte Carlo u A. Sokołowskiego, S. Denkowskiej [2000].

² Zob. [Denkowska 2006] – przykład empiryczny dotyczący testowania wielokrotnego w macierzy korelacji w celu wyboru zmiennych diagnostycznych do oceny przestrzennego zróżnicowania sytuacji mieszkaniowej w Polsce w roku 2002. Na tym przykładzie zaprezentowano dwa podejścia: jedno polegające na tradycyjnym wnioskowaniu o istotności współczynników korelacji w macierzy korelacji, w którym ignorowany jest efekt wielokrotności testowania, oraz takie podejście, w którym efekt ten jest kontrolowany przez zastosowanie wybranych procedur testowań wielokrotnych opartych na prawdopodobieństwach testowych. Przykład pokazuje, że lekceważenie testowania wielokrotnego powoduje, iż wykrywamy zbyt dużo istotnych współczynników korelacji, co skutkuje zbytnią redukcją zmiennych diagnostycznych.

2. Wprowadzenie w teorię testowań wielokrotnych

W celu unaocznienia efektu testowania wielokrotnego rozważmy k niezależnych testów istotności, za pomocą których przeprowadzamy weryfikację k prawdziwych hipotez zerowych. Testowania przeprowadzamy na poziomie istotności α . Prawdopodobieństwo tego, że nie odrzucimy żadnej hipotezy zerowej, wynosi $(1 - \alpha)^k$. Jeżeli przykładowo testujemy $k = 20$ niezależnych, prawdziwych hipotez zerowych, każdą na poziomie istotności 0,05, to prawdopodobieństwo odrzucenia przynajmniej jednej z nich wynosi 0,64, stąd wniosek, że bardziej prawdopodobne jest uzyskanie przynajmniej jednego istotnego statystycznie wyniku niż nieodrżucenie żadnej z hipotez zerowych. W badaniach praktycznych rzadko mamy do czynienia z niezależnymi testami, więc powyższe obliczenia należy traktować jako ostrzeżenie, że im więcej testowań przeprowadzamy, tym więcej występuje błędnych odrzuceń hipotez zerowych. Należy zatem się starać ograniczać w badaniach liczbę przeprowadzanych testów oraz stosować właściwą metodologię umożliwiającą kontrolowanie zjawiska wnioskowania wielokrotnego.

2.1. Rodzina wnioskowań

Właściwa kontrola efektu wielokrotności testowania jest zagadnieniem trudnym i kontrowersyjnym. Pierwszą, a zarazem kluczową decyzją jest wybór zbioru wnioskowań, który będzie tworzyć rodzinę wnioskowań. Y. Hochberg i A.C. Tachane [1987] zalecają, żeby w przypadku, gdy testowane hipotezy nie są ze sobą powiązane ani zawartością, ani późniejszym wykorzystaniem, traktować je oddzielnie, a nie łącznie. W przeciwnym wypadku istotne jest branie pod uwagę łącznego pomiaru błędów. Gdy wniosek końcowy wysnuwany jest na podstawie przeprowadzonych testów analizowanych łącznie i jego trafność zależy od łącznego pomiaru błędów dla danego zbioru wnioskowań, wtedy taki zbiór wnioskowań powinien być rozpatrywany łącznie jako **rodzina**.

W metodzie analizy grafów mamy do czynienia z rodziną wnioskowań dotyczących istotności współczynników korelacji, a wniosek końcowy dotyczący współzależności zmiennych powinien uwzględniać łączny pomiar błędów I rodzaju.

2.2. Miary błędu I rodzaju dla rodziny wnioskowań

W celu wprowadzenia miar błędu I rodzaju dla rodziny wnioskowań rozważmy problem jednoczesnego testowania k hipotez, wśród których k_0 hipotez jest prawdziwych. Zmienne losowe ułatwiające zdefiniowanie miar błędu I rodzaju dla rodziny wnioskowań wprowadza tab. 1.

