Egzamin z ekonometrii IiE 17.06.2010

1. Wyjaśnić, co to znaczy, że model A obejmuje model B. W jaki sposób testujemy obejmowanie?
2. Jakie trzy testy stosuje się do testowania hipotez parametrycznych postaci $\boldsymbol{h}(\boldsymbol{\theta}) = 0$ w kontekście estymacji MNW? Porównać wady i zalety tych testów.
3. Opisać dwustopniową procedurę estymacji modelu Heckmana.
4. Udowodnić, że estymator MZI jest estymatorem zgodnym.

ZADANIE 1 Na podstawie danych zebranych w ramach "Diagnozy Społecznej" z roku 2007 oszacowano za pomocą modelu probitowego model, w którym zmienną objaśnianą jest fakt korzystania z internetu (0-nie korzysta, 1 korzysta). Zmiennymi objaśniającymi są wiek, wiek^2, lata nauki, (lata_nauki)^2, interakcja między wiekiem i latami nauki oraz płeć. Wielkości statystyk opisowych oraz oszacowania parametrów znajdują się poniżej.

```
Probit regression                                 Number of obs   =      15141
                                                  LR chi2(6)      =    5211.40
                                                  Prob > chi2     =     0.0000
Log likelihood = -6733.6463                       Pseudo R2       =     0.2790
```

internet	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
wiek	-.0288243	.0049728	-5.80	0.000	-.0385708	-.0190779
wiek2	-.0001942	.0000569	-3.41	0.001	-.0003057	-.0000826
kobieta	.0161105	.0249495	0.65	0.518	-.0327896	.0650106
lata_nauki	.5789232	.0294019	19.69	0.000	.5212966	.6365499
lata_nauki2	-.015201	.0011747	-12.94	0.000	-.0175034	-.0128986
wiek*lata_nauki	.0011987	.0003316	3.62	0.000	.0005489	.0018486
_cons	-4.238429	.1896739	-22.35	0.000	-4.610183	-3.866675

```
link test statistics:       N(0,1)=1.13   [0.257]
Count R2:                   0.772
Adj Count R2:               0.258
McKelvey and Zavoina's R2:  0.641
```

Variable	Obs	Mean	Std. Dev.	Min	Max
wiek	15141	43.03778	19.38671	1	99
kobieta	15141	.5298197	.4991265	0	1
lata_nauki	15141	11.27937	3.408129	0	30

1. Podać założenia modelu probitowego.
2. Zinterpretować znak współczynnika przy zmiennej zerojedynkowej *kobieta* oraz sprawdzić, czy zmienna ta jest istotna statystycznie.
3. Zinterpretować wynik testu typu związku (link test).
4. Policzyć efekt cząstkowy dla wieku dla średnich poziomów zmiennych objaśniających w próbie.

   Podpowiedź: wartość funkcji gęstości rozkładu normalnego policzonego dla $\overline{\boldsymbol{x}}\widetilde{\boldsymbol{\beta}}$ wynosi 0.32.
5. Zinterpretować wielkość $R^2_{\text{liczebnościowego}}$, skorygowanego $R^2_{\text{liczebnościowego}}$ oraz $R^2_{\text{McKelveya i Zavoiny}}$.

Rozwiązanie:

1. Prawdopodobieństwo zaobserwowania y_i jest w modelu probitowym dane wzorem:

$$\Pr\left(y_i | \boldsymbol{x}_i\right) = \begin{cases} 1 - \Phi\left(\boldsymbol{x}_i\right) & \text{dla} \quad y_i = 0 \\ \Phi\left(\boldsymbol{x}_i\right) & \text{dla} \quad y_i = 1 \end{cases}$$

gdzie $\Phi\left(\right)$ jest dystrybuantą rozkładu normalnego. Alternatywnie zmienna ukryta jest dana wzorem

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

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

zmienna obserwowalna

$$y_i = \begin{cases} 0 & \text{dla} \quad y_i^* \leq 0 \\ 1 & \text{dla} \quad y_i^* > 0 \end{cases}.$$

2. Dodatni współczynnik przy zmiennej kobieta oznacza, że ceteris paribus, kobiety z wy ższym prawdopodbieństwem korzystają z internetu niż mężczyźni. Zmienna kobieta jest jednak nieistotna statystycznie (wartość p 0.58<0.05)

