Rozdział 10

Metodologia testowania hipotez

W tym rozdziale omówimy problem metodologicznie poprawnego testowania hipotez i wyboru prawidłowej liczby zmiennych do modelu. Prawidłowy metodologicznie sposób ma kluczowe znaczenie przy ustalaniu zbioru zmiennych, które wpływają na zmienną zależną oraz przy narzucaniu dodatkowych ograniczeń na model.

10.1 Obciążenie Lovella

Rozpatrzymy przypadek hipotezy złożonej o nieistotności wszystkich K zmiennych w modelu. Hipotezę tę można zapisać jako $H_0 : \beta_1 = \ldots = \beta_K = 0$ i przetestować jako hipotezę łączną przy poziomie istotności α. Czy hipotezę tę można także przetestować testując K hipotez prostych $H_0^1 : \beta_1 = 0; \ldots ; H_0^K : \beta_K = 0$, przy poziomie istotności α? W drugim przypadku odrzucimy hipotezę o łącznej nieistotności, gdy odrzucona zostaje jedna z hipotez prostych $H_0^i : \beta_i = 0$. Procedury te nie są równoważne mimo logicznej równoważności testowanych hipotez. Problem tkwi w trudnościach z określeniem rzeczywistego poziomu istotności dla drugiej procedury.

Rozważmy przypadek, kiedy statystyki testowe dla każdej z hipotez prostych są od siebie niezależne. Z definicji poziomu istotności wiemy, że jest on równy prawdopodobieństwu błędu I rodzaju a więc prawdopodobieństwu odrzucenia prawdziwej H_0. Zauważmy, że popełnimy błąd I rodzaju jeśli odrzucimy H_0 dla jakiejkolwiek z hipotez prostych, ponieważ w takim przypadku odrzucamy także hipotezę łączną. Prawdopodobieństwo, że nie popełnimy błędu I rodzaju jest więc równe prawdopodobieństwu, że żadna z prawdziwych hipotez nie zostanie odrzucona. Dla przypadku niezależnych statystyk testowych prawdopodobieństwo to jest równe $(1-\alpha)^K$. W konsekwencji prawdopodobieństwo popełnienia błędu pierwszego rodzaju jest równe:

$$\alpha^* = 1 - (1-\alpha)^K$$

Różnicę między założonym poziomem istotności α i prawdopodobieństwem α^* nazywamy obciążeniem Lovella. Zauważmy teraz, że $\lim\limits_{K \longrightarrow \infty} \alpha^* = 1$. Wynika z tego, że dla dużej liczby testowanych hipotez prostych prawdopodobieństwo popełnienia błędu drugiego rodzaju zbliża się do 1.

Problem ten związany jest z tak zwanym przekopywaniem danych (*data mining*). Przekopywanie danych służy znajdowaniu tych zmiennych, które są istotne przy wyjaśnianiu zmienności zmiennej zależnej. Jeśli jednak wyjdziemy od wystarczająco dużego zbioru zmiennych wyjściowych i badać będziemy kolejno istotność zmiennych, to prawie zawsze znajdziemy pewne zmienne o istotnych statystykach t, nawet wtedy, gdy wszystkie zmienne w danym zbiorze są w rzeczywistości nieistotne.

Przykład 10.1.1 *Grupa socjologów postanowiła przetestować hipotezę, że fakt urodzenia się pod konkretnym znakiem Zodiaku ma wpływ na losy respondentów. Do tego celu przeanalizowano bazę danych zawierającą datę urodzenia respondenta i jego prestiż jego zawodu w skali Treimana. Data urodzenia posłużyła do stworzenia* 12 *zmiennych zerojedynkowych determinujących znak Zodiaku, pod którym urodził się respondent. Zmienną zależną był prestiż wykonywanego zawodu, które stanowić miały miarę sukcesu życiowego. Uzyskano następujące wyniki estymacji:*

prestiż	*Współczynnik*	*(Błąd Std.)*	*t*	$\Pr(\lvert t\rvert > t^*)$
zodiak2	-1.186	(1.189)	-1.00	0.319
zodiak3	-1.163	(1.215)	-0.96	0.338
zodiak4	-1.944	(1.225)	-1.59	0.113
zodiak5	-0.844	(1.232)	-0.69	0.493
zodiak6	0.490	(1.249)	0.39	0.695
zodiak7	-1.736	(1.233)	-1.41	0.159
zodiak8	-2.517	(1.255)	-2.01	0.045
zodiak9	-1.772	(1.303)	-1.36	0.174
zodiak10	-1.710	(1.208)	-1.42	0.157
zodiak11	-0.063	(1.216)	-0.05	0.959
zodiak12	-0.446	(1.186)	-0.38	0.707
stała	39.486	(0.866)	45.60	0.000

