Egzamin z ekonometrii 17.06.2008
2 semestr
Informatyka i Ekonometria

ZADANIE 1 Oszacowano następujący model (grawitacyjny), wyjaśniający wielkość handlu między parami krajów OECD:

$$lGross\_Imp_{it} = \beta_0 + \beta_1 lrep\_GDP_{it} + \beta_2 lpart\_GDP_{it} + \beta_3 gatt\_wto_{it} + \beta_4 ldist_i + \beta_5 contig_i + c_i + \varepsilon_{it},$$

gdzie $lGross\_Imp$ jest logarytmem importu brutto, $lrep\_GDP_{it}$ jest logarytmem GDP kraju reportera (kraju importującego), $lpart\_GDP_{it}$ jest logarytmem GDP kraju partnera (kraju, którego dobra są importowane), $gatt\_wto_{it}$ jest zmienną zerojedynkową przyjmującą wartość 1, jeśli oba kaje należą do GATT, $ldist_i$ jest logarytmem odległości między krajem partnerem i reporterem, $contig_i$ zmienną zerojedynkową przyjmującą wartość 1, jeśli partner i reporter mają wspólną granicę. Indeks i przebiega po wszystkich parach krajów znajdujących się w bazie danych. Poniżej znajdują się wyniki estymacji tego modelu dokonanej za pomocą MNK, MNK z odporną macierzą wariancji, estymatora efektów losowych i estymatora efektów stałych.

MNK

```
      Source |       SS       df       MS              Number of obs =   26926
-------------+------------------------------           F(  5, 26920) =23920.77
       Model |  208926.011     5  41785.2022           Prob > F      =  0.0000
    Residual |  47024.3053 26920  1.74681669           R-squared     =  0.8163
-------------+------------------------------           Adj R-squared =  0.8162
       Total |  255950.316 26925  9.50604703           Root MSE      =  1.3217

------------------------------------------------------------------------------
  lGross_Imp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lrep_GDP |   .8787423    .004286   205.02   0.000     .8703414    .8871431
   lpart_GDP |   1.049853   .0042898   244.73   0.000     1.041445    1.058261
    gatt_wto |   .4614198   .0194011    23.78   0.000     .4233926    .4994471
       ldist |  -1.018892   .0082867  -122.96   0.000    -1.035134   -1.002649
      contig |   1.068317   .0370955    28.80   0.000     .9956078    1.141026
       _cons |  -30.09896    .154321  -195.04   0.000    -30.40144   -29.79649
------------------------------------------------------------------------------
```

MNK z odpornym, warstwowym estymatorem macierzy wariancji

```
Linear regression                                      Number of obs =   26926
                                                       F(  5,  2173) = 2142.46
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.8163
Number of clusters (id) = 2174                         Root MSE      =  1.3217

------------------------------------------------------------------------------
             |               Robust
  lGross_Imp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lrep_GDP |   .8787423   .0136073    64.58   0.000     .8520576    .9054269
   lpart_GDP |   1.049853   .0140416    74.77   0.000     1.022317    1.077389
    gatt_wto |   .4614198   .0512782     9.00   0.000     .3608603    .5619794
       ldist |  -1.018892   .0271707   -37.50   0.000    -1.072175   -.9656084
      contig |   1.068317   .1189752     8.98   0.000     .8349997    1.301634
       _cons |  -30.09896   .5179171   -58.12   0.000    -31.11463    -29.0833
------------------------------------------------------------------------------
```

Estymator efektów losowych

```
Random-effects GLS regression                   Number of obs      =     26926
Group variable (i): id                          Number of groups   =      2174

R-sq:  within  = 0.3977                         Obs per group: min =         1
       between = 0.8199                                        avg =      12.4
       overall = 0.8126                                        max =        14

Random effects u_i ~ Gaussian                   Wald chi2(5)       =  25896.14
corr(u_i, X)       = 0 (assumed)                Prob > chi2        =    0.0000

------------------------------------------------------------------------------
  lGross_Imp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lrep_GDP |   .8055094   .0106547    75.60   0.000     .7846266    .8263922
   lpart_GDP |   .8933495   .0104467    85.52   0.000     .8728744    .9138245
    gatt_wto |   .2426793    .015903    15.26   0.000       .21151    .2738486
       ldist |   -.868799   .0274303   -31.67   0.000    -.9225614   -.8150366
      contig |   1.450915   .1280123    11.33   0.000     1.200015    1.701814
       _cons |  -25.38166   .3412892   -74.37   0.000    -26.05057   -24.71274
-------------+----------------------------------------------------------------
     sigma_u |  1.2845759
     sigma_e |    .609029
         rho |   .8164736   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Wynik testu Breuscha-Pagana

```
Breusch and Pagan Lagrangian multiplier test for random effects:

        lGross_Imp[id,t] = Xb + u[id] + e[id,t]

