Egzamin z ekonometrii 17.06.2008
2 semestr

ZADANIE 1 Celem analizy jest wyznaczenie determinantów bycia bezrobotnym dla osób w wieku 18 - 65 lat (zmienna zależna: bezrobotny: 1 jeżeli osoba jest bezrobotna; 0 w pozostałych przypadkach). W tym celu postanowiono oszacować model logitowy. Jako zmienne oddziałujące na to, czy dana osoba jest bezrobotna uznano: *wiek* - wiek respondenta wyrażony w latach; *wiek*_2 - wiek podniesiony do kwadratu, *ang* - wartość 1 jeżeli, osoba zna biegle w mowie i piśmie język angielski, 0 w pozostałych przypadkach; *niem* - wartość 1 jeżeli, osoba zna biegle w mowie i piśmie język niemiecki, 0 w pozostałych przypadkach; *angXniem* - interakcja między zmienną ang i niem; zmienne zerojedynkowe dotyczące wykształcenia: *edu*_2 - wartość 1, jeżeli najwyższy poziom wykształcenia osiągnięty przez respondenta to wykształcenie zasadnicze lub gimnazjalne, 0 w pozostałych przypadkach; *edu*_3 - wartość 1, jeżeli najwyższy poziom wykształcenia osiągnięty przez respondenta to wykształcenie średnie, 0 w pozostałych przypadkach; *edu*_4 - wartość 1, jeżeli najwyższy poziom wykształcenia osiągnięty przez respondenta to wykształcenie wyższe, 0 w pozostałych przypadkach.
Za poziom bazowy w przypadku wykształcenia przyjęto wykształcenie podstawowe lub niższe.

Tabela 1

Marginal effects after logit

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.	]	X
wiek	.0108864	.00077	14.12	0.000	.009375	.012398	43.966
wiek_2	-.0001646	.00001	-18.21	0.000	-.000182	-.000147	2263.7
ang*	-.0233253	.0057	-4.09	0.000	-.034499	-.012152	.184885
niem*	-.0170236	.00763	-2.23	0.026	-.031985	-.002062	.120108
angXniem*	-.0241698	.01089	-2.22	0.026	-.045516	-.002823	.063833
edu_2*	-.014621	.0053	-2.76	0.006	-.02501	-.004232	.326991
edu_3*	-.0181968	.00538	-3.38	0.001	-.028742	-.007651	.296491
edu_4*	-.0394265	.00498	-7.91	0.000	-.049195	-.029658	.150742

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

Tabela 2

Prognozowane	Zaobserwowane 1	0	Total
1	453	1693	2146
0	314	4950	5264
Total	767	6643	7410

Tabela 3

Variable	model1	model2	model3	model wyĹ›ciowy
wiek	.20580622	.18246283	.19544186	.21522575
wiek_2	-.00314635	-.00285301	-.00300665	-.00325441
ang	-.65241519	-.80421091	-.6578719	-.53167519
niem	-.43072272	-.53014767	-.44393619	-.38179763
angXniem	-.51285321	-.39307318	-.52519984	-.59842639
edukacja	-.53476425		-.31875207	
edu_2	-.29362207			-.30209709
edu_3				-.38419648
edu_4				-1.0504967
_cons	-4.2303768	-4.1367982	-4.2676874	-4.4165494
aic	4409.1431	4426.0936	4413.6288	4389.782

```
log likelihood | -2196.5716   -2207.0468   -2199.8144    -2185.891
----------------------------------------------------------------------
```

1. Oszacowano wielkości efektów cząstkowych (tabela 1). Wyznaczyć i zinterpretować efekt cząstkowy dla wieku (przyjąć, iż średni wiek respondentów w analizowanej próbie wynosi 44 lata). Zinterpretować wielkości efektów cząstkowych dla zmiennych dotyczących znajomości języków obcych oraz wykształcenia.

2. Dla prawdopodobieństwa progowego wynoszącego $p^* = 0,15$ wyznaczono tablicę klasyfikacji (tabela 2). Na podstawie powyższej tabeli oszacować:

   (a) prawdopodobieństwo przewidzenia, iż osoba w rzeczywistości nie będącą bezrobotną zostanie uznana za bezrobotną;

   (b) R^2 liczebnościowe i je zinterpretować.

