Egzamin z ekonometrii 2.02.2010

Pytania teoretyczne

1. Pokaż, że w modelu ze stałą średnia wartość zmiennej zależnej równa jest średniej z wartości dopasowanych.
2. Podaj definicję elastyczności cząstkowej.
3. Wyprowadź postać macierzy wariancji kowariancji estymatora MNK.
4. Kiedy mówimy, że w modelu występuje problem równoczesności? Jakie są jej dwie najczęstsze przyczyny i jakie ma ona konsekwencje dla własności MNK?

ZADANIE 1 Załóżmy, że spełnione są założenia KMRL

1. Pokaż, że wartość oczekiwana $\mathrm{E}(\boldsymbol{e}) = 0$, gdzie $\boldsymbol{e}$ jest wektorem reszt z modelu oszcowanego MNK.
2. Oblicz estymator MNK b_e z regresji $\boldsymbol{y}$ na $\boldsymbol{e}$.

Rozwiązanie:

1. Jedno rozwiązanie
$$\mathrm{E}(\boldsymbol{e}) = \mathrm{E}(\boldsymbol{M}_x\boldsymbol{\varepsilon}) = \boldsymbol{M}_x\underbrace{\mathrm{E}(\boldsymbol{\varepsilon})}_{0} = 0$$
Inne rozwiązanie
$$\mathrm{E}(\boldsymbol{e}) = \mathrm{E}(\boldsymbol{y}-\widehat{\boldsymbol{y}}) = \mathrm{E}(\boldsymbol{x\beta} + \boldsymbol{\varepsilon} - \boldsymbol{Xb}) = \boldsymbol{X\beta}+\underbrace{\mathrm{E}(\boldsymbol{\varepsilon})}_{0} - \boldsymbol{x}\underbrace{\mathrm{E}(\boldsymbol{b})}_{\boldsymbol{\beta}} = \boldsymbol{0}$$

2. Jeden możliwy dowód: ponieważ $\boldsymbol{e'y} = \boldsymbol{e'}(\boldsymbol{Xb}+\boldsymbol{e}) = \underbrace{\boldsymbol{e'X}}_{0}\boldsymbol{b}+e'\boldsymbol{e} = \boldsymbol{e'e}$ a więc $b_e = (e'\boldsymbol{e})^{-1}\boldsymbol{e'}\,\boldsymbol{e} = 1$.

   Inny dowód:
$$b_e = (\boldsymbol{e'e})^{-1}\boldsymbol{e'}\,\boldsymbol{y} = \frac{\boldsymbol{y'M}_x\boldsymbol{y}}{\boldsymbol{y'M}_x\boldsymbol{y}} = 1$$
   przy czym skorzystaliśmy z tego, że macierz $\boldsymbol{M}_x$ jest idempotentna. Nieco trudniejszy jest dowód jeśli przyjmiemy, że w modelu jest także stała. Wtedy
$$\boldsymbol{b}_e = (\boldsymbol{e'e})^{-1}\boldsymbol{e'}\,\boldsymbol{y} = \begin{bmatrix} \boldsymbol{1'1} & \boldsymbol{e'1} \\ \boldsymbol{1'e} & \boldsymbol{e'e} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{e'y} \\ \boldsymbol{e'y} \end{bmatrix}$$
$$= \begin{bmatrix} \frac{\boldsymbol{e'y}}{N} \\ (\boldsymbol{e'e})^{-1}\boldsymbol{e'y} \end{bmatrix} = \begin{bmatrix} \frac{\boldsymbol{e'e}}{N} \\ 1 \end{bmatrix}$$

