Egzamin z ekonometrii – wersja ogólna
06-02-2019

Rozwiązania zadań

Zadanie 1

1. Należy wyznaczyć macierz wariancji – kowariancji dla wektora oszacowanych parametrów w przypadku sferyczności zaburzenia losowego: $\sigma^2(X'X)^{-1}$

$$Var\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \sigma^2(X'X)^{-1}$$

W modelu ze stałą i jedną zmienną objaśniającą:

$$X'X = \begin{bmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_N \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix} = \begin{bmatrix} N & \sum_{i=1}^{N} x_i \\ \sum_{i=1}^{N} x_i & \sum_{i=1}^{N} x_i^2 \end{bmatrix}$$

Wzór na odwrotność macierzy 2x2:

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \text{ i } \det(A) \neq 0 \text{ to } A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Zakładając, że macierz X ma pełen rząd kolumnowy (wystarczy, że istnieją takie $i \neq j$, że $x_i \neq x_j$):

$$(X'X)^{-1} = \frac{1}{N\sum_{i=1}^{N} x_i^2 - (\sum_{i=1}^{N} x_i)^2} \begin{bmatrix} \sum_{i=1}^{N} x_i^2 & -\sum_{i=1}^{N} x_i \\ -\sum_{i=1}^{N} x_i & N \end{bmatrix}$$

Czyli

$$Var(b) = \sigma^2(\mathrm{X'X})^{-1} = \frac{\sigma^2}{N\sum_{i=1}^{N} x_i^2 - (\sum_{i=1}^{N} x_i)^2}\begin{bmatrix} \sum_{i=1}^{N} x_i^2 & -\sum_{i=1}^{N} x_i \\ -\sum_{i=1}^{N} x_i & N \end{bmatrix}$$

oraz

$$cov(b_1, b_2) = -\frac{\sigma^2 \sum_{i=1}^{N} x_i}{N\sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i\right)^2}$$

2. Niech $y^* = cy$ i $X^* = aX$. Wówczas:

$$\hat{\gamma} = (X^{*\prime}X^*)^{-1}X^{*\prime}y^* =$$
$$\left((aX)' aX\right)^{-1}(aX)' cy = ca(a^2X'X)^{-1}X'y = ca\frac{1}{a^2}(X'X)^{-1}X'y = \frac{c}{a}b$$

Jeśli więc $\hat{\gamma} = b$ to spełniony musi być warunek $c = a$.

Zadanie 2

1. Występowanie heteroskedastyczności testujemy za pomocą testu White'a:
   i. hipoteza zerowa: homoskedastyczność składnika losowego.
   ii. wartość statystyki testowej wynosi chi2(14) = 10.99 oraz p-value = 0.6865 >0.05, czyli nie mamy podstaw do odrzucenia hipotezy zerowej o homoskedastyczności składnika losowego.

2. Autokorelację testujemy za pomocą:
   a) testu Durbina-Watsona:
      i. DW = 0.97 < 2, więc testujemy zestaw hipotez:

      hipoteza zerowa: brak autokorelacji rzędu pierwszego;
      hipoteza alternatywna: dodatnia autokorelacja rzędu pierwszego.

      ii. wartość statystyki testowej wynosi DW = 0,97 $< d_l = 1.41$, czyli odrzucamy hipotezę zerową o braku autokorelacji pierwszego rzędu;

   b) testu Breuscha-Godfreya:
      i. hipoteza zerowa: brak autokorelacji;
      hipoteza alternatywna: dodatnia autokorelacja.
      ii. wartość statystyki testowej wynosi chi2(4) = 24.22 oraz p-value = 0.0001 <0.01, czyli odrzucamy hipotezę zerową o braku autokorelacji.

3. Poprawność formy funkcyjnej testujemy za pomocą testu RESET:
   i. hipoteza zerowa: przyjęta postać funkcyjna modelu jest prawidłowa;
   hipoteza alternatywna: przyjęta postać funkcyjna modelu jest niepoprawna.
   ii. wartość statystyki testowej F(3, 51) = 3.45 i p-value = 0.0231 < 0,05, więc odrzucamy hipotezę zerową zakładającą poprawność przyjętej formy funkcyjnej.