3. W tabeli 3 znajdują się oszacowania parametrów oraz wielkości funkcji wiarygodności oraz kryterium informacyjnego AIC dla 4 modeli: modelu wyjściowego oraz trzech 3 modeli szczegółowych. Zmienna edukacja przyjmuje wartość 1, jeżeli najwyższy poziom wykształcenia osiągnięty przez respondenta to wykształcenie średnie lub wyższe, 0 w pozostałych przypadkach, natomiast pozostałe zmienne są kodowane jak w poprzednich podpunktach. Postanowiono przetestować hipotezę o tym, iż osoby, których najwyższy poziom wykształcenia to średnie lub wyższe, mają takie samo prawdopodobieństwo bycia bezrobotnym.

   (a) Zapisać postać hipotezy zerowej w postaci ograniczeń nałożonych na parametry „modelu wyjściowego”.

   (b) Który z modeli odpowiada ograniczeniu z hipotezy zerowej?

   (c) Przetestować tę hipotezę za pomocą testu ilorazu wiarogodności na poziomie istotności 0,05. Zinterpretować uzyskany wynik.
   Podpowiedź: Kwantyle rozkładu chi-kwadrat:
   $$\chi^2_{1;0,95} = 3.84;\ \chi^2_{2;0,95} = 5.99;\ \chi^2_{3;0,95} = 7.81$$

   (d) Na który z czterech wyżej wymienionych modeli wskazuje kryterium informacyjne AIC? Odpowiedź proszę uzasadnić.

Rozwiązanie:

1. Zaczynamy od wyznaczenia efektu cząstkowego dla wieku.
$$\frac{\partial F(\beta_1 wiek + \beta_2 wiek_2 + \beta Z)}{\partial wiek} = \underbrace{f(\beta_1 wiek + \beta_2 wiek_2 + \beta Z)}_{X\beta} \times (\beta_1 + 2\beta_2 wiek) =$$
$$= f(X\beta)\beta_1 + 2\beta_2 wiek \times f(X\beta)$$
gdzie wektor Z zawiera wszystkie pozostałe zmienne, poza wiekiem i wiekiem podniesionym do kwadratu, umieszczone w modelu. Efekty cząstkowe są liczone dla wektora średnich:
$$f(\bar{X}\beta)\beta_1 + 2wiek\beta_2 f(\bar{X}\beta) = 0,0108864 - 2 \times 44 \times 0,0001646 \approx -0,0036$$
Wzrost wieku z 44 na 45 lat powoduje spadek prawdopodobieństwa bycia bezrobotnym o $0,36\%$, przy założeniu pozostałych zmiennych na poziomie średniej z próby. Osoby znające biegle w mowie i piśmie język angielski mają o $2,33\%$ mniejsze prawdopodobieństwo bycia bezrobotnym, niż osoby nie znające języka angielskiego i niemieckiego. Osoby znające biegle w mowie i piśmie język niemiecki mają o $1,7\%$ mniejsze prawdopodobieństwo bycia bezrobotnym, niż osoby nie znające języka angielskiego i niemieckiego. Osoby znające biegle w mowie i piśmie język angielski i niemiecki mają o $6,45\%$ $(2,33\% + 1,7\% + 2,42\%)$ mniejsze prawdopodobieństwo bycia bezrobotnym, niż osoby nie znające języka angielskiego i niemieckiego. Równoważnie, osoby znające biegle w mowie i piśmie język angielski i niemiecki mają o $4,75\%$ $(2,33\% + 2,42\%)$ mniejsze prawdopodobieństwo bycia bezrobotnym w porównaniu z osobami, które znają biegle w mowie i piśmie tylko język niemiecki. Osoby o wykształceniu zawodowym lub gimnazjalnym mają o $1,46\%$ mniejsze prawdopodobieństwo bycia bezrobotnym w porównaniu z osobami o wykształceniu podstawowym lub niższym. Osoby o wykształceniu średnim mają o $1,82\%$ mniejsze prawdopodobieństwo bycia bezrobotnym w porównaniu z osobami o wykształceniu podstawowym lub niższym. Osoby o wykształceniu wyższym lub mają o $3,94\%$ mniejsze prawdopodobieństwo bycia bezrobotnym w porównaniu z osobami o wykształceniu podstawowym lub niższym. Powyższe interpretacje efektów cząstkowych dla poszczególnych zmiennych przy założeniu, że pozostałe zmienne przyjmują wartości na poziomie średniej z próby.

2.
$$P(\hat{y}=1|y=0)=\frac{P(\hat{y}=1\wedge y=0)}{P(y=0)}=\frac{1693/7410}{6643/7410}\approx 0,25$$
$$R^2_{liczebnościowe}=\frac{453+4950}{7410}\approx 73\%$$

Model prawidłowo przewiduje zdarzenia (bycie bezrobotnym lub nie) w 73%.

