

Egzamin z ekonometrii 12.06.2009

1 semestr, Informatyka i Ekonometria

Pytania teoretyczne

1. Wyjaśnij, co to znaczy, że między zmiennymi w modelu występują interakcje.
2. Jaki skutek może mieć pominięcie istotnej zmiennej w modelu?
3. Udowodnij, że estymator MNK jest estymatorem zgodnym. Wypisz założenia konieczne do tego dowodu.
4. Opisz proces szacowania modelu SUMNK na przykładzie usuwania heteroskedastyczności z modelu bądź estymacji parametrów modelu za pomocą estymatora Prais-Winstena.

ZADANIE 1 Pokaż, że dla prawdziwych założeń KMRL oraz założenia, że $\varepsilon \sim N(0, \sigma^2 \mathbf{I})$ rozkład statystyki Durбина-Watsona w małej próbie nie zależy od wielkości nieznanymi parametrów σ^2 i β a jedynie od postaci macierzy obserwacji \mathbf{X} .

Rozwiązanie:

$$DW = \frac{\sum_{t=2}^T e_t e_{t-1}}{\sum_{t=1}^T e_t^2} = \frac{\sum_{t=2}^T \left(\frac{e_t}{\sigma}\right) \left(\frac{e_{t-1}}{\sigma}\right)}{\sum_{t=1}^T \left(\frac{e_t}{\sigma}\right)^2}$$

Jednak wektor reszt równy jest $\frac{\varepsilon}{\sigma} = \mathbf{M}_X \frac{\varepsilon}{\sigma}$ i ma rozkład $\frac{\varepsilon}{\sigma} \sim N(0, \mathbf{M}_X)$. Ponieważ statystyka DW jest funkcją wektora $\frac{\varepsilon}{\sigma}$ a rozkład wektora $\frac{\varepsilon}{\sigma}$ niezależny od σ^2 i β więc rozkład statystyki DW także nie zależy od wielkości parametrów σ^2 i β niezależnie o wielkości próby. Rozkład statystyki DW zależy jednak od postaci macierzy \mathbf{X} , ponieważ $\mathbf{M}_X = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$.

ZADANIE 2 Mamy regresje oszacowane MNK:

- a. $y_i - x_{2i}$ na stałej i x_{1i}
- b. $y_i - x_{2i}$ na stałej $x_{1i} - x_{3i}$
- c. y_i na stałej i x_{1i}, x_{2i}, x_{3i}
- d. $y_i - x_{2i}$ na stałej i x_{2i}, x_{3i}

Odpowiedz na następujące pytania:

1. Jaka zachodzi relacja między RSS w tych regresjach?
2. Jaka zachodzi relacja między R^2 w tych regresjach?

Podpowiedź: Jedynie między pięcioma parami regresji można ustalić takie relacje. Zastanów się jak można sprowadzić rozpatrywane regresje do regresji szacowanej MNK z ograniczeniami nałożonymi na elementy wektora \mathbf{b} .

Rozwiązanie: Analizowane regresje można oszacować na podstawie regresji MNK y_i na stałej x_i, x_{2i}, x_{3i} przy następujących ograniczeniach:

- a. $\beta_2 = 1, \beta_3 = 0$
- b. $\beta_2 = 1, \beta_1 = -\beta_3$
- c. bez ograniczeń
- d. $\beta_2 = 1$

Ponieważ modele z ograniczeniami mają zawsze większe RSS i mniejsze R^2 niż modele bez ograniczeń więc:

1. $RSS_a \leq RSS_c, RSS_a \leq RSS_d, RSS_b \leq RSS_c, RSS_b \leq RSS_d, RSS_d \leq RSS_c$
2. $R_a^2 \geq R_c^2, R_a^2 \geq R_d^2, R_b^2 \geq R_c^2, R_b^2 \geq R_d^2, R_d^2 \geq R_c^2$

ZADANIE 3 Oszacowano następujący model (grawitacyjny) wyjaśniający eksport z kraju i (reportera) do kraju j (partnera)

$$exp_{ij} = \beta_1 + \beta_2 rep_gdp_i + \beta_3 part_gdp_j + \beta_4 dist_{ij} + \beta_5 rep_llocked_i + \beta_6 part_llocked_j + \varepsilon_{ij},$$

gdzie exp_{ij} jest logarytmem wielkości eksportu z kraju i do kraju j , rep_gdp_i , $part_gdp_j$ odpowiednio logarytmami wielkości PKB dla kraju i oraz kraju j , zmienna $dist_{ij}$ oznacza logarytm odległości między krajami i oraz j a zmienne zerojedynkowe $rep_llocked_i$ i $part_llocked_j$ przyjmują wartość 1 w przypadku, gdy odpowiednio reporter bądź partner nie mają dostępu do morza i 0 w przeciwnym przypadku. Poniżej znajdują się oszacowania uzyskane dla tego modelu na podstawie danych z 2006 roku.

Odpowiedzi proszę uzasadnić wielkością statystyk. Założony poziom istotności $\alpha = 0.05$.