Na poziomie istotności $\alpha = 0,05$ *stwierdzono, że znaki Zodiaku mają istotny w wpływ na kariery zawodowe respondentów, ponieważ istotna okazała się zmienna zerojedynkowa związana z* 8 *znakiem zodiaku.*

Wynik ten jest najprawdopodobniej rezultatem błędnej procedury testowania. Prawdziwy rzeczywisty poziom istotności (przy założeniu niezależności wielkości statystyk testowych) wynosi $\alpha^* = 1-(0.9)^{11} \approx 0,44$. *Oznacza to, że dla prawdziwej hipotezy zerowej o nieistotności wszystkich współczynników w* 44% *przypadków uzyskujemy co najmniej jedną statystycznie istotną statystykę* t. *Prawidłowa procedura testowania istotności zbioru zmiennych polega na łącznym przetestowaniu, łącznej hipotezy o nieistotności wszystkich zmiennych zerojedynkowych. Wartość statystyki testowej dla tej hipotezy wynosi* $F(11, 2106) = 1.08$, *policzony poziom istotności* 0.38. *Przy prawidłowym sposobie testowania hipotezy zerowej, że nie ma podstaw do odrzucenia* H_0, *że znaki zodiaku nie wpływają istotnie na karierę* *zawodową.*

Testowanie hipotez prostych zamiast testowania hipotezy łącznej nie zawsze musi prowadzić do wyższego prawdopodobieństwa odrzucenia hipotezy zerowej. Dobrym przykładem jest tu model, w którym występuje współliniowość między x_1 i x_2. W takim przypadku statystyki t dla x_1 i x_2 mogą być niskie ponieważ wariancje estymatorów b_1 i b_2 są wysokie. Jednak spadek sumy kwadratów reszt w przypadku usunięcia zmiennych x_1 i x_2 może być duży pod warunkiem, że obie te zmienne są silnie skorelowane z y. W takim przypadku może się zdarzyć, że nie będzie podstaw do odrzucenia odrzucenia $H_0^1 : \beta_1 = 0$ i $H_0^2 : \beta_2 = 0$ na poziomie istotności α ale odrzucimy, na tym samym poziomie istotności, hipotezę $H_0 : \beta_1 = \beta_2 = 0$.

Z powyższych rozważań wynika, że przy testowaniu hipotez łącznych należy unikać wielokrotnego testowania hipotez prostych zamiast testowania hipotez złożonych. Przekonamy się jednak, że wielokrotnego testowania hipotez jest w praktyce trudno uniknąć. Można jednak robić to tak, by przy każdym teście uzyskiwać prawidłowy poziom istotności.

Pytania:

1. Wyjaśnić co rozumiemy przez obciąznie Lovella.

10.2 Upraszczanie modelu

Standardowo, gdy zaczynamy badanie ekonometryczne, dysponujemy pewnym zbiorem zmiennych, które potencjalnie mogą się okazać istotne dla wyjaśnienia zmiennej zależnej. Pojawia się pytanie w jaki sposób powinniśmy ustalić listę istotnych zmiennych. Poza hipotezą $H_0 : \beta_1 = \ldots = \beta_K = 0$ interesują nas więc też hipotezy dotyczące poszczególnych współczynników $H_0^1 : \beta_1 = 0; \ldots ; H_0^K : \beta_K = 0$. Inaczej mówiąc, jest dla nas ważne nie tylko to, czy którakolwiek zmienne w modelu jest istotna ale także to, które z nich są istotne.