(a) Hipotezę o tym, iż osoby, których najwyższy poziom wykształcenia to przynajmniej wykształcenie średnie (czyli średnie lub wyższe), mają takie samo prawdopodobieństwo bycia bezrobotnym możemy zapisać w następujący sposób:
$$H_0:\beta_{edu\_3}=\beta_{edu\_4}.$$

(b) Hipotezie zerowej odpowiada model 1. W modelu tym umieszczono zmienną *edukacja*, która przyjmuje wartość 1 dla osób mających przynajmniej wykształcenie średnie. Umieszczając tę zmienną w modelu, zamiast zmiennych *edu*_3 i *edu*_4, zakładamy te samo prawdopodobieństwo bycia bezrobotnym dla osób o wykształceniu średnim i wyższym.

(c) Przeprowadzamy test ilorazu wiarogodności. Model bez ograniczeń: $\ln L(\hat{\theta})=-2185,891$. Model z ograniczeniami: $\ln L(\hat{\theta}_R)=-2196.5716$. Przeprowadzamy test ilorazu wiarogodności. Statystyka testowa:
$$LR=2\ln\left[\frac{L(\hat{\theta})}{L(\hat{\theta}_R)}\right]=2\ln L(\hat{\theta})-2\ln L(\hat{\theta}_R)=2\times(-2185,891+2196,5716)\approx 21,36$$
Wartość krytyczna: $\chi_{2*}=\chi^2_{0,95;1}=3,84$ (jeden stopień swobody, gdyż jest jedno ograniczenie). Hipotezę zerową odrzucamy, ponieważ $LR>\chi_{2*}$. Na podstawie wyniku testu stwierdzamy, że osoby o wykształceniu średnim i wyższym nie mają takiego samego prawdopodobieństwa bycia bezrobotnym.

(d) Kryterium informacyjne *AIC* wskazuje na „model wyjściowy”, gdyż właśnie dlatego modelu przyjmuje ono najmniejszą wartość.

ZADANIE 2 Rozpatrujemy liniowy model efektów nieobserwowalnych postaci:
$$y_{it}=\boldsymbol{x}_{it}\boldsymbol{\beta}+c_i+\varepsilon_{it}$$

1. Pokazać, że efekt indywidualny można wyeliminować z modelu stosując pierwsze różnice po czasie do lewej i prawej strony modelu.
2. Jak w modelu przekształconym tak jak w punkcie (1) można uzyskać estymator wektora parametrów $\boldsymbol{\beta}$? Przy jakich założeniach estymator ten będzie nieobciążony?
3. Pokazać, że estymator z punktu (2) dla przypadku, gdy $T=2$ (panelu z dwoma falami) jest algebraicznie równoważny estymatorowi efektów stałych.
4. Jakie założenia powinny być spełnione, aby estymator z punktu (2) był najlepszym liniowym i nieobciążonym estymatorem $\boldsymbol{\beta}$?

Rozwiązanie:

1.
$$\Delta y_{it}=\Delta\boldsymbol{x}_{it}\boldsymbol{\beta}+\Delta\varepsilon_{it}$$
2. Przekształcony model jest typowym modelem liniowym - można zastosować MNK. Warunkiem nieobciążoności jest nielosowość $\Delta\boldsymbol{x}_{it}$ bądź warunek, że $\mathrm{E}\left(\Delta\varepsilon_{it}|\,\Delta\boldsymbol{x}_{it}\right)=0$
3. Estymator efektów stałych dla panelu z dwoma falami liczymy na podstawie równania przekształconego za pomocą transformacji efektów stałych
$$y_{it}-\overline{y}_{i\cdot}=\left(\boldsymbol{x}_{it}-\overline{\boldsymbol{x}}_{i\cdot}\right)\boldsymbol{\beta}+\left(\varepsilon_{it}-\overline{\varepsilon}_{i\cdot}\right)$$

dla przypadku, gdy $T = 2$ średnie $\overline{y}_{i\cdot} = \frac{y_{i1}+y_{i2}}{2}$, $\overline{\boldsymbol{x}}_{i\cdot} = \frac{\boldsymbol{x}_{i1}+\boldsymbol{x}_{i2}}{2}$, $\overline{\varepsilon}_{i\cdot} = \frac{\varepsilon_{i1}+\varepsilon_{i2}}{2}$. W konsekwencji dla $T = 1$

$$\frac{1}{2}(y_{i2} - y_{i1}) = \frac{1}{2}(\boldsymbol{x}_{i2} - \boldsymbol{x}_{i1})\boldsymbol{\beta} + \frac{1}{2}(\varepsilon_{i2} - \varepsilon_{i1})$$
$$\Delta y_{i2} = \Delta \boldsymbol{x}_{i2}\boldsymbol{\beta} + \Delta\varepsilon_{i2}$$

a dla $T = 2$

$$\frac{1}{2}(y_{i1} - y_{i2}) = \frac{1}{2}(\boldsymbol{x}_{i1} - \boldsymbol{x}_{2})\boldsymbol{\beta} + \frac{1}{2}(\varepsilon_{i1} - \varepsilon_{i2})$$
$$\Delta y_{i2} = \Delta \boldsymbol{x}_{i2}\boldsymbol{\beta} + \Delta\varepsilon_{i2}$$

