Egzamin z ekonometrii dla IiE 11.06.2007
II semestr

Pytania teoretyczne

1. Jakie trzy testy stosujemy do testowania hipotez parametrycznych postaci $\boldsymbol{h}(\boldsymbol{\theta}) = 0$ w kontekście estymacji MNW? Porównać wady i zalety tych testów (po jednej wadzie i zalecie dla każdego testu).
2. Pokazać, że estymator MNK jest estymatorem M, oraz, że dla $\mathrm{E}(\varepsilon_i | \boldsymbol{x}_i) = 0$ spełnia on założenia konieczne do zgodności estymatora M.
3. Kiedy może pojawić się problem nielosowej selekcji próby i jaki model można w tym przypadku zastosować?
4. Czym różni się forma strukturalna od formy zredukowanej?

ZADANIE 1

1. Pokazać, że podobnie jak w MNK, także w przypadku modelu logitowego oszacowanym MNW

$$\sum_{i=1}^{N} e_i \boldsymbol{x}_i = 0,$$

gdzie $\widehat{p}_i = F\left(\boldsymbol{x}_i \widehat{\boldsymbol{\beta}}\right)$, a $\widehat{e}_i = y_i - \widehat{p}_i$.

2. Pokazać, że w oszacowanym MNW modelu logitowym ze stałą udział sukcesów w ogóle obserwacji równy jest średniemu oszacowanemu prawdopodobieństwu sukcesu:

$$\sum_{i=1}^{N} e_i = 0$$

3. Pokazać, że w oszacowanym MNW modelu logitowym estymatorem macierzy wariancji $\widehat{\boldsymbol{\beta}}$ może być

$$\widehat{\boldsymbol{\Sigma}} = \left(\sum_{i=1}^{N} \widehat{\sigma}_i^2 \boldsymbol{x}_i \boldsymbol{x}_i' \right)^{-1}$$

gdzie σ_i^2 jest równa:

$$\sigma_i^2 = F\left(\boldsymbol{x}_i \widehat{\boldsymbol{\beta}}\right) \left[1 - F\left(\boldsymbol{x}_i \widehat{\boldsymbol{\beta}}\right)\right]$$

Podpowiedź: Skorzystaj ze wzoru, który mówi, że macierz informacyjna Fishera $\boldsymbol{I}(\boldsymbol{\beta}) = \mathrm{Var}\left(\frac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \middle| \boldsymbol{X} \right)$

Rozwiązanie:

1. W przypadku modelu logitowego logarytm funkcji wiarogodności ma postać:

$$\begin{aligned}
\ell(\boldsymbol{\beta}) &= \sum_{i=1}^{N} \left[y_i \ln \left(\frac{e^{\boldsymbol{x}_i \boldsymbol{\beta}}}{1 + e^{\boldsymbol{x}_i \boldsymbol{\beta}}} \right) + (1 - y_i) \ln \left(\frac{1}{1 + e^{\boldsymbol{x}_i \boldsymbol{\beta}}} \right) \right] \\
&= \sum_{i=1}^{N} y_i \boldsymbol{x}_i \boldsymbol{\beta} - \sum_{i=1}^{N} \ln \left(1 + e^{\boldsymbol{x}_i \boldsymbol{\beta}} \right)
\end{aligned}$$

Wektor pierwszych pochodnych funkcji wiarogodności będzie więc dany wzorem

$$\begin{aligned}
\left.\frac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}} &= \sum_{i=1}^{N} y_i \boldsymbol{x}_i - \sum_{i=1}^{N} \frac{e^{\boldsymbol{x}_i \widehat{\boldsymbol{\beta}}}}{1 + e^{\boldsymbol{x}_i \widehat{\boldsymbol{\beta}}}} \boldsymbol{x}_i = \sum_{i=1}^{N} y_i \boldsymbol{x}_i - \sum_{i=1}^{N} \widehat{p}_i \boldsymbol{x}_i \\
&= \sum_{i=1}^{N} (y_i - \widehat{p}_i) \boldsymbol{x}_i = \sum_{i=1}^{N} e_i \boldsymbol{x}_i = 0
\end{aligned}$$

