

Egzamin z ekonometrii 12.09.2006

II semestr

Regulamin egzaminu

Pytania teoretyczne

1. Wyjaśnić jak należy rozumieć przyczynowość w sensie Grangera i jak należy ją testować.

x jest przyczyną w sensie Grangera y , jeśli zmienną można lepiej prognozować przy użyciu opóźnionych wartości x niż bez ich użycia. Przyczynowość w sensie Grangera testujemy, testując hipotezę łączną $H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$ w modelu

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \dots + \alpha_k y_{t-k} + \beta_1 x_{t-1} + \dots + \beta_k x_{t-k} + \varepsilon_t$$

2. Podać interpretację poszczególnych współczynników w mechanizmie korekty błędem (*ECM*).

Mechanizm korekty błędem ma postać następującego modelu:

$$\Delta y_t = \alpha (y_{t-1} - \mathbf{x}_{t-1} \boldsymbol{\beta}) + \sum_{i=1}^{k-1} \theta_i \Delta y_{t-i} + \sum_{i=0}^{k-1} \Delta \mathbf{x}_{t-i} \boldsymbol{\gamma}_i + \varepsilon_t$$

W modelu tym $[1, -\boldsymbol{\beta}']'$ jest wektorem kointegrującym, który określa zależności długookresowe między y i \mathbf{x} , α jest współczynnikiem determinującym szybkość dostosowania do stanu równowagi a θ_i i $\boldsymbol{\gamma}_i$ opisują krótkookresową dynamikę modelu.

3. Co to są ilorazy szans i dlaczego w kontekście modelu logitowego lepiej jest używać ilorazów szans niż efektów krańcowych?

Szansa to iloraz prawdopodobieństwa porażki do sukcesu. W modelu logitowym jest ona równa

$$Odd_i = \frac{e^{\mathbf{x}_i \boldsymbol{\beta}}}{1 - e^{\mathbf{x}_i \boldsymbol{\beta}}} \bigg/ \frac{1}{1 - e^{\mathbf{x}_i \boldsymbol{\beta}}} = e^{\mathbf{x}_i \boldsymbol{\beta}}$$

Iloraz szans dla wartości zmiennych niezależnych równych \mathbf{x}_a i \mathbf{x}_b jest równy

$$\frac{Odd_a}{Odd_b} = e^{(\mathbf{x}_a - \mathbf{x}_b) \boldsymbol{\beta}}$$

Jeśli chcemy zbadać wpływ k -tej zmiennej na p-stwo sukcesu to możemy policzyć iloraz szans w przypadku, gdy zmienna ta zwiększy się o 1 w stosunku do scenariusza bazowego. W takim przypadku iloraz szans będzie równy e^{β_k} . Wartość ilorazu szans dla modelu logitowego nie zależy, w przeciwieństwie do efektu krańcowego/cząstkowego od wielkości \mathbf{x} .

4. Omówić zalety i wady estymatora efektów stałych i estymatora efektów losowych.

Estymator efektów stałych:

zalety:

- zgodny nawet wtedy, gdy efekt indywidualny skorelowany ze zmiennymi objaśniającymi

wady:

- nieefektywny, jeśli efekt indywidualny nieskorelowany ze zmiennymi objaśniającymi
- nie można za jego pomocą oszacować wpływu na zmienną objaśnianą zmiennych objaśniających, które dla jednostek nie zmieniają się w czasie

Estymator efektów losowych

zalety:

- efektywny, jeśli efekt indywidualny nieskorelowany ze zmiennymi objaśniającymi

wady:

- nie jest zgodny jeśli efekt indywidualny skorelowany ze zmiennymi objaśniającymi

ZADANIE 1 Badacz chce wyestymować dynamiczną wersję Keynsowskiej funkcji konsumpcji na szeregu czasowym dla Polski w latach 1994.1 do 2004.4.

1. Na początku badacz przeprowadził regresję ΔY_t na Y_{t-1} i stałej oraz regresję ΔC_t na C_{t-1} i stałej i uzyskał następujące wyniki:

Zmienna	Współczynnik	(Błąd Stand.)	t
Y_{t-1}	-0.067	(0.057)	-1.16
Stała	14982.179	(9454.617)	1.58

Zmienna	Współczynnik	(Błąd Stand.)	t
C_{t-1}	-0.033	(0.016)	-2.03
Stała	7450.293	(2211.963)	3.37

gdzie C jest konsumpcją ogółem w cenach bieżących a Y wielkością PKB w cenach bieżących.