Rzeczywiście przekształcenie efektów stałych jest więc w tym przypadku równoważne różnicowaniu

4. Estymator jest najlepszym liniowymi i niobciążonym estymatorem parametru $\boldsymbol{\beta}$ jeśli w odniesieniu do modelu przekształconego spełnione są założenia twierdzenia Gaussa-Markowa (a więc spełnione będą założnia $KMRL$). Zatem prawdą powinno być, że:

   - $\Delta y_{it} = \Delta \boldsymbol{x}_{it}\boldsymbol{\beta} + \Delta\varepsilon_{it}$
   - $\mathrm{E}\left(\Delta\varepsilon_{it} | \Delta \boldsymbol{x}_{it}\right) = 0$ (lub $\Delta \boldsymbol{x}_{it}$ nielosowe)
   - $\mathrm{Var}\left(\Delta\varepsilon_{it} | \Delta \boldsymbol{x}_{it}\right) = \sigma$ (bezwarunkowe wartości oczekiwane, jeśli $\Delta \boldsymbol{x}_{it}$ nielosowe)
   - $\mathrm{Cov}\left(\Delta\varepsilon_{it}, \Delta\varepsilon_{js} | \Delta \boldsymbol{x}_{it}\right) = 0$ (bezwarunkowe wartości oczekiwane, jeśli $\Delta \boldsymbol{x}_{it}$ nielosowe) dla wszystkich i, t, j, s poza przypadkiem, gdy $i = j$ i $t = s$

ZADANIE 3 Badacz postanowił modelować dzienną stopę zwrotu z indeksu WIG, zdefiniowaną jako różnica logarytmów i oznaczoną jako return.

Tabela 1

```
Dickey-Fuller test for unit root                       Number of obs   =      3267
                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)            -42.544            -3.430            -2.860            -2.570
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t)
```

Tabela 2

```
KPSS test for return
Critical values for H0: return is trend stationary
10%: 0.119  5% : 0.146  2.5%: 0.176  1% : 0.216
Lag order    Test statistic
    0           .363
    1           .285
    2           .264
    3           .254
    4           .245
    5           .238
    6           .233
    7           .229
    8           .224
    9            .22
   10           .216
```

Tabela 3

Model	nobs	ll(null)	ll(model)	df	AIC	BIC
arma22	3268	.	8017.331	6	-16022.66	-15986.11
arma21	3268	.	8017.052	5	-16024.1	-15993.64
arma11	3268	.	8016.992	4	-16025.98	-16001.62
arma01	3268	.	8016.989	3	-16027.98	-16009.7
arma00	3268	.	7860.439	2	-15716.88	-15704.69

Tabela 4

return	Coef.	OPG Std. Err
return		
L1	-.0041296	.0284394
_cons	.0011257	.0004792
ARMA		
ma		
L1	.3212883	.0287775
/sigma	.020814	.0001395

Tabela 5

zmienna	Coef.	OPG Std. Err.	z	P>|z|	[95% Conf.	Interval]
ARCH						
arch						
L1	.3224945	.0255351	12.63	0.000	.2724467	.3725423
L2	.1792855	.0230251	7.79	0.000	.1341572	.2244138
L3	.2349035	.0241403	9.73	0.000	.1875893	.2822177
_cons	.0001132	4.58e-06	24.69	0.000	.0001042	.0001222

Uwaga: *W poniższym zadaniu badacz nie zawsze musi wykonywać operacje logicznie – tzn. może z testów wyjść mu model ARMA(10,10), a on będzie dalej badał model ARMA(13,7).*
W całym zadaniu przyjąć 5% przedział ufności.

1. Na początku badacz postanowił sprawdzić, czy badana zmienna jest stacjonarna. W tym celu przeprowadził test ADF i KPSS. Wyniki znajdują się w tabelach 1 i 2. Zinterpretować wyniki tych testów?

