

Egzamin z ekonometrii IiE 22.06.2012

Pytania teoretyczne

1. Kiedy selekcja próby jest problemem i jaki model można stosować w przypadku samoselekcji próby?
2. Jakie są konieczne założenia, by estymator M był estymatorem zgodnym?
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)

		Robust				
	Coef.	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

roads_pc		.3857326	.1310661	2.94	0.004	.126434	.6450311
continent							
2		.804924	.194956	4.13	0.000	.4192269	1.190621
3		.331046	.2195079	1.51	0.134	-.1032241	.7653162
4		-.1632501	.3217617	-0.51	0.613	-.7998171	.4733169
6		.4860655	.2434216	2.00	0.048	.004485	.967646
year		.0214684	.0071025	3.02	0.003	.0074169	.0355199
_cons		-42.52438	14.10573	-3.01	0.003	-70.43087	-14.61788

Random-effects GLS regression		Number of obs	=	1341	
Group variable: iso_num		Number of groups	=	131	
R-sq: within	=	0.3018	Obs per group: min	=	1
between	=	0.3185	avg	=	10.2
overall	=	0.2846	max	=	17
corr(u_i, X)	=	0 (assumed)	Wald chi2(8)	=	582.01
			Prob > chi2	=	0.0000

gdp_pc		Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
lit		-1.154482	.2883367	-4.00	0.000	-1.719612 - .5893527
roads_p		.0043875	.0009166	4.79	0.000	.0025909 .006184
roads_pc		.1488956	.0286103	5.20	0.000	.0928205 .2049708
continent						
2		1.831437	.2517003	7.28	0.000	1.338114 2.324761
3		1.480978	.2240336	6.61	0.000	1.04188 1.920076
4		2.152185	.2889072	7.45	0.000	1.585937 2.718433
5		1.278694	.5734689	2.23	0.026	.1547158 2.402673
year		.0362314	.0022892	15.83	0.000	.0317446 .0407182
_cons		-71.62583	4.478067	-15.99	0.000	-80.40269 -62.84898
sigma_u		.90632673				
sigma_e		.20572419				
rho		.95100155	(fraction of variance due to u_i)			

Fixed-effects (within) regression		Number of obs	=	1341	
Group variable: iso_num		Number of groups	=	131	
R-sq: within	=	0.3243	Obs per group: min	=	1
between	=	0.3727	avg	=	10.2
overall	=	0.3348	max	=	17
corr(u_i, Xb)	=	-0.7939	F(4, 1206)	=	144.68
			Prob > F	=	0.0000

gdp_pc		Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
--------	--	-------	-----------	---	------	----------------------

```

      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)          |          (b-B)          |          sqrt(diag(V_b-V_B))          |
          |          fe          re          |          Difference          |          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 H_0 and H_a ; obtained from xtreg
B = inconsistent under H_a , efficient under H_0 ; obtained from xtreg

Test: H_0 : difference in coefficients not systematic