Tabela 1. Liczba błędów w czasie testowania k hipotez zerowych

Decyzje Prawda/fałsz	Nie ma podstaw do odrzucenia H_0	Odrzucić H_0	Suma
H_0 jest prawdziwa	U	V	k_0
H_0 jest fałszywa	T	S	$k - k_0$
Suma	$k - R$	R	K

Źródło: [Benjamini, Hochberg 1995].

W ostatnim latach w literaturze na temat testowań wielokrotnych najczęściej wyróżniane są dwie miary błędu I rodzaju dla rodziny wnioskowań:

- **FWE (familywise error rate)** $\text{FWE} = P(V \geq 1)$,
- **FDR (false discovery rate):**
$$\text{FDR} = \begin{cases} E\left(\frac{V}{R}\right), & \text{gdy } R > 0 \\ 0, & \text{gdy } R = 0 \end{cases}$$

Do roku 1995 kontrola FWE wydawała się najwłaściwszym rozwiązaniem we wnioskowaniu wielokrotnym. Kontrola FWE dla rodziny wnioskowań oznacza, iż prawdopodobieństwo odrzucenia przynajmniej jednej prawdziwej hipotezy zerowej jest nie większe od ustalonego z góry α . Jednak nie we wszystkich badaniach stosowanie procedur kontrolujących FWE daje dobre rezultaty. W przypadku licznych rodzin wnioskowań procedury FWE osłabiają moc indywidualnych wnioskowań, w wyniku czego otrzymujemy zbyt mało odrzuceń hipotez zerowych.

W roku 1995 Benjamini i Hochberg [1995] zaproponowali nową miarę błędu I rodzaju FDR (false discovery rate), której kontrola oznacza, iż wartość oczekiwana frakcji błędnych odrzuceń wśród wszystkich odrzuceń hipotez zerowych jest kontrolowana na ustalonym z góry poziomie.

W celu pokazania różnicy pomiędzy FWE i FDR rozpatrzmy sytuację, gdy rodzina wnioskowań obejmuje 1000 hipotez zerowych i odpowiadających im hipotez alternatywnych. Porównajmy sytuację odrzucenia 100 hipotez zerowych, z czego jedna jest prawdziwa, z sytuacją odrzucenia w tym zbiorze dwóch hipotez, z czego jedna jest prawdziwa. Z punktu widzenia FWE obie te sytuacje są tak samo niekorzystne, bo odrzucona została jedna prawdziwa hipoteza zerowa. Można jednak spojrzeć na ten wynik w ten sposób, iż tylko 1% odrzuceń było błędnych w pierwszej sytuacji, a aż 50% – w drugiej.

3. Procedury testowań wielokrotnych oparte na prawdopodobieństwach testowych

Kontrolę wspomnianych miar błędu I rodzaju dla rodziny wnioskowań w macierzy korelacji [Denkowska 2006] zapewnić mogą odpowiednie **procedury testowań wielokrotnych oparte na prawdopodobieństwach testowych**. Procedury te mogą być stosowane w przypadku skończonych rodzin hipotez składających się

tylko z hipotez minimalnych. Proces testowania z ich wykorzystaniem opiera się przede wszystkim na analizie indywidualnych prawdopodobieństw testowych, tak więc zakres zastosowań tych procedur jest bardzo szeroki. Kolejną zaletą tych procedur są nieduże wymagania co do założeń modelu statystycznego. Istotną kwestią jest jedynie typ zależności pomiędzy statystykami testowymi, bowiem w przypadku niezależnych statystyk testowych lub pewnych typów zależności można stosować procedury o większej mocy. Specyfika macierzy korelacji zawęża grupę procedur testowań wielokrotnych do procedur, które można stosować bez względu na typ zależności między statystykami testowymi.