2. W przypadku modelu ze stałą $x_{1i} = 1$ a co implikuje, że $\sum_{i=1}^{N} e_i = 0$, a to z kolei implikuje, że

$$\frac{1}{N}\sum_{i=1}^{N} y_i = \frac{1}{N}\sum_{i=1}^{N}\widehat{p}_i$$

3. Korzystając ze wzoru wyprowadzonego w poprzednim punkcie:

$$\boldsymbol{I}(\boldsymbol{\beta}) = \mathrm{Var}\left(\left.\frac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right| \boldsymbol{X}\right) = \mathrm{Var}\left(\left.\sum_{i=1}^{N} y_i \boldsymbol{x}_i - \sum_{i=1}^{N} p_i \boldsymbol{x}_i \right| \boldsymbol{X}\right)$$
$$= \sum_{i=1}^{N} \mathrm{Var}\left(y_i | \boldsymbol{x}_i\right) \boldsymbol{x}_i \boldsymbol{x}_i' = \sum_{i=1}^{N} \sigma_i^2 \boldsymbol{x}_i \boldsymbol{x}_i'$$

ponieważ poszczególne obserwacje są niezależne a $\mathrm{Var}\left(y_i | \boldsymbol{x}_i\right) = F\left(\boldsymbol{x}_i\boldsymbol{\beta}\right)\left[1 - F\left(\boldsymbol{x}_i\boldsymbol{\beta}\right)\right]$. Macierz wariancji $\widehat{\boldsymbol{\beta}}$ można oszacować przy użyciu odwrotności oszacowanej macierzy informacji Fishera

ZADANIE 2 Przeprowadzono regresję MNK na próbie panelowej skonstruowanej na próbie $BAEL$ (fala z roku 2000.2), która objaśnia liczbę przepracownych godzin w tygodniu ($HOURTOT$) za pomocą wieku (AGE), wieku do kwadratu ($AGE2$) oraz płci (SEX - 1 mężczyzna, 2 kobieta) i uzyskano następujące oszacowania:

```
      Source |       SS       df       MS              Number of obs =   20018
-------------+------------------------------           F(  3, 20014) =  642.80
       Model |  251123.954     3  83707.9847           Prob > F      =  0.0000
    Residual |  2606308.99 20014  130.224292           R-squared     =  0.0879
-------------+------------------------------           Adj R-squared =  0.0877
       Total |  2857432.94 20017  142.750309           Root MSE      =  11.412

------------------------------------------------------------------------------
    HOURSTOT |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         AGE |   1.064796   .0358711    29.68   0.000     .9944853    1.135106
        AGE2 |   -.013275    .000422   -31.46   0.000    -.0141021   -.0124479
     _ISEX_2 |  -4.912066   .1621085   -30.30   0.000    -5.229812    -4.59432
       _cons |   23.49594   .7304442    32.17   0.000     22.06421    24.92767
------------------------------------------------------------------------------
```

Poziom istotności $\alpha = 0.05$

1. Zinterpretować oszacowania parametrów (w przypadku wieku zintepretuj wielkość parametrów dla osoby mającej 20 lat i 50 lat)

2. Przeprowadzono powtórnie regresję MNK, przy czym użyto odpornej macierzy wariancji i kowarinacji i uzyskano następujące wyniki:

```
Regression with robust standard errors                 Number of obs =   20018
                                                       F(  3,  6401) =  218.81
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.0879
Number of clusters (ID) = 6402                         Root MSE      =  11.412

------------------------------------------------------------------------------
             |               Robust
    HOURSTOT |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         AGE |   1.064796   .0631098    16.87   0.000     .9410792    1.188512
        AGE2 |   -.013275   .0007669   -17.31   0.000    -.0147784   -.0117716
     _ISEX_2 |  -4.912066   .2596481   -18.92   0.000    -5.421063   -4.403069
       _cons |   23.49594   1.248603    18.82   0.000     21.04826    25.94362
------------------------------------------------------------------------------
```

Skąd wynikają tak duże rozbieżności między oszacowaniami błędów standardowych uzyskanymi za pomocą zwykłej macierzy wariancji i kowariancji i oszacowaniami uzyskanymi za pomocą macierzy odpornej?

3. Oszacowano model powtórnie przy użyciu estymatora efektów losowych i uzyskano następujące wyniki:

```
Random-effects GLS regression                   Number of obs      =     20018
Group variable (i): ID                          Number of groups   =      6402

R-sq:  within  = 0.0002                         Obs per group: min =         1
       between = 0.1108                                        avg =       3.1
       overall = 0.0878                                        max =         4

Random effects u_i ~ Gaussian                   Wald chi2(3)       =    782.88
corr(u_i, X)       = 0 (assumed)                Prob > chi2        =    0.0000

------------------------------------------------------------------------------
    HOURSTOT |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         AGE |   1.070623   .0548659    19.51   0.000     .9630881    1.178159
        AGE2 |   -.013242   .0006507   -20.35   0.000    -.0145174   -.0119666
     _ISEX_2 |  -5.032645   .2602534   -19.34   0.000    -5.542733   -4.522558
       _cons |   22.89486   1.105944    20.70   0.000     20.72725    25.06247
-------------+----------------------------------------------------------------
     sigma_u |  9.5763971
     sigma_e |  6.5749277
         rho |  .67963082   (fraction of variance due to u_i)
------------------------------------------------------------------------------

Breusch and Pagan Lagrangian multiplier test for random effects:

        HOURSTOT[ID,t] = Xb + u[ID] + e[ID,t]

        Estimated results:
                         |       Var     sd = sqrt(Var)
                ---------+-----------------------------
                HOURSTOT |   142.7503       11.94782
                       e |   43.22967       6.574928
                       u |   91.70738       9.576397

        Test:   Var(u) = 0
                             chi2(1) = 10155.65
                         Prob > chi2 =     0.0000
```

