

Egzamin z ekonometrii 22.06.2012

Pytania teoretyczne

1. Podaj ogólną postać modeli *DL* i *ADL*
2. Wyjaśnij jak należy rozumieć przyczynowość w sensie Grangera i jak jest testowana.
3. Jakie są wady liniowego modelu prawdopodobieństwa? Odpowiedź uzasadnij.
4. Kiedy mówimy, że w modelu występuje równoczesność? Za pomocą jakiej statystyki można przetestować występowanie równoczesności? Jakiej metody estymacji powinno się użyć w przypadku jej występowania?

ZADANIE 1 Przy użyciu danych dotyczących 131 krajów świata z lat 1990-2006 badacz zbudował panel niezbilansowany. Przy użyciu tego panelu badacz oszacował model, którego celem jest wyjaśnienie zróżnicowania poziomu PKB per capita w poszczególnych krajach. Zmienną zależną jest zmienna *gdp_pc* oznaczająca logarytm PKB per capita (wyrażone w tysiącach dolarów) a zmiennymi objaśniającymi poziom alfabetyzacji (*lit* - wyrażony w procentach udział ludności umiejącej czytać), *roads_pc* (liczba kilometrów dróg na głowę), *roads_p* (procent dróg o nawierzchni utwardzonej), trend czasowy (*year* - rok z którego pochodzą dane). Dodatkowo do regresji włączono zmienną zerojedynkową *continent* związany z geograficznym położeniem danego kraju (1 Afryka, 2 Ameryka, 3 Azja, 4 Europa, 5 rejon Pacyfiku). Model został przez badacza oszacowany za pomocą MNK, estymatora efektów losowych i estymatora efektów stałych. Wyniki estymacji znajdują się na następnej stronie.

Założony poziom istotności przy testowaniu hipotez statystycznych $\alpha = 0.05$. Uzyskane wyniki testów należy uzasadnić wielkościami odpowiednich statystyk bądź wartościami *p*.

1. Wyjaśnij, dlaczego badacz użył w regresji MNK odpornego warstwowego estymatora macierzy wariancji i kowariancji.
2. Weźmy pod uwagę jedynie wyniki dla MNK i estymatora efektów stałych. Który z nich jest estymatorem efektywnym w przypadku rozpatrywanego problemu? Odpowiedź uzasadnij odpowiednią statystyką testową.
3. Na podstawie znajdujących się na wydruku statystyk testowych wybierz spośród estymatorów POLS, RE i FE estymator, który powinno się użyć w kontekście analizowanego problemu. Wyjaśnij jakie hipotezy zerowe testujemy za pomocą użytych testów.
4. Dlaczego w przypadku estymatora efektów stałych nie udało się oszacować współczynnika dla zmiennej *continent*?
5. Zinterpretuj wielkości wszystkich trzech statystyk R^2 uzyskanych dla estymatora efektów stałych.
6. Zinterpretuj wielkość współczynników przy zmiennych *lit*, *roads_pc* i *year* w modelu efektów stałych.
7. Czy w modelu efektów stałych poprawne byłoby pominięcie efektów indywidualnych dla krajów? Odpowiedź uzasadnij wielkością odpowiedniej statystyki.
8. Dlaczego zarówno w modelu efektów stałych jak i w modelu efektów losowych uzyskujemy dwa oszacowania błędów standardowych czynników losowych (*sigma_u*, *sigme_e*)?

Linear regression

Number of obs = 1341
F(8, 130) = 38.21
Prob > F = 0.0000
R-squared = 0.5872
Root MSE = .90472

(Std. Err. adjusted for 131 clusters in iso_num)

		Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]
gdp_pc						
lit		1.275277	.5295954	2.41	0.017	.2275362 2.323018
roads_p		.0231543	.003162	7.32	0.000	.0168988 .0294099


```

      lit | -3.135022   .3397581   -9.23   0.000   -3.801605   -2.46844
    roads_p | .0013333   .000923    1.44   0.149   -.0004775   .0031441
    roads_pc | .0801028   .0287958    2.78   0.005    .0236074   .1365982
          |
    continent |
      2 | (omitted)
      3 | (omitted)
      4 | (omitted)
      5 | (omitted)
          |
      year | .0487547   .0025059   19.46   0.000    .0438382   .0536711
    _cons | -94.18744   4.828643  -19.51   0.000  -103.6609  -84.71396
-----+-----
    sigma_u | 1.8640004
    sigma_e | .20572419
      rho | .98796571   (fraction of variance due to u_i)
-----+-----
F test that all u_i=0:      F(130, 1206) =   188.88      Prob > F = 0.0000