2. Po przeprowadzeniu tych regresji badacz wyestymował Keynsowską funkcję konsumpcji z dodatkowymi zmiennymi zerojedynkowymi związanymi z sezonowością i uzyskał następujące oszacowania:

Zmienna	Współczynnik	(Błąd Stand.)	t
Y	0.849	(0.015)	55.06
kw_2	-5986.197	(1898.578)	-3.15
kw_3	-9853.442	(1906.045)	-5.17
kw_4	-25261.582	(1966.185)	-12.85
Stała	5738.341	(2621.667)	2.19

W kolejnym kroku badacz wygenerował reszty z tak oszacowanego modelu, następnie przeprowadził regresję pierwszych różnic reszt Δe_t na resztach opóźnionych e_{t-1} i uzyskał następujące oszacowania:

Zmienna	Współczynnik	(Błąd Stand.)	t
e_{t-1}	-1.061	(0.158)	-6.70
Stała	243.258	(612.030)	0.40

3. Kończąc badanie badacz przeprowadził regresję ΔY_t na opóźnionych resztach, opóźnieniach ΔY_t oraz opóźnieniach ΔC_t i uzyskał następujący wynik:

Zmienna	Współczynnik	(Błąd Stand.)	Pr ($ t > t^*$)
e_{t-1}	-0.109	(0.177)	0.545
ΔY_{t-1}	-0.167	(0.178)	0.359
ΔY_{t-2}	-0.181	(0.145)	0.222
ΔY_{t-3}	-0.432	(0.144)	0.006
ΔY_{t-4}	0.514	(0.167)	0.005
ΔC_{t-1}	0.891	(0.233)	0.001
ΔC_{t-2}	0.569	(0.303)	0.072
ΔC_{t-3}	0.379	(0.289)	0.202
ΔC_{t-4}	-0.670	(0.283)	0.026
Stała	1209.524	(1350.171)	0.379

Pytania:

1. Wyjaśnić jaki wniosek badacz uzyskał w punkcie 1 (odpowiedź uzasadnić) i czemu służyły te regresje.
2. Wyjaśnić jaki wniosek uzyskał badacz w punkcie 2 (odpowiedź uzasadnić).
3. Podać interpretację elementów w równaniu z punktu 3.

Wartości krytyczne:

- test pierwiastka jednostkowego ADF , $\alpha = 0.05$, 25 obs.: bez wyrazu wolnego: -1.95 , z wyrazem wolnym: -3.00 , z wyrazem wolnym i trendem: -3.60 (Fuller, 1976),

- test na kointegrację *ADF*, $\alpha = 0.05$: bez wyrazu wolnego: -2.76 , z wyrazem wolnym: -3.37 (*Philips i Ouliaris, 1990*).

Rozwiązanie:

1. Pierwsze dwie regresje służą przeprowadzeniu testu Dickey-Fullera (DF), na istnienie pierwiastka jednostkowego (niestacjonarność zmiennych). Tylko dla niestacjonarnych zmiennych sens ma testowanie kointegracji. Porównując dla Y_t wielkość statystyki t z wartością krytyczną $-1.16 > -3.00$ (mamy stałą w regresji) dochodzimy do wniosku, że H_0 o istnieniu pierwiastka jednostkowego dla Y_t nie można odrzucić. Podobnie dla C_t statystyka $-2.03 > -3.00$ a więc także nie da się odrzucić H_0 o istnieniu pierwiastka jednostkowego. Wniosek: obie zmienne są niestacjonarne, można badać występowanie kointegracji między tymi zmiennymi.
2. W tym punkcie badacz bada kointegrację między zmiennymi. Jeśli występuje kointegracja między zmiennymi Y_t i C_t , to reszty z regresji C_t na Y_t powinny być stacjonarne. Wielkość statystyki t dla regresji reszt na resztach opóźnionych $-6.70 < -3.37$, że rzeczywiście reszty są stacjonarne, zatem występuje kointegracja między zmiennymi
3. W ostatniej regresji badacz estymuje mechanizm korekty błędem (ECM) objaśniający szybkość dostosowań do równowagi długookresowej oszacowanej w regresji na poziomach. Współczynnik przy opóźnionych resztach jest współczynnikiem szybkości dostosowań do stanu równowagi, pozostałe elementy związane są z dynamiką krótkookresową. Wielkość statystyki t dla współczynnika szybkości dostosowania do stanu równowagi długookresowej sugeruje, elementy związane z dostosowaniem do równowagi długookresowej są nieistotne w modelu.