Podać liczbę osób przebadanych w ramach panelu, zinterpretować wszystkie R^2, przetestować łączną istotność zmiennych.

4. Zintepretować oszacowania σ_ε^2, σ_u^2 i ρ oraz wynik testu na istnienie efektów losowych

5. Oszacowano model za pomocą estymatora efektów stałych dla tej samej próby i uzyskano następujące oszacowania:

```
Fixed-effects (within) regression               Number of obs      =     20018
Group variable (i): ID                          Number of groups   =      6402

R-sq:  within  = 0.0018                         Obs per group: min =         1
       between = 0.0002                                        avg =       3.1
       overall = 0.0000                                        max =         4

                                                F(2,13614)         =     12.15
```

```
corr(u_i, Xb)  = -0.3536                        Prob > F          =    0.0000

------------------------------------------------------------------------------
    HOURSTOT |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         AGE |   .6787975   .2518456     2.70   0.007     .1851453     1.17245
        AGE2 |  -.0042814   .0029958    -1.43   0.153    -.0101535    .0015907
     _ISEX_2 |  (dropped)
       _cons |   21.02294   5.261653     4.00   0.000     10.70937     31.3365
-------------+----------------------------------------------------------------
     sigma_u |  11.734469
     sigma_e |  6.5749277
         rho |  .76106622   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(6401, 13614) =     7.29         Prob > F = 0.0000
```

Zbadać łączną istotność zmiennych i istotność poszczególnych zmiennych, sprawadzić czy efekty indywidualne są istotne.

6. Wyjaśnić dlaczego w tym przypadku nie uzyskaliśmy oszacowania współczynnika dla zmiennej SEX.

7. Przeprowadzono test Hausmana, dla tych dwóch estymatorów i uzyskano następujący wynik:

```
                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |       fe           re         Difference          S.E.
-------------+----------------------------------------------------------------
         AGE |    .6787975     1.070623       -.3918259         .2457965
        AGE2 |   -.0042814     -.013242        .0089606         .0029242
------------------------------------------------------------------------------
                           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

                  chi2(2) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =       31.74
                Prob>chi2 =      0.0000
```

Zinterpretować wynik tego testu i wyjaśnić jakie są wnioski wynikające z wyniku testu jeśli chodzi o prawidłowy sposób estymacji rozpatrywanego modelu.

Rozwiązanie:

1. Wielkość efektu cząstkowego dla wieku jest równa $\frac{\partial HOURSTOT}{\partial AGE} = \beta_{AGE} + 2AGE\beta_{AGE^2}$. Dla osoby 20 letniej mamy $\frac{\partial HOURSTOT}{\partial AGE} = 1.064 + 2 \times 20 \times (-.013) = .544$. Dla 50 letniej $\frac{\partial HOURSTOT}{\partial AGE} = 1.064 + 2 \times 50 \times (-.013) = -.236$. Oznacza to, że dodatkowy rok życia w przypadku osoby 20 letniej zwiększa długość czasu poświęconego pracy o $.544$ godziny a dla 50 letniej skraca o $.236$ godz. Kobiety pracują w tygodniu o 4.91 godz krócej niż mężczyźni.

2. Różnice te wynikają z tego, że w próbie panelowej z racji na występowanie nieobserwowalnych efektów indywidualnych, (łączne) błędy losowe są skorelowane dla poszczególnych jednostek a tym samym standardowa, uzyskana za pomocą MNK macierz wariancji kowariancji nie jest prawidłowym oszacowaniem macierzy wariancji błędu losowych. Prawidłowe oszacowanie można uzyskać właśnie za pomocą macierzy odpornej.