```

Test Hausmana

```

-----+-----
          |          Coefficients          |
          |          (b)          (B)          |
          |          fe          re          |
          |          Difference          |
          |          sqrt(diag(V_b-V_B))          |
          |          S.E.          |
-----+-----
      lit | -3.135022   -1.154482   -1.98054   .1797152
    roads_p | .0013333   .0043875   -.0030542   .0001079
    roads_pc | .0801028   .1488956   -.0687928   .0032632
      year | .0487547   .0362314   .0125233   .0010193
-----+-----

```

b = consistent under Ho and Ha; obtained from xtreg
B = inconsistent under Ha, efficient under Ho; obtained from xtreg

Test: Ho: difference in coefficients not systematic

$$\begin{aligned}
\chi^2(4) &= (b-B)' [(V_b-V_B)^{-1}] (b-B) \\
&= 49.61 \\
\text{Prob}>\chi^2 &= 0.0000 \\
& (V_b-V_B \text{ is not positive definite})
\end{aligned}$$

Rozwiązanie:

1. Badacz użył w regresji MNK warstwowego estymatora odpornego, ponieważ w przypadku regresji na panelu można spodziewać się niediagonalnej macierzy wariancji i kowariancji ze względu na występowanie efektów indywidualnych
2. Estymatorem nieobciążonym i efektywnym jest w tym przypadku estymator efektów stałych, z wartości statystyki $F = 188.88$ [0.0000] dla hipotezy $H_0 : u_i = 0$ dla $i = 1, \dots, N$ wynika, że efekty indywidualne są istotne w modelu, zatem estymator efektów stałych jest nieobciążony i efektywny w przeciwieństwie do estymatora MNK.
3. Podstawą wyboru odpowiedniego estymatora w rozpatrywanym przypadku powinien być wynik testu Hausmana. Hipotezą zerową w tym teście jest warunek konieczny dla zgodności estymatora efektów losowych to jest brak korelacji między efektem indywidualnym a zmiennymi objaśniającymi $\text{Cov}(u_i, \mathbf{X}_i) = 0$. W naszym przypadku hipoteza ta jest odrzucana 49.61 [0.0000] a tym samym jedynym zgodnym estymatorem jest estymator efektów stałych.

4. Za pomocą estymatora efektów stałych nie jest możliwe oszacowanie wpływu zmiennych, które nie zmieniają się w czasie. Zmienna zerojedynkowa oznaczająca, że dany kraj był bądź jest krajem komunistycznym nie zmienia się w czasie.
5. Wielkość R_{within}^2 oznacza, że 30% zróżnicowania wewnątrz obiektowego (to jest zmian wartości PKB per capita) udało się wyjaśnić za pomocą zróżnicowania zmiennych objaśniających dla tego kraju. Wielkość $R_{between}^2$ oznacza, że 31% zróżnicowania PKB per capita między krajami udało się wyjaśnić za pomocą różnic w wielkościach zmiennych objaśniających pomiędzy krajami, $R_{overall}^2$ oznacza, że 28% całkowitej zmienności zmiennej zależnej udało się wyjaśnić zmiennością zmiennych niezależnych.
6. Wzrost poziomu alfabetyzacji o 1 punkt procentowy zmniejsza PKB per capita o 3.13% (ten wniosek z estymowanego modelu przeczy intuicji), zwiększenie udziału dróg utwardzonych o 1 punkt procentowy powoduje wzrost PKB per capita o 0.001%, wzrost długości dróg per capita o 1 km na głowę powoduje wzrost PKB na głowę o 0.08%.
7. W modelu efektów stałych nie można pominąć charakterystyk indywidualnych krajów ponieważ są one istotne co wnioskujemy z wyniku testu, że wszystkie $u_i = 0$ (statystyka 188.88 [0.0000]) a zarazem wiemy z wyniku testu Hausmana, że efekty indywidualne są skorelowane ze zmiennymi objaśniającymi. Pominięcie efektów indywidualnych wywołałoby w tym przypadku pojawienie się problemu endogeniczności.
8. Ponieważ w przypadku liniowego modelu efektów nieobserwowalnych mamy dwa składniki losowe: efekt indywidualny u_i oraz błąd czystolosowy ε_i . Oszacowane odchylenia standardowe odpowiadają odchyleniom standardowym u_i oraz ε_i .