ZADANIE 2 Badacz oszacował model, który wyjaśnia stopę wzrostu wyrażoną w procentach (growth), stosunkiem zagranicznych inwestycji bezpośrednich do PKB (FDI_GDP) wyrażonym w procentach, procentem populacji która jest piśmienna ($literacy$), stosunkiem inwestycji zagranicznych do PKB (inv_GDP) wyrażonym w procentach oraz zmienną zerojedynkową związaną z członkowstwem w Unii Europejskiej ($eu = 0$ nie jest członkiem, $eu = 1$ jest członkiem). Uzyskane przez badacza wyniki regresji wraz z wynikami testów diagnostycznych znajdują się poniżej

growth	b	se	t	p
FDI_GDP	-0.001	0.013	-0.055	0.956
literacy	0.050	0.020		
inv_GDP	0.124	0.049	2.520	0.013
eu	-1.176176	1.040	-1.130	0.261
stała	-2.336	1.847	-1.265	0.208
N	123			
F(4, 118)	3.840			0.006
R^2	.115			

Breusch-Pagan test chi2(1) = 31.68 [0.0000]
Jarque - Bera test chi2(2) = 2.71 [0.2585]
RESET F(3, 115) = 1.13 [0.3392]

UWAGA: testy przeprowadzamy na poziomie istotności $\alpha = 0.05$. Odpowiedzi należy uzasadnić wielkościami odpowiednich statystyk testowych.

1. Zintepretuj wielkość R^2. Sprawdź, czy zmienne w modelu są łącznie istotne.
2. Policz wielkość statystyki t dla zmiennej $literacy$. Zweryfikuj hipotezę o istotności tej zmiennej oraz o istotności zmiennej FDI_GDP.

   Podpowiedź: $t_{0.975}(118) = 1.96$.
3. Zintepretuj wielkości współczynników przy zmiennych inv_GDP oraz eu.
4. Badacz postanowił sprawdzić, poprawność obserwacji użytych w modelu. Policzył w tym celu dźwignię, standaryzowane reszty oraz odległość Cooka i uporządkował obserwacje według odległości Cooka. Czy w uzyskanej tabeli zawierającej obserwacje dla 3 największych odległości Cooka jest coś niepokojącego? W jaki sposób tego rodzaju problem należy rozwiązać?

	FDI_gdp	growth	inv_gdp	eu	literacy	cook	dźwignia	reszta stand.
Gwinea Równikowa	19.331	-7.758	41.654	0	87.0	.187	.058	-3.872
Azerbejdżan	-2.941	33.032	31.560	0	99.3	.246	.026	6.789
Luksemburg	304.948	5.016	17.797	1	99.0	23.109	.962	2.130

5. Zinterpretuj wielkość statystyk diagnostycznych.
6. Jakie są konsekwencje problemu wykrytego przez statytstyki diagnostyczne? Wymień przynajmniej dwie metody radzenia sobie z tym problemem.

Rozwiązanie:

1. 11.5% zmienności zmiennej zależnej została wyjaśniona w modelu za pomocą zmiennych niezależnych. Hipoteza o nieistotności wszystkich zmiennych należy odrzucić wartość p = 0.006<0.05.
2. Statystyka t dla zmiennej literacy jest równa $\frac{0.050}{0.020} = 2.5 > 1.96$, hipotezę o nieistotności tej zmiennej należy odrzucić. Dla zmiennej FDI_GDP, wartość p jest równa 0.956>0.05, w tym przypadku hipotezy o nieistotności nie da się odrzucić.
3. Wzrost sotosunku iwestycji do PKB o jeden punkt % powoduje podniesienie stopy wzrostu gospodarczego o 0.124 punktu procentowego, kraje przynależące do EU mają wzrost cetearis paribus niższy od krajów nielażących do tego ugrupowania o 1.176 punktu procentowego.
4. Bardzo niepokojąca jest wielkość statyki Cooka dla Luksemburga. Jest ona znacznie większa zarówno od 0.5 jak i od 1. Obserwacja ta jest więc bardzo wpływowa. Wydaje się jednak równośnie obserwacją błędną ponieważ trudno sobie wyobrazić by w roku 2006 bezpośrednie inwestycje w Luksemburgu były trzykrotnie większe od jego dochodu narodowego. Obserwację tę należy jeszcze raz sprawdzić w razie ustalenia, że jest błędna usunąc z próby.
5. Wielkość statystyki Breusch-Pagana zmusza na do odrzucenia H_0 o homoskedastyczności i przyjęcia H_1 o heteroskedastyczności (wartość p = 0.000<0.05), wartość statstyki dla testu Jarque-Bera nie pozwala na odrzucenie H_0 o normalności rozkładu czynnika losowego (0.2585>0.05), wartość statytyki hipotezy RESET nie pozwala na odrzucenie hipotezy o poprawności formy funkcyjnej (0.3392>0.05)
6. Testy diagnostyczne wykryły heteroskedastyczność w modelu. Konsekwencją heteroskedastyczności są błędne (obciążone i niezgodne) oszacowania macierzy wariancji kowariancji MNK. Z problem tym można sobie poradzić przy użyciu SUMNK lub odpornej macierzy wariancji kowariancji.