Source	SS	df	MS			
Model	63553.5604	5	12710.7121	Number of obs =	6656	
Residual	24393.3521	6650	3.66817324	F(5, 6650) =	3465.13	
Total	87946.9124	6655	13.2151634	Prob > F =	0.0000	
				R-squared =	.	
				Adj R-squared =	0.7224	
				Root MSE =	1.9152	

exp	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
rep_GDP	1.220763	.0136375	89.52	0.000	1.194029	1.247497
part_GDP	.872535	.0112611	77.48	0.000	.8504597	.8946104
dist	-1.417857	.0272906	.	.	-1.471356	-1.364359
rep_llocked	.1583344	.0634953	2.49	0.013	.0338632	.2828056
part_llocked	-.8607977	.0595425	-14.46	0.000	-.9775201	-.7440754
_cons	-31.29849	.5187627	-60.33	0.000	-32.31543	-30.28155

RESET F(3, 6650) = 43.47 [0.0000]
 Breusch-Pagan chi2(1) = 681.64 [0.0000]
 Jarque-Bera chi2(2) = 22072 [0.0000]

UWAGA: w nawiasach kwadratowych znajdują się wartości p.

1. Policz i zinterpretuj wielkość R^2 . Sprawdź, czy zmienne w modelu są łącznie istotne.
2. Policz wielkość statystyki t dla zmiennej $dist_{ij}$. Zweryfikuj hipotezę o istotności tej zmiennej oraz o istotności zmiennej rep_gdp_i .
Podpowiedź: $t_{0.975}(6650) = 1.96$.
3. Zinterpretuj wielkości współczynników przy zmiennych rep_gdp_i , $dist_{ij}$ oraz $rep_llocked_i$. Czy znaki tych współczynników są zgodne z twoją intuicją? Odpowiedź uzasadnij.
4. Do modelu dodano dodatkowe zmienne objaśniające: ilość ziemi uprawnej w kraju i oraz kraju j . Uzyskano wielkość RSS na poziomie 24357. Przeprowadź test łącznej istotności tych dwóch zmiennych.
Podpowiedź: $F_{0.95}(2, 6650) = 2.997$.
5. Zinterpretuj wielkość statystyk diagnostycznych.
6. Poniżej znajdują się oszacowania uzyskane z użyciem odpornej macierzy wariancji White'a. W jakim celu przeprowadzono tę dodatkową estymację?

	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
lGross_Exp						
lrep_GDP	1.220763	.0149171	81.84	0.000	1.191521	1.250006
lpart_GDP	.872535	.0116294	75.03	0.000	.8497376	.8953325
ldist	-1.417857	.0266666	-53.17	0.000	-1.470132	-1.365582
rep_land_l~d	.1583344	.0622102	2.55	0.011	.0363823	.2802864
part_land_~d	-.8607977	.0630951	-13.64	0.000	-.9844845	-.737111
_cons	-31.29849	.553028	-56.59	0.000	-32.3826	-30.21438

7. W jaki sposób próbowałbyś rozwiązać problemy, które ujawnione zostały przez testy diagnostyczne? Odpowiedź uzasadnij.

Rozwiązanie:

1. $R^2 = \frac{63553.5604}{87946.9124} = 0.72264$. Zmienność zmiennej objaśnianej została wyjaśniona w 72% przez zmienność zmiennych objaśniających w modelu. Hipoteza o łącznej nieistotności zmiennych jest odrzucana [wartość p = 0.000]
2. $t = \frac{-1.417857}{0.0272906} = -51.954$. Hipoteza zerowa o nieistotności zmiennej $dist_{ij}$ jest odrzucana ponieważ $|-51.954| > 1.96$.
3. 1% wzrost pkb kraju reportera zwiększa eksport do partnera o 1.22%. Wynik ten wydaje się intuicyjny - większe i bogatsze kraje eksportują więcej. Wzrost odległości między krajami o 1% zmniejsza eksport z kraju i do kraju j o -1.4% . Wynik ten wydaje się intuicyjny - im większa odległość, tym większy koszt transportu, który zniechęcać będzie do eksportu. Kraj reporter nie posiadający granicy morskiej ma większy eksport niż kraj posiadający granicę morską i pozostałe charakterystyki takie same o 16%. Ten wynik wydaje się sprzeczny z intuicją - transport morski jest tani, więc kraje z dostępem do morza powinny mieć wyższy eksport.
4. $F = \frac{(S_R - S)/g}{S/(n-k)} = \frac{(24393 - 24357)/2}{24357/6650} = 4.9144 > 2.997$ a więc hipoteza o łącznej nieistotności zmiennych ilość ziemi uprawnej w kraju i oraz kraju j jest odrzucana.
5. Test RESET odrzuca H_0 o poprawności formy funkcyjnej (wartość p = 0.000), test Breusch-Pagana odrzuca H_0 o homoskedastyczności, test Jarque-Bera odrzuca H_0 o normalności rozkładu czynnika losowego.
6. Estymator White'a macierzy wariancji kowariancji estymatora MNK jest odporny na heteroskedastyczność. Dodatkową regresję przeprowadzono, ponieważ w naszym modelu występuje heteroskedastyczność i użycie tej metody umożliwiłoby uzyskanie poprawnych oszacowań. Niestety metoda ta nie rozwiązuje problemu związanego z niepoprawnością formy funkcyjnej.
7. Problem związany z niepoprawnością formy funkcyjnej można rozwiązać modyfikując model i sprawdzając inne formy funkcyjne. Problem związany z heteroskedastycznością można rozwiązać stosując UMNK lub odporną macierz wariancji kowariancji White'a.