ZADANIE 2 Na podstawie wyników badań Polskiego Generalnego Sondażu Społecznego starano się zidentyfikować zmienne, które wpływają na poziom szczęścia osób przebadanych. Na podstawie odpowiedzi na ankietę stworzono zmienną *hapunhap*, która przyjmuje wartość 1 dla osób, które zadeklarowały się jako bardzo szczęśliwe, 2 dla osób raczej szczęśliwych, 3 dla osób niezbyt szczęśliwych i 4 dla osób nieszczęśliwych. W modelu znalazły się następujące zmienne objaśniające: *sex* (1 mężczyzna, 2 kobieta), *age* (wiek), *marital* (1 w związku, 2 wdowiec, 3 rozwiedziony, 4 w separacji, 5 kawaler/panna). Poniżej znajdują się oszacowania wielkości parametrów i oszacowania efektów cząstkowych dla alternatywy 2 dla zmiennej *hapunhap*, modelowanej za pomocą uporządkowanego probita.

Testy przeprowadzamy na poziomie istotności $\alpha = 0.05$.

```
Ordered probit estimates                                Number of obs   =      8692
                                                       LR chi2(5)      =      945.49
                                                       Prob > chi2     =      0.0000
Log likelihood = -8036.7191                            Pseudo R2      =      0.0556
```

hapunhap	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
age	.0085278	.0008877	9.61	0.000	.006788 .0102677
_Imarital_2	.7935595	.0422345	18.79	0.000	.7107814 .8763377
_Imarital_3	.9263212	.0622339	14.88	0.000	.804345 1.048298
_Imarital_4	1.141439	.1201056	9.50	0.000	.9060366 1.376842
_Imarital_5	.441339	.0369661	11.94	0.000	.3688867 .5137912

_cut1	-.4960711	.0437028			(Ancillary parameters)
_cut2	1.484603	.0458395			
_cut3	2.652354	.0535389			

Marginal effects after oprobit

y = Pr(hapunhap==2) (predict, outcome(2))

= .67490831

variable	dy/dx	Std. Err.	z	P> z	[95% C.I.]	X
age	-.0004643	.00007	-6.58	0.000	-.000602	-.000326		46.196
_Imari~2*	-.142372	.01238	-11.50	0.000	-.166634	-.11811		.121261
_Imari~3*	-.2045613	.0211	-9.69	0.000	-.245926	-.163197		.039807
_Imari~4*	-.2879605	.04236	-6.80	0.000	-.370978	-.204943		.010124
_Imari~5*	-.0543852	.00734	-7.41	0.000	-.068779	-.039991		.155545

(*) dy/dx is for discrete change of dummy variable from 0 to 1

1. Wypisać założenia modelu uporządkowanego probita.
2. Zinterpretować wielkość pseudo R^2 . Sprawdzić, czy wszystkie zmienne w modelu są łącznie istotne.
3. Wyjaśnić dlaczego w przypadku takiego badania stosuje się model uporządkowanego probita, a nie zwykłą regresję liniową lub zwykły model probitowy.
4. Zinterpretować znaki przy zmiennych *age* oraz przy zmiennej zerojedynkowej, związanej ze zmienną *marital_5*.
5. Zinterpretować uzyskane wielkości efektów cząstkowych dla alternatywy 2 i zmiennych *age* i *marital_5*. Wyjaśnić, jaka jest relacja między znakami tych efektów cząstkowych a znakami parametrów modelu.
6. Podać listę zmiennych, które okazały się istotne w tym modelu (uzasadnić to wielkością statystyk i p-value).
7. Badany model oszacowano powtórnie po dodaniu zmiennych zerojedynkowych, związanych z poziomem wykształcenia (10 możliwych poziomów). Uzyskana wielkość logarytmu funkcji wiarygodności wyniosła -7974.7428 . Zweryfikuj hipotezę łączną, że poziom osiągniętego wykształcenia nie wpływa na poziom szczęścia badanych.