Przykład 10.2.1 *Rozważmy następujący model dynamiczny dla funkcji konsumpcji:*

$$kons_t = \mu + \beta_0 pkb_t + \beta_1 pkb_{t-1} + \ldots + \alpha_p pkb_{t-p} + \varepsilon_t$$

W modelu tym obecna konsumpcja zależy nie tylko od dochodu w obecnym okresie ale także od dochodu z poprzednich okresów. Problemem przy specyfikacji tego modelu jest nie tylko to, że nieznane są wielkości parametrów $\alpha_{1,} \ldots, \alpha_p$ ale nieznana jest także liczba opóźnień p, które powinniśmy umieści w modelu.

Szukanie prawidłowej specyfkcaji modelu można rozumieć jako szukanie takiej specyfikacji, która przy możliwe małej liczbie parametrów dobrze opisuje analizowany zbiór danych. Obecnie uważa się, że prawidłową metodą szukania prawidłowej specyfikacji modelu jest metoda od ogólnego do szczegółowego (*general to specific*). Wybór prawdiłowego modelu w ramach tej metody można dokonać za pomocą klasycznych statystyk testowych ale także za pomocą tak zwanych kryteriów informacyjnych.

10.3 Metoda od ogólnego do szczegółowego

Nieusystematyzowane przeszukiwanie przeszukiwanie zbioru danych w celu znalezienia istotnych zmiennych niezależnych może nas doprowadzić do uznania przypadkowych zmiennych za zmienne istotne. Szczególnie łatwo może się tak stać w przypadku, gdy liczba obserwacji jest stosunkowo niewielka. Kluczowe dla osiągnięcia sukcesu przy przekopywaniu danych jest właściwe ustrukturyzowanie tych poszukiwań.

Proponowanym obecnie rozwiązaniem problemu szukania prawidłowej specyfikacji jest metodologia od ogólnego do szczegółowego. Polega ona na stopniowym upraszczaniu możliwie najogólniejszego modelu początkowego poprzez narzucanie na niego coraz bardziej rozbudowanych zagnieżdzonych hipotez. Mówimy, że hipotezy są zagnieżdzone jeśli można uszeregować tak, że H_0^1 zawiera najmniej ograniczeń, H_0^2 zawiera ograniczenia zawarte w H_0^1 plus pewne dodatkowe ograniczenia i tak dalej aż do H_0^K zawierającej najwięcej ograniczeń. Sytuację taką zapisujemy jako $H_0^1 \subset H_0^2 \subset \ldots \subset H_0^K$. Upraszczając model sekwencyjnie testyjemy hipotezy od H_0^1 aż do momentu, kiedy hipoteza H_0^i zostanie odrzucona.

Przykład 10.3.1 *Przypuśćmy, że modelujemy poziom dochodu gospodarstwa domowego na podstawie następujących charakterystyk demograficzno-społecznych: miejsce zamieszkania (miasto, wieś), płeć głowy gospodarstwa (mężczyzna, kobieta) oraz poziomu wykształcenia głowy (podstawowe, średnie, wyższe). Łacznie w takim modelu będziemy mieć* 4 *zmiennych zero-jedynkowych. Załóżmy, że interesują nas następujące hipotezy:*

H_0^1 : *dla poziomu dochodu gospodarstwa nie ma znaczenia płeć głowy gospodarstwa.*

H_0^2 : *dla poziomu dochodu gospodarstwa nie ma znaczenia płeć głowy gospodarstwa i miejsce zamieszkania.*

H_0^3 : *dla poziomu dochodu gospodarstwa nie mają znaczenia obie charakterystyki.*

W modelu tym $H_0^1 \subset H_0^2 \subset H_0^3$. *Stosując metodologię od ogólnego do szczegółowego powinniśmy najpierw przetestować hipotezę łączną, że zmienna zero-jedynkowe związana z płcią są nieistotne. W przypadku braku podstaw do odrzucenia tej hipotezy testujemy hipotezę łączną, że* 2 *zmienne zero-jedynkowe związane z miejscem zamieszkania i płcią są nieistotne. Jeśli z kolei nie ma podstaw do odrzucenia tej hipotezy, to powinniśmy przetestować hipotezę o łącznej nieistotności wszystkich zmiennych.*