3.1. Skorygowane prawdopodobieństwa testowe

Założmy, że rozpatrujemy rodzinę 'k' minimalnych hipotez zerowych $H_{0,1}, H_{0,2}, \dots, H_{0,k}$ z odpowiadającymi im prawdopodobieństwami testowymi $p_1, p_2, \dots, p_k$. Uporządkujmy prawdopodobieństwa testowe $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(k)}$ i niech $H_{(0,1)}, H_{(0,2)}, \dots, H_{(0,k)}$ oznaczają odpowiadające im hipotezy zerowe.

Analogicznie do definicji zwykłych (nieskorygowanych) prawdopodobieństw testowych p **skorygowane prawdopodobieństwo** $\tilde{p}_i$ dla hipotezy $H_{0,i}$ vs. $H_{A,i}$ równe jest najmniejszej wartości FWE, dla której $H_{0,i}$ może zostać odrzucona, gdy cała rodzina hipotez jest rozpatrywana. Jeżeli mamy wyznaczone³ skorygowane prawdopodobieństwa testowe $\tilde{p}_i$ ($i = 1, \dots, k$) dla każdego testu $H_{0,i}$ vs. $H_{A,i}$, decyzja o odrzuceniu hipotezy $H_{0,i}$ na poziomie FWE równym α podejmowana jest, gdy $\tilde{p}_i \leq \alpha$.

Skorygowane prawdopodobieństwa testowe są definiowane analogicznie z FDR.

3.2. Procedury testowań wielokrotnych kontrolujące FWE

Do uniwersalnych procedur testowań wielokrotnych kontrolujących FWE zaliczamy jednoetapową **procedurę Bonferroniego** oraz jej wieloetapową modyfikację – **procedurę Holma**. Procedura Bonferroniego jest najprostszą procedurą testowań wielokrotnych, a jej algorytm można przedstawić następująco: *odrzucaamy $H_{0,i}$ wtedy, gdy $p_i \leq \frac{\alpha}{k}$.*

Metoda Bonferroniego jest bardzo konserwatywna, czyli jest metodą o małej mocy. Skorygowane prawdopodobieństwa testowe dla metody Bonferroniego wyznaczane są ze wzoru:

$$\tilde{p}_j = \min(k p_j; 1) \text{ dla } j = 1, \dots, k. \quad (1)$$

³ Sposoby wyznaczania skorygowanych prawdopodobieństw testowych dla poszczególnych procedur testowań wielokrotnych zostaną omówione w dalszej części artykułu.

Procedura Holma jest mniej konserwatywna niż metoda Bonferroniego, bowiem każda hipoteza odrzucona przez metodę Bonferroniego jest odrzucona również przez metodę Holma, hipotezy odrzucone przez metodę Holma zaś mogą nie zostać odrzucone przez metodę Bonferroniego. Skorygowane prawdopodobieństwa testowe dla metody Holma wyznaczone są ze wzorów [Westfall i in. 1999, s. 30]:

$$\tilde{p}_{(1)} = \min(1; k p_{(1)}) \text{ oraz } \tilde{p}_{(j)} = \min\left(1; \max\left(\tilde{p}_{(j-1)}, (k-j+1)p_{(j)}\right)\right) \text{ dla } j = 2, \dots, k. \quad (2)$$

3.3. Procedura testowań wielokrotnych kontrolująca FDR

Uniwersalną procedurą kontrolującą FDR (*false discovery rate*) jest **procedura Yekuteliego-Benjaminiego** [Benjamini, Yekutieli 2001]. Jest ona modyfikacją procedury LSU (*linear step-up*) Hochberga-Benjaminiego [Benjamini, Hochberg 1995] zapewniającą kontrolę **FDR** bez względu na typ zależności pomiędzy statystykami testowymi. Niestety procedura ta jest procedurą konserwatywną w porównaniu z procedurą LSU. Skorygowane prawdopodobieństwa testowe dla metody Yekuteliego-Benjaminiego wyznaczamy ze wzorów:

$$\tilde{p}_{(k)} = \min\left(1; p_{(k)} \sum_{i=1}^k \frac{1}{i}\right) \text{ oraz } \tilde{p}_{(k-j)} = \min\left(\tilde{p}_{(k-j+1)}, p_{(k-j)} \frac{k}{k-j} \sum_{i=1}^k \frac{1}{i}\right) \text{ dla } j = 1, \dots, k-1. \quad (3)$$