Podpowiedź: $\chi_{0.95}^2(8) = 15.51$, $\chi_{0.95}^2(9) = 16.91$, $\chi_{0.95}^2(10) = 18.31$

Rozwiązanie:

1. Zmienna ukryta

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

$$\varepsilon_i \sim N(0, 1)$$

i poszczególne obserwacje są niezależne. Obserwujemy y , który powstaje w sposób następujący:

$$\begin{aligned} y_i = 1 & \quad \text{jeśli} \quad y_i^* \leq \alpha_1 \\ y_i = 2 & \quad \text{jeśli} \quad \alpha_1 < y_i^* \leq \alpha_2 \\ & \quad \vdots \\ y_i = J & \quad \text{jeśli} \quad y_i^* > \alpha_J \end{aligned}$$

gdzie $\alpha_1 < \alpha_2 < \dots < \alpha_J$ są nieznanne.

2. 5.56% zmienności w modelu wyjaśniona przez zmienne niezależne, odrzucamy na postawie statystyki LR hipotezę o tym, że wszystkie zmienne w modelu są nieistotne [$945.49, 0.000 < 0.05$].
3. Zmienna zależna jest zmienną dyskretną o dobrze zdefiniowanym porządku i liczbie możliwych alternatyw (4 alternatywy). W tym przypadku nie możemy stosować modelu probitowego (więcej niż dwie alternatywy) nie powinniśmy też stosować regresji liniowej, ponieważ chcemy wyjaśnić prawdopodobieństwo alternatyw (wartości dopasowane z regresji liniowej będą trudne do zinterpretowania, mogą być np. ujemne)
4. Dodatni znak przy zmiennej *AGE* oznacza, że wraz ze wzrostem wieku rośnie prawdopodobieństwo że respondenta zadeklaruje się jako nieszczęśliwy i maleje prawdopodobieństwo, że zadeklaruje się jako bardzo szczęśliwy. Podobnie kawalerowie i panny z wyższym prawdopodobieństwem niż osoby w związku deklarują się jako nieszczęśliwe i rzadziej deklarują się jako bardzo szczęśliwe.

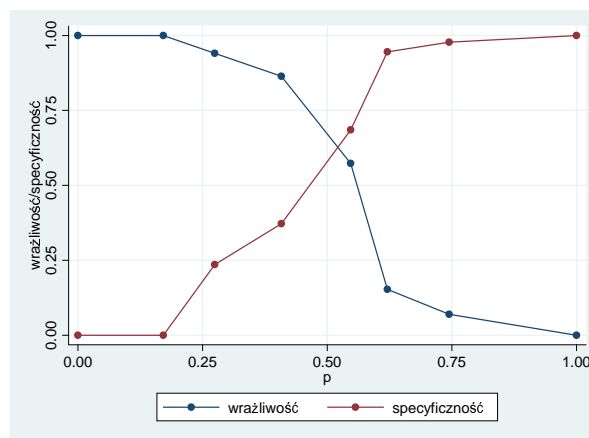
- Każdy dodatkowy rok zmniejsza życia prawdopodobieństwo zadeklarowania się jako osoba raczej szczęśliwa o .046%. Panny i kawalerowie deklarują się jako osoby raczej szczęśliwie z prawdopodobieństwem mniejszym 5.43% niż osoby pozostające w związku. W przypadku pośrednich wyborów w modelu dla zmiennych uporządkowanych nie można podać jednoznacznej relacji między wielkościami parametrów i wielkościami efektów cząstkowych.
- W modelu istotne okazały się zmienne: AGE [9.61, 0.000 < 0.05], marital_2 [18.79, 0.000 < 0.05], marital_3 [14.88, 0.000 < 0.05], marital_4 [9.50, 0.000 < 0.05], marital_5 [11.94, 0.000 < 0.05]
- Statystyka testu LR ma postać: .

$$LR = 2(-7974.7428 + 8036.7191) = 123.95 > \chi_{0.95}^2(9) = 16.91$$

Testujemy zerowość 9 współczynników (jeden poziom zmiennej dyskretnej został usunięty jako bazy) a więc właściwą wartością krytyczną jest $\chi_{0.95}^2(9)$. Odrzucamy hipotezę o tym, że łącznie wszystkie współczynniki są nieistotne. Wykształcenie istotnie wpływa na deklarowany poziom szczęścia badanych respondentów.