Ćwiczenie 10.3.2 (c.d. 10.2.1) *Ustalamy postać funkcji konsumpcji dla Polski na podstawie danych kwartalnych z lat* 1995.1 *do* 2005.2. *Testujemy najpierw szereg hipotez zagnieżdzonych związanych z istotnością opóźnień. Dodatkowo do przetestowanych hipotez dołączamy hipotezę o niestotności nieopóźnionej wielkości dochodu w modelu oraz o niestotności stałej. Uwględnimy więc w modelu maksymalnie* 5 *opóźnień. Model jest na zmiennych zlogarytmowanych. Pierwsza specyfikacja daje nam następujące oszacowania:*

$$kons_t = \underset{(0.331)}{0.446} - \underset{(0.145}{0.177} pkb_t + \underset{(0.137)}{0.119} pkb_{t-1} + \underset{(0.040)}{0.317} pkb_{t-2} + \underset{(0.042)}{0.232} pkb_{t-3} - \underset{(0.152)}{0.142} pkb_{t-4} + \underset{(0.137)}{0.227} pkb_{t-4}$$

Test dla hipotezy, że parametr przy pkb_{t-5} *jest równy zeru daje policzony poziom istotności* $\alpha^* = 0.11$. *Test hipotezy, że parametry przy* pkb_{t-5} *i* pkb_{t-4} *są równe zeru daje* $\alpha^* = 0.24$. *Test dla hipotezy, że* *parametry przy pkb_{t-5}, pkb_{t-4} i pkb_{t-3} są równe zeru daje $\alpha^* = 0.000$.Wynika z tego, że powinniśmy w modelu uwzględnić 3 opóźnienia. Z kolei test hipotezy, że parametry przy pkb_t, pkb_{t-4}, pkb_{t-5} są równe zeru daje $\alpha^* = 0.22$. Test, że stała oraz współczynniki przy pkb_t, pkb_{t-4} są równe zeru daje $\alpha^* = 0.09$. Ostateczna forma modelu, jest więc następująca:*

$$kons_t = \underset{(0.039)}{0.374} pkb_{t-1} + \underset{(0.038)}{0.349} pkb_{t-2} + \underset{(0.041)}{0.243} pkb_{t-3}$$

Przy testowaniu kolejnych hipotez jako hipotezę alternatywną przyjmujemy model początkowy. Inaczej mówiąc w kolejnych krokach testujemy coraz bardziej rozbudowane hipotezy łączne.

Często, gdy przekopujemy duże ilości zmiennych, łącznej istotności wszystkich zmiennych jest niemożliwe z powodu zbyt małej ilości stopni swobody. Należy więc zaczynać od zbioru danych, których ewentualna przydatność w objaśnianiu danego zjawiska da się teoretycznie lub intuicyjnie uzasadnić. Sposób uszeregowania hipotez zagnieżdzonych może mieć także istotny wpływ na wynik testowania. Uszeregowanie hipotez powinno, mieć uzasadnienie teoretyczne.

Pytania:

1. Opisać procedurę od ogólnego do szczegółowego na przykładzie doboru liczby opóźnienień w modelu.

10.4 Kryteria informacyjne (selekcja modelu)

W poprzednim rozdziale pakazaliśmy w jaki sposób metoda od ogólnego do szczegółowego może być wykorzystana do znalezienia właściewej specyfikacji modelu przy użyciu klasycznych testów statystycznych. Metody tej często używa się stosując nie tyle klasyczne testy statystyczne ale tak zwane kryteria informacyjne.

W kontekście MNK zdefiniowaliśmy $\overline{R}^2$, która pozwalaja porównać jakość modeli opisujących zmienność danej zmiennej zależnej. Niestety miarę tę można zdefiniować jedynie w kontekście MNK. W przypadku ogólniejszej klasy modeli szacowanych za pomocą **M**etody **N**ajwiększej **W**iarygodności MNW definiujemy tak zwane kryteria informacyjne, które pozwalają podobnie jak $\overline{R}^2$ porównywać różne modele dla zmiennej zależnej. W przecieństwie do $\overline{R}^2$ w przypadku kryteriów informacyjnych przyjęta została konwencja, że najlepszym modelem jest model, dla którego wartość kryterium inofrmacyjnego jest najniższa.