ZADANIE 2 Na podstawie danych z badania dochodów i wydatków konsumpcyjnych starano się wyjaśnić wielkość miesięcznych wydatków poniesionych w restauracjach i kawiarniach przez osoby samotne za pomocą wysokości dochodu, wieku i płci. Poniżej znajdują się oszacowania parametrów i efektów cząstkowych dla modelu tobitowego oszacowanego dla tego problemu.

```
Tobit regression                                Number of obs   =          5089
                                                LR chi2(3)      =          800.00
                                                Prob > chi2     =          0.0000
Log likelihood = -4375.8865                    Pseudo R2       =          0.0838
```

wyd	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
dochg	.0555346	.0034797	15.96	0.000	.048713	.0623562
wiek	-3.371001	.2278059	-14.80	0.000	-3.817598	-2.924403
1.kobieta	-54.79897	7.322538	-7.48	0.000	-69.15429	-40.44364
_cons	1.77951	13.57973	0.13	0.896	-24.84262	28.40164
/sigma	136.1669	4.530283			127.2856	145.0482

Obs. summary: 4530 left-censored observations at wyd<=0
 559 uncensored observations
 0 right-censored observations

Efekt cząstkowy dla Pr(wyd>0)

	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
dochg	.0000597	3.78e-06	15.81	0.000	.0000523	.0000671
wiek	-.0036249	.0002232	-16.24	0.000	-.0040623	-.0031875
kobieta	-.0661504	.0097708	-6.77	0.000	-.0853007	-.0470001

Efekt czastkowy dla E(wyd*)

	dy/dx	Std. Err.	z	P> z	[95% Conf. Interval]	
dochg	.005689	.0003851	14.77	0.000	.0049341	.0064438
wiek	-.3453266	.0261011	-13.23	0.000	-.3964837	-.2941695
kobieta	-6.279588	.9481177	-6.62	0.000	-8.137864	-4.421311

Testy przeprowadzamy na poziomie istotności $\alpha = 0.05$. Wyniki testów należy uzasadnić liczbami z wydruku.

1. Podaj założenia modelu tobitowego.
2. Jaką nietypową cechą będzie się najprawdopodobniej charakteryzować rozkład zmiennej zależnej?
3. Dlaczego policzenie regresji liniowej dla tej całej obserwowanej próby da najprawdopodobniej wartości dopasowane, które dla części obserwacji będą nieinterpretowalne?
4. Zbadaj istotność poszczególnych zmiennych w modelu oraz ich łączną istotność.
5. Zinterpretuj efekty cząstkowe dla dochodu i wieku policzone dla zmiennej obserwowalnej.
6. Zinterpretuj efekty cząstkowe policzone dla prawdopodobieństwa niezerowych wydatków.
7. Poniżej znajduje się oszacowanie elastyczności dochodowej wydatków w restauracjach i kawiarniach policzone dla średniej wartości dochodu w próbie (a więc $\frac{\partial E(y|x)}{\partial x} / \frac{y}{x}$ policzone dla $x = \bar{x}$). Zinterpretuj tę wartość i oceń, czy jest ona zgodna z intuicją ekonomiczną.

Elastycznosc dla E(wyd*)

	ey/ex	Std. Err.	z	P> z	[95% Conf. Interval]	
dochg	.8098296	.0492449	16.44	0.000	.7133113	.9063479

8. Do zmiennych w modelu dodano zmienną $wiek^2$ i uzyskano wielkość logarytmu funkcji wiarygodności na poziomie -4363.6564 . Zweryfikuj hipotezę, że ta dodatkowa zmienna jest nieistotna w modelu.

Podpowiedź: $\chi_{0.95}^2(1) = 3.84$, $\chi_{0.95}^2(2) = 5.99$, $\chi_{0.95}^2(3) = 7.81$.

Rozwiązanie:

1. Założenia modelu tobitowego są następujące:

$$y_i^* = \mathbf{x}_i \boldsymbol{\beta} + \varepsilon_i$$

$$\varepsilon_i \sim N(0, \sigma^2)$$

$$y_i = \begin{cases} y_i^* & \text{dla } y_i^* > 0 \\ 0 & \text{dla } y_i^* \leq 0 \end{cases}$$

i poszczególne ε_i są niezależne.

2. Nie wszystkich stać na jedzenie poza domem, więc dla niektórych gospodarstw wydatki poniesione w restauracjach i kawiarniach wyniosą 0. Na podstawie wydruku widzimy, że na 5089 przebadanych gospodarstw, jedynie 559 poniosło w danym miesiącu wydatki tego typu.
3. W przypadku oszacowania dla tego modelu zwykłej regresji liniowej najprawdopodobniej część wartości dopasowanych będzie ujemna - co jednak jest bez sensu ponieważ wydatki mogą być wyłącznie zerowe bądź dodatnie.