3. Test typu związku wskazuje, że nie można odrzucić hipotezy o prawidłowości formy funkcyjnej przyjętaj dla modelu

4. W modelu probitowym efekt cząstkowy jest równy

$$\frac{\partial \Phi\left(\boldsymbol{x}\boldsymbol{\beta}\right)}{\partial x_k} = \phi\left(\boldsymbol{x}\boldsymbol{\beta}\right) \frac{\partial \boldsymbol{x}\boldsymbol{\beta}}{\partial x_k}$$

przy czym efekt ten jest liczony dla średnich w próbie. Dla $x_k = wiek$

$$\begin{aligned} \frac{\partial \boldsymbol{x}\boldsymbol{\beta}}{\partial x_k} &= \frac{\partial\left(\beta_0 + \beta_1 \text{wiek} + \beta_2 \text{wiek}^2 + \beta_3 \text{kobieta} + \beta_4 \text{lata\_nauki} + \beta_5 \text{lata\_nauki}^2 + \beta_6 \text{wiek} \times \text{lata\_nauki}\right)}{\partial \text{wiek}} \\ &= \beta_1 + 2\beta_2 \text{wiek} + \beta_6 \text{lata\_nauki} \end{aligned}$$

W rezultacie efekt cząstkowy dla wieku policzony dla średnich wartości zmiennych objaśniających w próbie jest równy

$$\begin{aligned} \phi\left(\overline{\boldsymbol{x}}\boldsymbol{\beta}\right) \left.\frac{\partial \boldsymbol{x}\boldsymbol{\beta}}{\partial x_k}\right|_{\boldsymbol{x}=\overline{\boldsymbol{x}}} &= 0.32 * (-.0288243 - 2 * 0.0001942 * 43.03778 + .0011987 * 11.27937) \\ &= -0.01\,024\,6 \end{aligned}$$

5. $R^2_{\text{liczebnościowe}}$: 77% procent obserwacji dla zmiennej zależnej zostało prawidłowo przewidzianych przez model

   skorygowane $R^2_{\text{liczebnościowe}}$: 26% procent obserwacji zostało prawdiłowo przewidzianych przez model za pomocą zmienności zmiennych niezależnych

   $R^2_{\text{McKelveya i Zavoiny}}$: 64% zmienności zmiennej ukrytej zostałoby wyjaśnionych przez model, gdyby zmienna ta była obserwowalna

ZADANIE 2 Niech y_i oraz d_i będą zmiennymi zerojedynkowymi. Wartości tych zmiennych zebrane są w tablicy częstości empirycznych:

		d	
		1	0
y	1	n_1	n_2
	0	n_3	n_4

gdzie n_1 oznacza liczbę obserwacji w próbce, dla których zmienne y_i oraz d_i przyjmują wartość 1 itd. Niech $F(\cdot)$ oznacza dystrybuantę pewnego rozkładu ciągłego i symetrycznego (czyli $F(-x) = 1 - F(x)$). Chcemy oszacować następujący model (model bez stałej ze zmienną zerojedynkową jako zmienną objaśniającą):

$$\Pr\left(Y_i = y_i | d_i\right) = \begin{cases} 1 - F\left(\alpha d_i\right) & \text{dla} \quad y_i = 0 \\ F\left(\alpha d_i\right) & \text{dla} \quad y_i = 1 \end{cases}$$

1. Pokazać, iż logarytm funkcji wiarogodności dla tego modelu możemy zapisać w następujący sposób:

$$\ell(\alpha) = n_1 \ln[F(\alpha)] + n_3 \ln[1 - F(\alpha)] + (n_2 + n_4)\ln(0.5)$$

2. Pokazać, że estymator Metody Największej Wiarogodności w tym modelu wyraża się wzorem:

$$\alpha = F^{-1}\left(\frac{n_1}{n_1 + n_3}\right)$$

gdzie $F^{-1}(\cdot)$ oznacza odwrotność dystrybuanty $F(\cdot)$

Rozwiązanie:

1. Funkcję $\Pr(Y_i = y_i | d_i)$ można zapisać następująco:

$$\Pr(Y_i = y_i | d_i) = [F(\alpha d_i)]^{y_i} [1 - F(\alpha d_i)]^{y_i - 1}$$

Wynika z tego, że funkcja wiarogodności ma postać

$$L(\alpha) = \prod_{i=1}^{n} [F(\alpha d_i)]^{y_i} [F(-\alpha d_i)]^{y_i - 1}$$

a logarytm funkcji wiarygodności

$$\ell(\alpha) = \sum_{i=1}^{n} \{y_i \ln[F(\alpha d_i)] + (y_i - 1)\ln[1 - F(\alpha d_i)]\}$$