        Estimated results:
                         |       Var     sd = sqrt(Var)
                ---------+-----------------------------
               lGross~mp |   9.506047       3.083188
                       e |   .3709163        .609029
                       u |   1.650135       1.284576

        Test:   Var(u) = 0
                             chi2(1) = 80953.34
                         Prob > chi2 =     0.0000
```

Estymator efektów stałych

```
Fixed-effects (within) regression               Number of obs      =     26926
Group variable (i): id                          Number of groups   =      2174

R-sq:  within  = 0.3994                         Obs per group: min =         1
       between = 0.6355                                        avg =      12.4
       overall = 0.6276                                        max =        14

                                                F(3,24749)         =   5485.90
corr(u_i, Xb)  = 0.0831                         Prob > F           =    0.0000

------------------------------------------------------------------------------
  lGross_Imp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lrep_GDP |   .8938499    .020053    44.57   0.000     .8545447     .933155
   lpart_GDP |   .6866236   .0194959    35.22   0.000     .6484105    .7248368
    gatt_wto |   .2695054   .0163246    16.51   0.000     .2375082    .3015027
       ldist |  (dropped)
      contig |  (dropped)
       _cons |  -29.06958   .3601506   -80.72   0.000     -29.7755   -28.36367
-------------+----------------------------------------------------------------
     sigma_u |  1.8738165
     sigma_e |  .60901669
         rho |  .90445833   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(2173, 24749) =    46.96         Prob > F = 0.0000
```

Wynik testu Hausmana

```
                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |      b_fe         b_re         Difference          S.E.
-------------+----------------------------------------------------------------
    lrep_GDP |    .8938499     .8055094         .0883404         .0169883
   lpart_GDP |    .6866236     .8933495        -.2067258         .0164608
    gatt_wto |    .2695054     .2426793         .0268262         .0036863
------------------------------------------------------------------------------
                      b = consistent under Ho and Ha; obtained from xtreg
           B = inconsistent under Ha, efficient under Ho; obtained from xtreg