ZADANIE 3 Badacz postanowił oszacować wpływ zmian w systemie zasiłków rodzinnych na zachowania prokreacyjne. Badacz oznaczył zmienną otrzymywanie zasiłku przez z przyjmującą wartość 1 jeśli gospodarstwo miało prawo do zasiłku i 0 jeśli nie miało prawa do zasiłku. Pod koniec 2005 dochód na członka gospodarstwa, poniżej którego wypłacano zasiłek rodzinny został obniżony z 548 zł do 504 zł. W konsekwencji rodziny o dochodzie między 504 i 548 utraciły prawo do zasiłku. Badacz dysponuje obserwacjami z roku 2005 i z roku 2006 dotyczącymi liczby narodzin y i sytuacji społeczno ekonomicznej rodzin o dochodzie niższym niż 548 zł. Dla tych obserwacji badacz zdefiniwał zmienną zerojedynkową g, przyjmującą wartość 1 jeśli dana rodzina miała dochód w przedziale 504 do 548 i 0 dla rodzin o niższym dochodzie, oraz zmienną t, która przyjmuje wartość 0 dla obserwacji z roku 2005 i 1 dla obserwacji z roku 2006. Badacz poczynił następujące założenia upraszczające:

a. Oczekiwana liczba narodzin we wszystkich rodzinach o dochodzie niższym niż 504 w roku 2005 byłaby równa β_0, gdyby rodziny te nie otrzymywały zasiłku:

$$\mathrm{E}\left(y|\, t=0, g=0, z=0\right) = \beta_0$$

b. Oczekiwana liczba narodzin między rokiem 2005 i 2006 zmieniłaby się o tyle samo dla rodzin o dochodzie poniżej 504 i o dochodzie między 504 i 548, gdyby nie wprowadzono zmian w systemie zasiłków rodzinnych:

$$\begin{aligned} \mathrm{E}\left(y|\, t=1, g=0, z=0\right) - \mathrm{E}\left(y|\, t=0, g=0, z=0\right) &= \beta_1 \\ \mathrm{E}\left(y|\, t=1, g=1, z=0\right) - \mathrm{E}\left(y|\, t=0, g=1, z=0\right) &= \beta_1 \end{aligned}$$

c. Oczekiwana liczba narodzin w rodzinach o dochodzie niższym od 504 i dochodzie między 504 i 548 różni się o β_2. Różnica ta nie zmieniłaby się między rokiem 2005 i 2006, gdyby nie wprowadzono zmian w systemie zasiłków rodzinnych:

$$\begin{aligned} \mathrm{E}\left(y|\, t=0, g=1, z=0\right) - \mathrm{E}\left(y|\, t=0, g=0, z=0\right) &= \beta_2 \\ \mathrm{E}\left(y|\, t=1, g=1, z=0\right) - \mathrm{E}\left(y|\, t=1, g=0, z=0\right) &= \beta_2 \end{aligned}$$

d. Wpływ wprowadzonego zasiłku sumuje się z wpływem różnic dochodowych i wpływem czasu.

$$\mathrm{E}\left(y|\, t, g, z\right) = \mathrm{E}\left(y|\, t, g\right) + \beta_3 z$$

Pytania:

1. Na podstawie podanych powyżej założeń znajdź rozwiązania dla $\mathrm{E}\left(y|\, t=0, g=0, z=0\right)$, $\mathrm{E}\left(y|\, t=1, g=0, z=0\right)$, $\mathrm{E}\left(y|\, t=0, g=1, z=0\right)$, $\mathrm{E}\left(y|\, t=1, g=1, z=0\right)$ w kategoriach β_0, β_1 i β_2.
2. Na podstawie wyników uzyskanych w punkcie (1) i założenia (d) sformułuj model liniowy dla $\mathrm{E}\left(y|\, t, g\right)$.