Zauważmy, że dla $d_i = 0$, $F(\alpha d_i) = 0.5$ a dla $d_i = 1$, $F(\alpha d_i) = F(\alpha)$ wynika z tego, że

$$\ell(\alpha) = n_1 \ln[F(\alpha)] + n_3 \ln[1 - F(\alpha)] + (n_2 + n_4)\ln(0.5)$$

ponieważ dla n_1 obserwacji mamy $y_i = 1$, $d_i = 1$; dla n_3 obserwacji mamy $y_i = 0$, $d_i = 1$ i dla $(n_2 + n_4)$ obserwacji mamy $d_i = 0$.

2. Warunki pierwszego rzędu mają więc postać

$$\frac{\partial \ell(\alpha)}{\partial \alpha} = \frac{n_1}{F(\alpha)} - \frac{n_3}{1 - F(\alpha)} = 0$$

Rozwiązując dla $F(\alpha)$ uzyskujemy:

$$F(\alpha) = \frac{n_1}{n_1 + n_3}$$

a więc istotnie $\widehat{\alpha} = F^{-1}\left(\frac{n_1}{n_1+n_3}\right)$

ZADANIE 3 Mamy następujący model wielorównaniowy

$$C_t = \alpha_1 + \alpha_2 Y_t + u_t$$
$$Y_t = C_t + I_t + G_t$$

przy czym zakładamy, że C_t (konsumpcja) i Y_t (dochód narodowy) są endogeniczne, a I_t (inwestycje) i G_t (wydatki rządowe) są egzogeniczne.

1. Zbadać, czy równanie konsumpcji jest zidentyfikowane.

2. Wyjaśnić, z jakiego powodu w modelu tym występuje sprzężenie zwrotne i dlaczego powoduje ono, że estymator MNK nie jest zgodny.

3. Jak nazywamy równanie o postaci takiej jak równanie na Y_t w tym modelu? Czym wyróżniają się równania tego typu?

4. Czy estymator wektora parametrów $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)'$ postaci

$$\widehat{\boldsymbol{\alpha}} = \left(\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\boldsymbol{Z}'\boldsymbol{y}$$

gdzie $x_i = (1, y_i)$ a $z_i = (1, I_t)$ jest estymatorem zgodnym? Przy jakich założeniach? Odpowiedź uzasadnij.

5. Czy można w jakiś sposób wykorzystać jednocześnie I_t i G_t w estymacji równania konsumpcji?

Rozwiązanie:

1.

$$\begin{array}{ll} \textit{zmienne egzogeniczne} & 1, I_t, G_t \\ \textit{zmienne endogeniczne} & C_t, Y_t \end{array}$$

$$K = 3 \quad G_1 = 2 \quad K_1 = 1 \quad \text{zidentyfikowane } 3 = K \geq G_1 + K_1 - 1 = 2$$

Identyfikacji drugiego równania nie badamy bo nie ma w nim parametrów do oszacowania

2. C_t wpływa na Y_t a Y_t wpływa na C_t, mamy więc w modelu sprzężenie zwrotne. Ponieważ wpływ C_t na Y_t oznacza, że u_t wpływa na Y_t więc Y_t i u_t są skorelowane a tym samym w modelu występuje problem równoczesności, który powoduje, że estymator MNK jest obciążony i jest niezgodny.

3. Równanie na Y_t jest tożsamością wynikającą z rachunkowości dochodu narodowego. Tożsamości odróżniają się od innych równań tym, że nie występuje w nich błąd losowy ani parametry do oszacowania.

4. Estymator $\widehat{\boldsymbol{\alpha}}$ jest estymatorem MZI, będzie on zgodny jeśli zmienna I_t jest rzeczywiście egzogeniczna a więc nieskorelowana z zaburzeniem losowym u_t.

5. Można zastosować obie te zmienne jako zmienne instrumentalne w uogólnionym estymatorze zmiennych instrumentalnych postaci: $\boldsymbol{b}_{MZI} = \left(\widehat{\boldsymbol{X}}'\widehat{\boldsymbol{X}}\right)^{-1}\widehat{\boldsymbol{X}}'\boldsymbol{y}$, gdzie $\widehat{\boldsymbol{X}} = \boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{X}$ i $\boldsymbol{Z}$ jest macierzą obserwacji dla instrumentów zawierającą stałą, I_t, G_t