4. Modyfikacja metody analizy grafów

Dla jasności późniejszych rozważań przypomnijmy oryginalną metodę analizy grafów [Bartosiewicz 1980, s. 21], wyróżniając 3 etapy:

1. Na podstawie danych empirycznych dotyczących n obserwacji dokonanych na $(l+1)$ zmiennych buduje się macierz obserwacji **X** o wymiarze $n \times (l+1)$ (pierwsza kolumna macierzy **X** zawiera informacje o danych empirycznych zmiennej objaśnianej). Z macierzy **X** wyznacza się macierz korelacji **W**, a następnie usuwa się z niej pierwszy wiersz i pierwszą kolumnę, zawierające współczynniki korelacji z próby pomiędzy zmienną objaśnianą a zmiennymi objaśniającymi. Otrzymaną w ten sposób macierz oznacza się przez **R** i jest to macierz korelacji o wymiarze $l \times l$ zawierająca współczynniki korelacji pomiędzy l zmiennymi objaśniającymi.

2. Etap kolejny polega na weryfikacji istotności współczynników korelacji pomiędzy zmiennymi objaśniającymi. Sprawdzamy $\frac{l(l-1)}{2}$ hipotez o postaci:

$H_0: \rho_{ij} = 0 \quad (i \neq j) \text{ vs. } H_A: \rho_{ij} \neq 0$. Sprawdzianem tej hipotezy jest statystyka

$t = \frac{r_{ij}}{\sqrt{1-r_{ij}^2}} \sqrt{n-2}$, której wartość empiryczna otrzymana dla próby jest porówny-

wana z wartością krytyczną $t_{\alpha;n-2}$ odczytaną z tablic t -Studenta dla kwantyla $1 - \frac{\alpha}{2}$ i $(n-2)$ stopni swobody. W praktyce zalecane jest, by wyznaczyć wartość kry-

tyczną r^* ze wzoru: $r^* = \sqrt{\frac{t_{\alpha;n-2}^2}{n-2+t_{\alpha;n-2}^2}}$ i następnie wszystkie współczynniki kore-

lacji r_{ij} ($i, j=1, \dots, l$) w macierzy $\mathbf{R}$, które spełniają nierówność $|r_{ij}| \leq r^*$ dla $i \neq j$, zastąpić zerami, otrzymując w ten sposób nową macierz $\mathbf{R}'$.

3. Wykorzystując macierz $\mathbf{R}'$, buduje się grafy, w których wierzchołkami są zmienne, a łukami – współczynniki korelacji. Następnie określa się stopień każdego wężła grafu, tzn. liczbę łuków, którymi jest on związany z innymi wierzchołkami. Dla wierzchołków izolowanych stopień wynosi zero. W każdym grafie spójnym wyróżniamy wierzchołek o maksymalnym stopniu. Jeżeli w danym grafie spójnym jest kilka takich wierzchołków, to wybiera się spośród nich ten, który reprezentuje zmienną najsilniej skorelowaną ze zmienną objaśnianą. Ostatecznie zmiennymi objaśniającymi pozostaną zmienne wyróżnione z grafów spójnych oraz zmienne reprezentowane przez wierzchołki izolowane.

Ponieważ uzyskanie prawdopodobieństw testowych dla poszczególnych testowań nie stanowi problemu, często w praktyce etap drugi sprowadzany jest do porównań otrzymanych prawdopodobieństw testowych z przyjętym na wstępie poziomem istotności α .

Prawdopodobieństwa testowe wykorzystamy do zmodyfikowania metody analizy grafów.