2. Niezależnie od wyniku testów, badacz przyjął, że zmienna jest stacjonarna. Następnie badacz postanowił wybrać model ARMA metodą od ogólnego do szczegółowego, zaczynając od modelu $ARMA(2,2)$. Następnie testował modele $ARMA(2,1)$, $ARMA(1,1)$, $ARMA(0,1)$, a na koniec $ARMA(0,0)$. W tabeli 1 znajduje się zestawienie wyników.

   (a) Jaki model powinien wybrać badacz kierując się metodologią od ogólnego do szczegółowego?
   Podpowiedź: Wartości krytyczne $\chi_1 = 3.84, \chi_2 = 5.99, \chi_3 = 7.81, \chi_4 = 9.48, \chi_5 = 11.07, \chi_6 = 12.59$

   (b) Jaki model powinien wybrać kierując się kryteriami informacyjnymi? Wybór uzasadnij.

3. Badacz postanowił zająć się modelem $ARMA(1,1)$ i uzyskał oszacowania parametrów znajdujące się w tabeli 4.

(a) Policzyć prognozę na jeden okres wprzód, wiedząc, że ostatnia badana wartość wyniosła $return = 0.0144648$, a ostatnie reszty 0.01282963. Zapisać równanie model, z którego skorzystano.

(b) Policzyć rozwiązanie długookresowe (o ile istnieje).

Wiedząc, że dane finansowe często charakteryzują się warunkową heteroskedastycznością, badacz postanowił sprawdzić, czy w jego modelu ona występuje.

4. Jaką zmienną badacz powinien poddać badaniu, żeby stwierdzić, że w jego modelu występuje warunkowa heteroskedastyczność?

5. Badacz przeprowadził na tej zmiennej model ARCH(3) i otrzymał wyniki wyniki znajdujące się w tabeli 5.

   (a) Policzyć jednookresową prognozę wariancji dla tego modelu wiedząc, że ostatnio obserwowane wartości wariancji były następujące: $\varepsilon^2_T = 0.0001646, \varepsilon^2_{T-1} = 0.00000257, \varepsilon^2_{T-2} = 0.00000106$.

   (b) Policzyć wariancję bezwarunkową w tym modelu.

Rozwiązanie:

1. Na podstawie testu ADF odrzucamy H0, która mówi o niestacjonarności zmiennej. Wobec tego przyjmujemy, że zmienna jest stacjonarna.W teście KPSS sytuacja jest odwrotna H0, mówi o stacjonarności zmiennej. Tu jednak również odrzucamy H0, na zadanym poziomie istotności. Zatem KPSS wskazuje na niestacjonarność zmiennej

2.

   (a) Obliczając różnice logarytmów wiarygodności
   ARMA22 vs ARMA21 LR=0.56<X(1)=3.84
   ARMA22 vs ARMA11 LR=0.68<X(2)=5.99
   ARMA22 vs ARMA01 LR=0.69<X(3)=7.81
   ARMA22 vs ARMA00 LR=313.79>X(4)=9.48
   Badacz powinien wybrać model ARMA(0,1)

   (b) Kierując się kryteriami informacyjnymi badacz powinien wybrać model o najmniejszej wartości kryteriów informacyjnych. Zarówno dla AIC i SBC jest to model ARMA(0,1).

3. Prognoza wynosi

   (a)
   $$y_{T+1} = 0.0011257 - 0.0041296 \times 0.0144648 + 0.3212883 \times 0.01282963 = 0.00518798$$

   (b) Rozwiązanie długookresowe rozwiązujemy przy założeniu, że $\mathrm{E}(y_t) = \mathrm{E}(y_{t+1})$, i $\mathrm{E}(\varepsilon) = 0$,
   $$E(y_t) = \frac{0.0011257}{1 + 0.0041296} = 0.00112107$$

4. Badacz powinien zbadać kwadraty reszt z modelu ARMA.

5.

   (a)
   $$\begin{aligned}\mathrm{Var}(y_{T+1}) &= 0.0001132 + 0.3224945 \times 0.0001646 \\ &+ 0.1792855 \times 0.00000257 \\ &+ 0.2349035 \times 0.00000106 \\ &= 0.00016699\end{aligned}$$

   UWAGA: w orginalnym tekście zadania było: $\sigma^2_T = 0.0001646, \sigma^2_{T-1} = 0.00000257, \sigma^2_{T-2} = 0.00000106$. Ponieważ zapis ten jest niepoprawny, więc podpunkt ten został potraktowany jako fakultatywny.

   (b) Wyprowadzając wzór na wariancję bezwarunkową otrzymamy, że
   $$\sigma^2 = \frac{\mu}{1 - \alpha_1 - \alpha_2 - \alpha_3} = \frac{0.0001132}{1 - 0.3224945 - 0.1792855 - 0.2349035} = 0.0004299$$