4.
   a) Nie jest spełnione założenie o braku autokorelacji zaburzenia losowego oraz założenie o liniowej zależności między zmienną zależną i zmiennymi niezależnymi.
   b) Konsekwencje dla interpretacji modelu i wnioskowania statystycznego są następujące:
      i. w przypadku nie spełnienia założenia o braku autokorelacji zaburzenia losowego, estymator b jest nieobciążony i zgodny, ale nieefektywny. Estymator macierzy wariancji-

kowariancji *b* jest już obciążony i niezgodny. Macierz wariancji-kowariancji *b* jest wykorzystywana do testowania hipotez na temat istotności zmiennych, więc poprawność wnioskowania statystycznego jest podważona.

ii. Niespełnienie założenia liniowej zależności między zmienną zależną a zmiennymi niezależnymi podważa interpretację ekonomiczną modelu (interpretacja oszacowanych parametrów). Takie własności jak nieobciążoność czy efektywność estymatora MNK wyprowadzane są przy założeniu liniowej zależności między zmienną zależną a zmiennymi niezależnymi.

c)

i. Problem autokorelacji można rozwiązać za pomocą Stosowalnej UMNK lub odpornego estymatora Newey'a-Westa macierzy wariancji-kowariancji.

ii. Niepoprawna forma funkcyjna: możemy próbować poprawić formę funkcyjną modelu wprowadzając do modelu interakcje między zmiennymi, dokonać przekształceń zmiennych, zastosować model wielomianowy, schodkowy lub krzywej łamanej

Zadanie 3

1. Wprowadzenie do modelu interakcji umożliwia uchwycenie różnych „premii” uzyskiwanych przez kobiety o innej rasie za wykształcenie.
2. W modelu umieszczono zmienną *wiek* podniesioną do kwadratu, ponieważ spodziewano się, że zależność między płacą a wiekiem nie jest liniowa.

   Hipoteza o tym, że wiek nie wpływa na poziom dochodu ma następującą postać:

   $$H_0: \beta_{wiek} = \beta_{wiek\_2} = 0$$

   W celu obliczenia wartości statystyki testowej korzystamy z modelu wyjściowego (model bez ograniczeń) i modelu z ograniczeniami, który otrzymamy po usunięciu z wyjściowej regresji zmiennej *wiek* i *wiek_2*.

   Obliczamy statystykę testową zgodnie ze wzorem:

   $$F = \frac{(S_R - S)/g}{S/(N-K)} = \sim F(g, N-K)$$

   gdzie: $S_R -$ suma kwadratów reszt modelu z ograniczeniami
   $S -$ suma kwadratów rest zmodel bez ograniczeń
   $g -$ liczba ograniczeń

   Jeśli statystyka testowa jest większa od wartości krytycznej $F\sim(g, N-K)$, to odrzucamy hipotezę zerową.
3. Zmienna *wykształcenie* będzie skorelowana ze zmiennymi zerojedynkowymi dotyczącymi wykształcenia. Najprawdopodobniej wystąpi problem współliniowości.
4. Po dodaniu do modelu zmiennej *srednie_1* macierz X nie będzie miała pełnego rzędu kolumnowego, czyli nie będzie można odwrócić macierzy $X'X$ i tym samym dokonać estymacji takiej regresji. Wystąpi problem dokładnej współliniowości.
5.
   a) Kobieta rasy białej o wykształceniu wyższym zarabia o 37,51% więcej niż kobieta rasy białej o wykształceniu średnim.

   b) Kobieta rasy innej niż biała o wykształceniu średnim zarabia o 39,53% więcej niż kobieta rasy innej niż biała o wykształceniu podstawowym.

6. Hipoteza ma następującą postać:

$$H_0: \beta_{rasa} + \beta_{srednie\_rasa} + \beta_{wyzsze\_rasa} = 0$$

W celu obliczenia wartości statystyki testowej korzystamy z modelu wyjściowego (model bez ograniczeń) i modelu z ograniczeniami.
Obliczamy statystykę testową zgodnie ze wzorem:

$$F = \frac{(S_R - S)/g}{S/(N-K)} = \sim F(g, N-K)$$

gdzie: $S_R -$ suma kwadratów reszt modelu z ograniczeniami
$S -$ suma kwadratów rest zmodel bez ograniczeń
$g -$ liczba ograniczeń

Jeśli statystyka testowa jest większa od wartości krytycznej $F \sim (g, N-K)$, to odrzucamy hipotezę zerową.