Modyfikacja procedury analizy grafów polega na zastąpieniu etapu drugiego w algorytmie procedury analizy grafów etapem 2':

2'. Analogicznie jak w oryginalnej wersji przeprowadzamy weryfikację istotności współczynników korelacji pomiędzy zmiennymi objaśniającymi. W rezultacie otrzymujemy zbiór prawdopodobieństw testowych: $p_1, \dots, p_k$, gdzie $k=1, 3, \dots, \frac{l(l-1)}{2}$. Następnie porządkujemy prawdopodobieństwa testowe tak, aby: $p_{(1)} \leq \dots \leq p_{(k)}$. Dla $p_{(1)}, \dots, p_{(k)}$ wyznaczamy odpowiadające im skorygowane prawdopodobieństwa testowe $\tilde{p}_{(1)}, \dots, \tilde{p}_{(k)}$, wykorzystując **jedną z procedur**:

- **procedurę Holma kontrolującą FWE** – wzory (2),
- **procedurę Yekutielięgo-Benjaminiego kontrolującą FDR** – wzory (3).

Następnie wszystkie współczynniki korelacji w macierzy $\mathbf{R}$, dla których spełniona jest nierówność $\tilde{p}_{(i)} > \alpha$ ($i=1, \dots, \frac{l(l-1)}{2}$), zastępujemy zerami i w ten sposób otrzymujemy nową macierz $\mathbf{R}'$.

Wybór procedury testowań wielokrotnych zastosowanej do zmodyfikowania metody analizy grafów zależy od poczynionego na wstępie wyboru miary kontroli błędu I rodzaju dla rodziny wnioskowań dotyczących istotności współczynników

korelacji w macierzy korelacji. Kontrolę prawdopodobieństwa błędnego stwierdzenia jednej lub więcej istotnych zależności na poziomie α zapewniają procedury Bonferroniego i Holma. Do modyfikacji metody analizy grafów została jednak zaproponowana metoda Holma, gdyż ma większą moc niż metoda Bonferroniego. Ponieważ w przypadku rozpatrywania dużej macierzy korelacji moc procedury Holma znacznie spada, warto rozważyć, czy nie wystarczy kontrola frakcji błędnych odrzuceń wśród wszystkich odrzuceń hipotez zerowych i zastosowanie procedury Yekutieli-Benjaminiego, która dopuszcza pewien niewielki procent (α 100%) błędnie stwierdzonych istotnych zależności wśród wszystkich zależności uznanych za istotne, ale za to unikamy zachowawczej postawy procedur kontrolujących FWE.

5. Podsumowanie

Niezaprzeczalnie kontrola efektu testowania wielokrotnego w macierzy korelacji jest wskazana. Ignorowanie efektu testowania wielokrotnego powoduje, iż wykrywamy zbyt dużo istotnych współczynników korelacji, co w przypadku metody analizy grafów skutkować może tym, iż w finalnym zbiorze zmiennych egzogenicznych zabraknie ważnych zmiennych objaśniających [Denkowska 2006].