```

      chi2(4) = (b-B)' [(V_b-V_B)^(-1)] (b-B)
              =          49.61
    Prob>chi2 =          0.0000
(V_b-V_B is not positive definite)

```

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 $Cov(u_i, \mathbf{X}_i) = 0$. 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:

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

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

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

i poszczególne obserwacje są niezależne.

2. Samochód jest dobrem trwałym, więc nie każde gospodarstwo domowe w każdym roku kupuje samochód. W rezultacie próba będzie zawierała dużo zerowych obserwacji. Na podstawie wydruku widzimy, że na 35952 przebadanych gospodarstwach, jedynie 293 poniosło wydatki na zakup nowego samochodu.
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 dodatnie.
4. Na podstawie wydruku efektów cząstkowych dla prawdopodobieństwa dochodzimy do wniosku, że prawdopodobieństwo zakupu samochodu wzrośnie o 0.0628 punkta procentowego jeśli dochód wzrośnie o 1000 zł.

- Całkowity wzrost wydatków na samochód odczytujemy z tablicy efektów cząskowych dla bezwarunkowego y . Wzrost oczekiwanego wydatku na samochód przy 1000 zł wzroście dochodu wynosi 3.4 zł.
- Oszacowana wielkość oznacza, że przy 1% wzroście dochodu oczekiwane wydatki na samochód wzrosną o .167%. Oszacowanie to wydaje się bardzo niskie wzięwszy pod uwagę, że w Polsce samochód jest ciągle raczej dobrem luksusowym - oszacowanie elastyczności powinno wyjść większe od 1.
- 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 następujący model wielorównaniowy w formie strukturalnej:

$$\begin{aligned} C_t &= \alpha_0 + \alpha_1 Y_t + \varepsilon_{1t} \\ I_t &= \beta_0 + \beta_1 r_t + \varepsilon_{2t} \\ T_t &= \phi_0 + \phi_1 C_t + \phi_2 I_t + \varepsilon_{3t} \\ Y_t &= C_t + I_t + G_t \end{aligned}$$

gdzie C_t oznacza konsumpcję, Y_t oznacza dochód narodowy, I_t inwestycje, T_t dochody podatkowe państwa, G_t wydatki rządowe a r_t stopę procentową. Zakładamy, że G_t i r_t są egzogeniczne.

- Które równania w tym modelu są zidentyfikowane a które nie są? Odpowiedź uzasadnij.
- Jak nazywamy równanie o postaci takiej, jak równanie dla Y_t w tym modelu? Czym wyróżniają się równania tego typu?
- Wyjaśnij, z jakiego powodu w równaniu konsumpcji występuje sprzężenie zwrotne i dlaczego powoduje ono, że estymator MNK nie jest zgodny. Czy analogiczny problem występuje w przypadku równania inwestycji?
- Czy estymator wektora parametrów $\alpha' = (\alpha_0, \alpha_1)$

$$\hat{\alpha} = (\mathbf{Z}'\mathbf{X})^{-1} \mathbf{Z}'\mathbf{y}$$

gdzie $\mathbf{X}_t = (1, Y_t)$ a $\mathbf{Z}_t = (1, G_t)$ jest estymatorem zgodnym? Przy jakich założeniach? Odpowiedź uzasadnij.

- Zapisz postać równania konsumpcji formy zredukowanej powyższego modelu.
- Policz asymptotyczne obciążenie estymatora MNK parametru α_1 w równaniu konsumpcji.

Podpowiedź: wykorzystaj wzór na granicę według prawdopodobieństwa estymatora MNK w modelu z jedną zmienną objaśniającą $plim(\hat{\alpha}_1) = \alpha_1 + \frac{\text{Cov}(Y_t, \varepsilon_{1t})}{\text{Var}(Y_t)}$. Załóż, że wariancja Y_t jest znana. Zwróć uwagę na to, że kowariancja między ε_{1t} i ε_{2t} może być niezerowa.

Rozwiązanie:

1.

$$\begin{array}{ll} \text{zmiennne egzogeniczne} & 1, g_t, r_t \\ \text{zmiennne endogeniczne} & c_t, i_t, T_t, y_t \end{array}$$

$$\begin{array}{lll} K = 3 & G_1 = 2 & K_1 = 1 \\ & G_2 = 1 & K_2 = 2 \\ & G_3 = 4 & K_3 = 1 \end{array}$$

Identyfikacja:

$$\begin{array}{ll} 3 \geq 2 + 1 - 1 & (+) \\ 3 \geq 1 + 2 - 1 & (-) \\ 3 \geq 4 + 1 - 1 & (+) \end{array}$$

Drugie równanie tego modelu nie jest zidentyfikowane.

2. Jest to identyczność/tożsamość (dochodu narodowego). Charakteryzuje się brakiem parametrów do wyestymowania i brakiem błędu losowego
3. W tym równaniu występuje sprzężenie zwrotne ponieważ konsumpcja zależy od dochodu ale dochód narodowy zależy od konsumpcji (z racji na identyczność dochodu narodowego). Skoro zmienna objaśniająca dochód zależy od konsumpcji a sama konsumpcja zależy od zaburzenia losowego ε_1 to i dochód musi zależeć od wielkości ε_1 . Jeśli jednak dochód zależy do ε_1 to w równaniu konsumpcji występuje równoczesność, ponieważ zmienna objaśniająca dochód jest skorelowana z zaburzeniem losowym. W równaniu inwestycji problem ten nie występuje, ponieważ w tym równaniu jedyną zmienną objaśniającą jest stopa procentowa r_t , o której założono, że jest egzogeniczna.
4. Wskazany estymator jest to estymator MZI z użytym instrumentem G_t . Estymator ten będzie estymatorem zgodnym jeśli spełnione będą założenia, przy których estymatory MZI są zgodne. Po pierwsze G_t powinno być nieskorelowane z błędem losowym (egzogeniczne) - ale to zakładamy. Po drugie powinno być skorelowane z z dochodem narodowym Y_t korelacja ta jednak wynika z tożsamości dochodu narodowego.
5. Zredukowana forma modelu ma po lewej stronie wektor zmiennych endogenicznych a po prawej kombinację liniową wszystkich zmiennych egzogenicznych. Równanie konsumpcji w formie zredukowanej będzie miała następującą formę:

$$C_t = \pi_0 + \pi_1 G_t + \pi_2 r_t + \varepsilon_{1t}$$

6. Z podanego wzoru wynika, że asymptotyczne obciążenie estymatora MNK jest równe

$$plim(\hat{\alpha}_1) - \alpha_1 = \frac{\text{Cov}(Y_t, \varepsilon_{1t})}{\text{Var}(Y_t)}$$

Podstawiając

$$Y_t = C_t + I_t + G_t = \alpha_0 + \alpha_1 Y_t + \varepsilon_{1t} + \beta_0 + \beta_1 r_t + \varepsilon_{2t} + G_t$$

i rozwiązując dla Y_t

$$Y_t = \frac{1}{1 - \alpha_1} (\alpha_0 + \beta_0 + \beta_1 r_t + G_t + \varepsilon_{1t} + \varepsilon_{2t})$$

Liczymy kowariancję między Y_t i ε_{1t}

$$\begin{aligned} \text{Cov}(Y_t, \varepsilon_{1t}) &= \frac{1}{1 - \alpha_1} \text{Cov}(\alpha_0 + \beta_0 + \beta_1 r_t + G_t + \varepsilon_{1t} + \varepsilon_{2t}, \varepsilon_{1t}) \\ &= \frac{1}{1 - \alpha_1} [\text{Cov}(\varepsilon_{1t}, \varepsilon_{1t}) + \text{Cov}(\varepsilon_{2t}, \varepsilon_{1t})] \\ &= \frac{1}{1 - \alpha_1} [\text{Var}(\varepsilon_{1t}) + \text{Cov}(\varepsilon_{2t}, \varepsilon_{1t})] \end{aligned}$$

przy czym skorzystaliśmy z założonego braku korelacji między r_t , G_t i ε_{1t} .