Najpopularniejszymi kryteriami informacyjnymi jest kryterium informacyjne Akaike AIC (***A**kaike **I**nformation **C**riterion*) oraz Bayesowskie kryterium informacyjne Schwartza BIC (**B**ayes **I**nformation **C**riterion)[1]. Wzory na te kryteria można sformułować ogólnie w kategoriach logarytmu funkcji wiarogodności lub też w przypadku MNK w kategoriach sumy kwadratów reszt $\boldsymbol{e}'\boldsymbol{e}$ w przypadku MNK.

[1] W przypadku Bayesowskieo kryterium informacyjnego Szwartza spotyka się też skróty SC, SBC, SIC

	AIC	BIC
pełen model	-192.3156	-181.0392
$H_0 : \beta_5 = 0$	-191.2822	-181.6167
$H_0 : \beta_5 = \beta_4 = 0$	-185.5504	-177.4958
$H_0 : \beta_5 = \beta_4 = \beta_3 = 0$	-162.1931	-155.7494
$H_0 : \beta_5 = \beta_4 = \beta_1 = 0$	-161.2049	-156.3722
$H_0 : \beta_5 = \beta_4 = \beta_1 = \mu = 0$	-161.9610	-158.7392

Odpowiednie wzory dla kryteriów informacyjnych są następujące:

$$BIC = -\frac{2\ell\left(\widehat{\theta}\right)}{N} + \frac{K\log(N)}{N} = \log\left(\frac{\boldsymbol{e}'\boldsymbol{e}}{2}\right) + \frac{K\log(N)}{N}$$
$$AIC = -\frac{2\ell\left(\widehat{\theta}\right)}{N} + \frac{2K}{N} = \log\left(\frac{\boldsymbol{e}'\boldsymbol{e}}{2}\right) + \frac{2K}{N}$$

gdzie $\ell\left(\widehat{\theta}\right)$ jest logarytmem funkcji wiarygodności dla oszacowanego wektora parametrów a K jest liczbą parametów modelu a N liczbą obserwacji.

Spróbujmy przenalizować wzory na kryteria informacyjne pod kątem ogólnych zasad, które powinny obowiązywać przy wyborze modeli. Jak już wcześniej pisaliśmy dobry model powinien spełniać dwa podstawowe warunki: powinien być dobrze dopasowany i możliwie jak najprostszy. Prostotę modelu można mierzyć za pomocą liczby parametrów, które się w nim pojawiają. Jeśli przeanalizujemy teraz kryteria informacyjne w wersji dla MNK, to przekonamy się, że rosną one wraz z pogarszaniem się jakości dopasowania miarzoną przez $\log\left(\frac{\boldsymbol{e}'\boldsymbol{e}}{2}\right)$ oraz wraz ze wzrostem liczby parametrów. Przy liczeniu kryteriów informacyjnych należy zwrócić uwagę na to, czy próba na której szacowany jest model jest za każdym razem taka sama.

Różnica między kryterium AIC i BIC polega na innym ważeniu jakości dopasowania i prostoty modelu. Drugi element sumy we wzorach na kryteria informacyjne mierzy prostotę modelu. W obu przypadkach element ten rośnie wraz ze wzrostem liczby parametrów i wzrost ten jest tym większy im mniejsza jest liczba obserwacji. Takie zdefiniowanie kryterium informacyjnych związane jest z faktem, że prostota modelu jest szczególnie ważna w przypadku modeli szacowanych na małych próbach. Jakkolwiek asymptotycznie oba kryteria wybierać będą jako prawidłowy model, model prawdziwy, to jadnak w małych próbach ich wskazania mogą się znacząco różnić. W literaturze sugeruje się, że kryterium AIC ma tendencję do wybierania modelu o zbyt dużej liczbie parametrów.

Przykład 10.4.1 (c.d. 10.3.2) *W modelu dla funkcji konsumpcji model najlepszy można znaleźć na postawie wielkości kryteriów informacyjnych dla różnych specyfikacji:*

Jak widać na podstawie kryterium informacyjnego Akaike wybralibyśmy jako najlepszy pełen model a na podstawie kryterium Bayesowskiego, model ze stała i 4 *opóźnieniami. Tak jak już wspomniano* AIC

wybiera z reguły większe modele. Obie wybrane specyfikacje różnią się od tych uzyskanych przy użyciu formalnych testów.

Pytania:

1. Wymić znane ci kryteria informacyjne i opisać w jaki sposób używa się ich do wyboru najlepszego modelu.