   Podpowiedź: skorzystaj z tego, że t i g są zmiennymi zerojedynkowymi.
3. Na podstawie wyników z punktu (2) oraz założenia (d) sformułuj model dla $\mathrm{E}\left(y|\, t, g, z\right)$.
4. Zastanów się, które gospodarstwa domowe nie otrzymują zasiłku. Korzystając z tego spróbuj zapisać z jako funkcję t i g i sformułuj model równoważny modelowi z punktu (3), ale wykorzystujący jako zmienne objaśniające jedynie t i g.

Rozwiązanie:

1. Wychodząc założenia (d) dochodzimy oraz z założeń (b) i (c) dochodzimy do wniosku, że:

$$\begin{aligned} \mathrm{E}\left(y|\, t=0, g=0, z=0\right) &= \beta_0 \\ \mathrm{E}\left(y|\, t=1, g=0, z=0\right) - \mathrm{E}\left(y|\, t=0, g=0, z=0\right) &= \beta_1 \\ \mathrm{E}\left(y|\, t=1, g=1, z=0\right) - \mathrm{E}\left(y|\, t=0, g=1, z=0\right) &= \beta_1 \\ \mathrm{E}\left(y|\, t=0, g=1, z=0\right) - \mathrm{E}\left(y|\, t=0, g=0, z=0\right) &= \beta_2 \\ \mathrm{E}\left(y|\, t=1, g=1, z=0\right) - \mathrm{E}\left(y|\, t=1, g=0, z=0\right) &= \beta_2 \end{aligned}$$

   Daje to nam układ 5 równań z 4 niewiadomymi. Jedno równanie można odrzucić ponieważ jest liniowo zależne od pozostałych. Rozwiązaniem jest:

$$\begin{aligned} \mathrm{E}\left(y|\, t=0, g=0, z=0\right) &= \beta_0 \\ \mathrm{E}\left(y|\, t=1, g=0, z=0\right) &= \beta_0 + \beta_1 \\ \mathrm{E}\left(y|\, t=0, g=1, z=0\right) &= \beta_0 + \beta_2 \\ \mathrm{E}\left(y|\, t=1, g=1, z=0\right) &= \beta_0 + \beta_1 + \beta_2 \end{aligned}$$

2. Dla zmiennych zerojedynkowych t i g równania z poprzedniego punktu można zapisać w sposób następujący:
$$\mathrm{E}\left(y|\, t, g, z=0\right) = \beta_0 + \beta_1 t + \beta_2 g$$
   Korzystając znowu z założenia (d) dochodzimy do wniosku, że
$$\mathrm{E}\left(y|\, t, g, z=0\right) = \mathrm{E}\left(y|\, t, g\right) = \beta_0 + \beta_1 t + \beta_2 g$$

3. Korzystając jeszcze raz z założenia (d) wnioskujemy, że
$$\mathrm{E}\left(y|\, t, g, z\right) = \mathrm{E}\left(y|\, t, g\right) + \beta_3 z = \beta_0 + \beta_1 t + \beta_2 g + \beta_3 z$$

4. W grupie analizowanych gospodarstw zasiłku nie otrzymywały jedynie gospodarstwa o dochodzie między 504 i 548 w roku 2006. Wynika z tego, że w naszej bazie danych zmienna $z = 1 - gt$. Ostatecznie model będzie miał postać:
$$\mathrm{E}\left(y|\, t, g\right) = \beta_0 + \beta_1 t + \beta_2 g + \beta_3 \left(1 - gt\right)$$
   lub
$$y = \beta_0 + \beta_1 t + \beta_2 g + \beta_3 \left(1 - gt\right) + \varepsilon$$
   dla ε spełniającego $\mathrm{E}\left(\varepsilon|\, t, g_i\right) = 0$.