W artykule zaproponowano modyfikacje metody analizy grafów uwzględniające efekt wielokrotności testowania w macierzy korelacji. W zależności od istoty zagadnienia ekonomicznego można wybierać spośród procedur kontrolujących FWE lub FDR. Gdy istotne jest, by prawdopodobieństwo błędnego stwierdzenia jednej lub więcej istotnych statystycznie zależności było kontrolowane na poziomie α , najlepszym rozwiązaniem jest zastosowanie modyfikacji opartej na metodzie Holma. Kontrola FWE na poziomie α oznacza, że uzyskane istotne wyniki możemy z dużym zaufaniem przyjąć, gdyż prawdopodobieństwo błędnego alarmu jest nie większe od α . Natomiast w sytuacji licznej rodziny wnioskowań warto rozważyć, czy nie wystarczy kontrola (FDR) wartości oczekiwanej frakcji błędnych odrzuceń wśród wszystkich odrzuceń na z góry ustalonym poziomie. W przypadku bowiem „dużych” macierzy korelacji, a tym samym – licznej rodziny wnioskowań, gwałtownie maleje moc indywidualnych testów w sytuacji zastosowania procedur kontrolujących FWE i w efekcie zastosowania tych procedur otrzymujemy zbyt mało istotnych wyników. Zastosowanie procedur kontrolujących FDR pozwala uniknąć zachowawczej postawy FWE. Wybór przez badacza tej miary błędu I rodzaju dla rodziny wnioskowań oznacza, iż dopuszcza on i akceptuje pewien niewielki procent α 100% błędnych odrzuceń wśród wszystkich odrzuceń, a w zamian za to oczekuje mniej ostrożnej postawy niż w przypadku kontroli FWE i w efekcie większej liczby wykrytych istotnych statystycznie zależności. Z pewnością jest to lepszym rozwiązaniem, niż ignorowanie efektu wielokrotności testowania.

Warto na koniec podkreślić, że ignorowanie efektu wielokrotności prowadzi do wykrywania wielu zupełnie przypadkowych zależności. Zależności takie są później często eksponowane w naukowych publikacjach jako ciekawe czy wręcz zdumiewające wyniki badań. To z kolei budzi sceptycyzm wobec metod statystycznych, podczas gdy źródłem nieporozumienia są niewłaściwie przeprowadzone badania, nieuwzględniające efektu testowania wielokrotnego.

Literatura

- Bartosiewicz S. (red.) (1980), *Metody ekonometryczne. Przykłady i zadania*, PWE, Warszawa.
- Benjamini Y., Hochberg Y. (1995), *Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing*, „Journal of the Royal Statistical Society, Ser. B”, vol. 57, nr 1.
- Benjamini Y., Yekutieli D. (2001), *The Control of the False Discovery Rate in Multiple Testing under Dependency*, „Annals of Statistics” 29.
- Denkowska S. (2005), *Zastosowanie procedur testowań wielokrotnych opartych na uporządkowanych prawdopodobieństwach testowych do wydzielania jednorodnych podgrup wartości przeciętnych*, „Przegląd Statystyczny” vol. 52, nr 1.
- Denkowska S. (2006), *Multiple Testing in a Correlation Matrix*, [w:] J. Pociecha (red.), *A Comparative Analysis of the Socio-Economic Consequences of Transition Process in Central and Eastern European Countries*, Cracow University of Economics.
- Hochberg Y., Tamhane A.C. (1987), *Multiple Comparison Procedures*, John Wiley & Sons, NY.
- Sokołowski A., Denkowska S. (2000), *Testowanie wielokrotne w macierzy korelacji – eksperyment symulacyjny*, [w:] A. Zeliaś (red.), materiały z XXXV Konferencji Statystyków, Ekonometryków i Matematyków Akademii Ekonomicznych Polski Południowej (Osieczany, 23-25.03.1999 r.), Kraków.
- Westfall P.H., Tobias R.D., Rom D., Wolfinger R.D., Hochberg Y. (1999), *Multiple Comparisons and Multiple Tests, Using the SAS System*, SAS Institute Inc.

MODIFICATION OF GRAPH ANALYSIS METHOD

Summary

Graph analysis method is the commonly known procedure of independent variables selection in linear regression model. It is based on traditional attitude to the significance testing of correlation coefficients in the correlation matrix, in which multiple testing effect is ignored because the decisions of significance are taken each on α -level. Proposed modification of graph analysis method ensures the con-

trol of multiple testing effect in a correlation matrix. This modification is based on multiple testing procedures which ensure control of probability of at least one wrong rejection when the entire family of inferences is considered (**FWE** *Family-wise Error Rate*), or the control of **FDR** (*False Discovery Rate*) which is the expected proportion of erroneously rejected null hypotheses among the rejected ones.