4. Wszystkie zmienne w modelu są istotne (wartość p są dla wszystkich zmiennych mniejsze od 0.05), stała jest nieistotna ($0.896 > 0.05$). Łącznie wszystkie zmienne są istotne o czym świadczy wartość p dla statystyki LR ($0.000 < 0.05$).
5. Oczekiwanyrost wydatków odczytujemy z tablicy efektów cząstkowych dla y^* . Wzrost dochodu o 1000 zł. powoduje oczekiwany wzrost wydatków na restauracje i kawiarnie o 5 zł na miesiąc, wzrost wieku o 1 rok powoduje spadek tego typu wydatków o 34 gr.
6. Na podstawie wydruku efektów cząstkowych dla prawdopodobieństwa dochodzimy do wniosku, że prawdopodobieństwo poniesienia większych od zera wydatków w restauracjach i kawiarniach wzrośnie o 5.97 punkta procentowego jeśli dochód wzrośnie o 1000 zł, wzrost wieku o rok powoduje spadek tego p-stwa o 0.36 punktu procentowego oraz, że kobiety mają 6.61 punktu procentowego niższe p-stwo poniesienia tego typu wydatków.
7. Oszacowana wielkość oznacza, że przy 1% wzroście dochodu oczekiwane wydatki na restauracje i kawiarnie wzrosną o 0.80%. Oszacowanie to wydaje się niskie ponieważ wydatki tego typu wydają się raczej wydatkami na dobra luksusowe (powinny rosnąć wraz z dochodem) - a więc oszacowanie elastyczności powinno wyjść większe od 1.
8. Statystyka testu LR ma postać: .

$$LR = 2(-4375.8865 + 4363.6564) = 24.46 > \chi_{0.95}^2(1) = 3.84$$

Testujemy zerowość 1 współczynnika (a więc właściwą wartością krytyczną jest $\chi_{0.95}^2(1)$). Odrzucamy H_0 o nieistotności dochodu².

ZADANIE 3 Mamy następujące procesy

$$y_t = \phi y_{t-1} + \varepsilon_{1t} + \theta_1 \varepsilon_{1t-1}$$

$$x_t = x_{t-1} + \varepsilon_{2t}$$

$$p_t = \alpha_1 x_t + \alpha_2 y_t + u_{1t}$$

$$q_t = \gamma y_t + u_{2t}$$

gdzie $\varepsilon_{1i} \sim IID(0, \sigma^2)$, $\varepsilon_{2i} \sim IID(0, \sigma^2)$, $u_{1i} \sim IID(0, \sigma^2)$, $u_{2i} \sim IID(0, \sigma^2)$ i $\varepsilon_1, \varepsilon_2, u_1, u_2$ nieskorelowane.

1. Jakie warunki musi spełniać ϕ , żeby miało sens badanie kointegracji między p_t i q_t ?
2. Jaki warunek musi koniecznie spełniać α_1 , by mogła zaistnieć kointegracja między p_t i q_t ?
3. Jeśli warunki sformułowane powyżej są spełnione, jaką wartość przyjmie wektor kointegrujący β ?

Rozwiązanie:

1. Badanie kointegracji będzie miało sens jeśli procesy będą $I(1)$. Aby q_t mogło być $I(1)$, zmienna y_t musi być $I(1)$. Z kolei y_t może być $I(1)$ tylko wtedy, gdy $\phi = 1$.
2. Szukamy takiej kombinacji liniowej p_t i q_t , która jest stacjonarna.

$$\begin{aligned} p_t - \beta q_t &= \alpha_1 x_t + \alpha_2 y_t + u_{1t} - \beta \gamma y_t - \beta u_{2t} \\ &= \alpha_1 x_t + (\alpha_2 - \beta \gamma) y_t + u_{1t} - \beta u_{2t} \end{aligned}$$

Zuważmy, że x_t i y_t są niekorelowane. Kombinacja liniowa niekorelowanych zmiennych $I(1)$ musi być $I(1)$. Jedyną możliwością uzyskania kointegracji jest $\alpha_1 = 0$ i $\alpha_2 - \beta \gamma = 0$.

3. Wartość $\beta = \frac{\alpha_2}{\gamma}$. Wektor kointegrujący to taki wektor β , że $\beta x_t \sim I(0)$. W tym przypadku $x_t = \begin{bmatrix} p_t \\ q_t \end{bmatrix}$ a wektor kointegrujący ma postać

$$\beta = \begin{bmatrix} 1 \\ -\beta \end{bmatrix} = \begin{bmatrix} 1 \\ -\frac{\alpha_2}{\gamma} \end{bmatrix}$$

wektorem kointegrującym będzie też każdy wektor postaci $a\beta$ dla dowolnego a .