ZADANIE 3 Oszacowano model probitowy wyjaśniający fakt wykonywania pracy (0 - nie pracował, 1 pracował). Analizowana próba zawierała obserwacje dotyczące pracujących i bezrobotnych z Badania Ekonomicznej Aktywności Ludności (BAEL). Oszacowania uzyskane z modelu ze zmiennymi objaśniamić płęć (0 mężczyzna, 1 kobieta) oraz wykształcenie (0 wyższe, 1 średnie, 2 podstawowe) są następujące:

	Współczynnik	Błąd Std.	Efekt cząstkowy	t	Pr(t > t*)
płęć	-.350172	.0486801	-.1362704	-7.19333	0.000
educ1	-.539261	.0838674	-.2106847	-6.42992	0.000
educ2	-1.256692	.0933853	-.4226368	-13.45707	0.000
stała	.657053	.0832691		7.89072	0.000



Rysunek 1: Wrażliwość i specyficznosc

Prognozowane	Zaobserwowane			
		1	0	Razem
	1	705	520	1225
	0	525	1131	1656
Razem	1230	1651	2881	

Tablica 1: Trafienia dla $p^* = 0.5$

- Zinterpretować uzyskane wielkości efektów cząstkowych.
- Zinterpretować częstości uzyskane w tabeli (1) i policzyć na jej podstawie wrażliwość i specyficznosc modelu dla punktu ucięcia $p^* = 0.5$.

3. Na podstawie rysunku (1) określić w przybliżeniu, dla jakiego p^* model będzie w 25% przypadków prawidłowo przewidywał, że osoba o określonych charakterystykach nie posiada pracy.
4. Na podstawie rysunku (1) określić w przybliżeniu, dla jakiego p^* model będzie w 25% przypadków przewidywał, że osoba w rzeczywistości posiadająca pracę jest bezrobotna.

Rozwiązanie:

1. Efekt cząstkowy dla płci oznacza, że kobiety mają o 13.6 punkta procentowego mniejsze prawdopodobieństwo znalezienia pracy niż mężczyźni o tym samym poziomie wykształcenia, efekt cząstkowy dla *educ1* oznacza, że osoby z wykształceniem średnim mają o 21.0 punkta procentowego mniejsze prawdopodobieństwo wykonywania pracy, efekt cząstkowy dla *educ2* oznacza, że osoby z wykształceniem podstawowym mają o 42.2 punkta procentowego mniejsze prawdopodobieństwo wykonywania pracy
2. Wrażliwość to prawdopodobieństwo przewidzenia sukcesu jeśli rzeczywiście mamy sukces. Oszacować te prawdopodobieństwo można na podstawie udziału takich obserwacji w ogóle sukcesów. Dla $p^* = 0.5$ oszacowanie wrażliwości wynosi więc $\frac{705}{1230} = 0.573$. Analogicznie specyficzność to prawdopodobieństwo przewidzenia porażki jeśli zaobserwowaliśmy porażkę. Oszacowanie specyficzności wynosi: $\frac{1131}{1651} = 0.685$.
3. Z rysunku możemy odczytać, że specyficzność (zaobserwowana porażka, przewidywana porażka) jest równa 0.25 dla $p^* \approx 0.28$.
4. Prawdopodobieństwa uznania sukcesu za porażkę można zapisać jako $\Pr(\text{porażka} | \text{sukces})$. Ponieważ wrażliwość jest równa $\Pr(\text{sukces} | \text{sukces})$ więc poszukiwane prawdopodobieństwo jest równe:

$$\Pr(\text{porażka} | \text{sukces}) = 1 - \Pr(\text{sukces} | \text{sukces}) = 1 - \text{wrażliwość}$$

Prawdopodobieństwo to można znaleźć, znajdując takie p^* , dla którego wrażliwość jest równa $1 - 0.25 = 0.75$. Z wykresu odczytujemy, że zachodzi to dla $p^* \approx 0.6$.