3. W ramach panelu przebadano 6402, 0.02% zróżnicowania obserwacji dla poszczególnych gospodarst została wyjaśniona przez model, 11.08% zróznicowania między gospodarstwami została wyjaśniona przez model, 8.7% ogólnego zróżnicowania została wyjaśniona przez model. Odrzucamy hipotezę o łącznej nieistotności wszystkich zmiennych na podstawie testu Walda $[782.88, 0.000 < 0.05]$

4. Odchylenie standardowe efektu indywidualnego wynosi 9.57, ochylenie standardowe czystego błędu losowego wynosi 6.57, udział wariancji efektu indywidualnego w łącznym błędzie losowym (współczynnik korelacji dla łącznych błędów losowych dla dwóch różnych obserwcji dla tej samej jednostki) wynosi $.67$. Wielkość statystyki $[10155.65, 0.000 < 0.05]$ pozwala nam odrzucić hipotezę zerową o braku efektów losowych

5. Odrzucamy H_0 o łącznej nieistotności zmiennych objaśnijących $[12.15, 0.000 < 0.05]$. Stała i wiek są istotne $[4.00, 0.000 < 0.05]$, $[2.70, 0.007 < 0.05]$. Odrzucamu H_0 o nieistotności efektów indywidualnych musimy odrzucić $[7.29, 0.000 < 0.05]$

6. Zmienna SEX dla poszczególnych jednostek nie zmienia się w czasie. Za pomocą estymatora efektów stałych nie da się oszacować współczynników dla takich zmiennych.

7. Na postawie testu Hausmana musimy odrzucić H_0 o tym, że efekty indywidualne są nieskorelowane ze zmiennymi objaśniającymi. Tym samym do estymacji tego modelu nie powinno się stosować estymatora efektów losowych, ponieważ nie będzie on zgodny.

ZADANIE 3 Mamy następujący model

$$y_i = (\alpha + \beta x_i)\, \varepsilon_i$$

gdzie ε_i ma rozkład wykładniczy o dystrybuancie $F(\varepsilon_i | x_i) = 1 - \exp(-\varepsilon_i)$ i $\varepsilon_i, \varepsilon_j$ są niezależne dla $i \neq j$.

1. Wyprowadzić wzory analityczne dla estymatorów UMM w tym modelu.

   Podpowiedź: Wartość oczekiwana w rozkładzie wykładniczym o dystrybuancie $F(z) = 1 - \exp(-\beta z) = \frac{1}{\beta}$

2. Opisać, (bez wyprowadzania wzorów analitycznych), w jaki sposób można uzyskać bardziej precyzyjne oszacowania α i β stosując estymator UMM z optymalną macierzą wag.

Rozwiązanie:

1. Wartość oczekiwana $\mathrm{E}(y_i | x_i) = \alpha + \beta x_i$. Tak więc dla $f(\alpha, \beta) = y_i - \alpha - \beta x_i$ mamy, że $\mathrm{E}[f(\alpha, \beta) | x_i] = 0$. Jeśli za instrumenty przyjmiemy stałą i x_i to uzyskamy bezwarunkowe ograniczenia na momenty

$$\begin{aligned}\mathrm{E}[m_{1i}(\alpha, \beta)] &= \mathrm{E}[f_i(\alpha, \beta)] = \mathrm{E}(y_i - \alpha - \beta x_i) = 0\\ \mathrm{E}[m_{2i}(\alpha, \beta)] &= \mathrm{E}[f_i(\alpha, \beta)\, x_i] = \mathrm{E}\left(y_i x_i - \alpha x_i - \beta x_i^2\right) = 0\end{aligned}$$

   Zamieniając momenty teoretyczne empirycznymi uzyskujemy

$$\begin{aligned}\overline{y} - \widehat{\alpha} - \widehat{\beta}\overline{x} &= 0\\ \overline{yx} - \widehat{\alpha}\overline{x} - \widehat{\beta x^2} &= 0\end{aligned}$$

   Co po rozwiązaniu daje nam następujące wzory $\widehat{\beta} = \frac{\overline{yx} - \overline{y}\overline{x}}{\overline{x^2} - \overline{x}^2}$ i $\widehat{\alpha} = \overline{y} - \widehat{\beta}\overline{x}$.

2. Aby precyzyjniejsze oszacowania α i β można zastosować estymator UMM z optymalną macierzą wag. W tym celu:

   (a) liczymy estymatory $\widehat{\alpha}, \widehat{\beta}$ ze wzorów z punktu 1

   (b) znajdujemy macierz wag $\widehat{\boldsymbol{A}}_{kl} = \sum_{i=1}^{n} m_k\left(\widehat{\alpha}, \widehat{\beta}\right) m_l\left(\widehat{\alpha}, \widehat{\beta}\right)$ i znajdujemy estymatory α i β minimalizując względem α, β formę kwadratową

$$Q(\alpha, \beta) = \boldsymbol{m}'(\alpha, \beta)\, \widehat{\boldsymbol{A}}\boldsymbol{m}(\alpha, \beta)$$

   gdzie $\boldsymbol{m}(\alpha, \beta) = \sum_{i=1}^{n} [m_{1i}(\alpha, \beta), m_{1i}(\alpha, \beta)]$.