                  chi2(3) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =      392.73
                Prob>chi2 =      0.0000
```

1. Zinterpretować oszacowania parametrów uzyskanych z estymacji przeporwadzonej za pomocą MNK.

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

3. Na podstawie oszacowań uzyskanych za pomocą estymatora efektów losowych podać liczbę par krajów 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. Zbadać łączną istotność zmiennych i istotność poszczególnych zmiennych uzyskanych na podstawie estymatora efektów stałych. Sprawadzić, czy efekty indywidualne są istotne.

6. Wyjaśnić, dlaczego w przypadku estymatora efektów stałych nie uzyskano oszacowania współczynników dla zmiennych $ldist$ i $contig$.

7. Zinterpretować wynik testu Hausmana i wyjaśnić, jakie są wnioski wynikające z wyniku testu testu dotyczące prawidłowego sposóbu estymacji rozpatrywanego modelu.

Rozwiązanie:

1. wspólczynnik przy $lrep\_GDP$ jest elastycznością - import wzrośnie o 0.87%; jeśli GDP reportera wzrośnie o 1%; wspólczynnik przy $lpart\_GDP$ jest elastycznością - import wzrośnie o 1.39% jeśli GDP partnera wzrośnie o 1%; współczynnik przy zmiennej $gatt\_wto$ mówi, że import jest większy o $[\exp(0.46) - 1] \times 100\% = 58.4\%$, jeśli oba kraje należą do GATT; współczynnik przy $ldist$ jest elastycznością - import spadłby o 1.01%, jeśli odległołość między krajami wzrosłaby o 1%, współczynnik przy $contig$ mówi, że import z kraju, z którym reporter ma wspólną granicę jest o $[\exp(1.07) - 1] \times 100\% = 191.5\,\%$ większy niż z krajmi, z którymi nie ma wspólnej granicy.

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 26926 par krajów, 39.77% zróżnicowania importu dla poszczególnych par została wyjaśniona przez zmienną wewnątrzobiektową dla tych par, 81.99% zróżnicowania importu różnymi parami krajów została wyjaśniona przez zróżnicowanie między parami, 81.26% ogólnego zróżnicowania została wyjaśniona przez model. Hipotezę o łącznej nieistotności wszystkich zmiennych na podstawie testu Walda $[25896.14, 0.000 < 0.05]$

4. Odchylenie standardowe efektu indywidualnego wynosi 1.28, ochylenie standardowe czystego błędu losowego wynosi 0.61, udział wariancji efektu indywidualnego w łącznym błędzie losowym (współczynnik korelacji dla łącznych błędów losowych dla dwóch różnych obserwcji dla tej samej jednostki) wynosi $.82$. Wielkość statystyki $[80953.34, 0.000 < 0.05]$ pozwala nam odrzucić hipotezę zerową o braku efektów losowych

5. Odrzucamy H_0 o łącznej nieistotności zmiennych objaśniających $[5485.9, 0.000 < 0.05]$. Stała, $lrep\_GDP$, $lpart\_GDP$, $gatt\_wto$ są istotne $[-29.06, 0.000 < 0.05]$, $[44.57, 0.000 < 0.05]$, $[35.33, 0.000 < 0.05]$, $[16.51, 0.000 < 0.05]$. Odrzucamy H_0 o nieistotności efektów indywidualnych $[46.96, 0.000 < 0.05]$

6. Zmienne $ldist$ i $contig$ dla poszczególnych par krajów nie zmieniają 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 $[392.73; 0.000 < 0.05]$. Tym samym do estymacji tego modelu nie powinno się stosować estymatora efektów losowych, ponieważ nie będzie on zgodny.

ZADANIE 2 Zmienna losowa y_i pochodzi z rozkładu Poissona z parametrem $\lambda_i = \exp(\boldsymbol{x}_i\boldsymbol{\beta})$ oraz y_i i y_j niezależnymi dla $i \neq j$.

1. Pokazać, że estymator MNW w tym modelu znajdujemy rozwiązując względem $\boldsymbol{\beta}$ następujący układ równań:
$$\sum_{i=1}^{N} e_i \boldsymbol{x}_i = \mathbf{0} \tag{*}$$
gdzie $e_i = y_i - \lambda_i$.

   Podpowiedź: $\Pr(y = k) = \frac{e^{-\lambda_i}\lambda_i^k}{k!}$

2. Pokazać, że estymator UMM w tym modelu możemy znaleźć rozwiązując układ równań identyczny do układu (*).

   Podpowiedź: W rozpatrywanym modelu $\mathrm{E}(y_i|\,\boldsymbol{x}_i) = \lambda_i$.

3. Udowodnić, że estymator Nieliniowej Metody Najmnieszych Kwadratów ($NMNK$) parametru $\boldsymbol{\beta}$ znajdujemy rozwiązując układ równań:
$$\sum_{i=1}^{N} e_i \boldsymbol{x}_i \lambda_i = \mathbf{0}$$

4. Pokazać, że w przypadku zastosowania NMNK występuje problem heteroskedastyczności czynnika losowego. Wykazać, że stosując estymator analogiczny do estymatora ważonej MNK, uzykamy znowu estymator MNW.

5. W przypadku modelu Poissona podnosi się często, że istotnym jego ograniczeniem jest założenie, że $\mathrm{E}\left(y_i | \boldsymbol{x}_i\right) = \mathrm{Var}\left(y_i | \boldsymbol{x}_i\right)$. Na podstawie wyników z poprzednich punktów wykazać, że estymator MNW parametru $\boldsymbol{\beta}$ uzyskany z modelu Poissona jest zgodny nawet wtedy, gdy założenie o równości wartości oczekiwanej i wariancji nie jest spełnione.

Rozwiązanie:

1.
$$\ln \Pr\left(y_i = k\right) = \ln\left(\frac{e^{-\lambda}\lambda_i^k}{k!}\right) = -\lambda_i + k\ln\lambda_i - \ln\left(y_i!\right)$$

   Logarytm funkcja wiarygodności jest więc równy

$$\ell\left(\boldsymbol{\beta}\right) = \sum_{i=1}^{N} \ln \Pr\left(y = y_i\right) = \sum_{i=1}^{N} \left[y_i \ln\lambda_i - \lambda_i - \ln\left(y_i!\right)\right]$$

   Ponieważ

$$\frac{\partial \ln \Pr\left(y = y_i\right)}{\partial \lambda_i} = \frac{y_i}{\lambda_i} - 1$$

   więc logarytm funkcji wiarygodności jest więc równy

$$\begin{aligned}\frac{\partial \ell\left(\boldsymbol{\beta}\right)}{\partial \boldsymbol{\beta}} &= \frac{\partial \sum_{i=1}^{N} \ln \Pr\left(y = y_i\right)}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{N} \frac{\partial \ln \Pr\left(y = y_i\right)}{\partial \lambda_i}\frac{\partial \lambda_i}{\partial \boldsymbol{\beta}} \\ &= \sum_{i=1}^{N}\left(\frac{y_i}{\lambda} - 1\right)\lambda_i \boldsymbol{x} = \sum_{i=1}^{N}\left(y_i - \lambda_i\right)\boldsymbol{x} = \sum_{i=1}^{N} e_i \boldsymbol{x}\end{aligned}$$

2. Moment warunkowy

$$\mathrm{E}\left(y_i | \boldsymbol{x}_i\right) = \lambda_i$$

   wynika z tego, że

$$\mathrm{E}\left(y_i - \lambda_i | \boldsymbol{x}_i\right) = \mathrm{E}\left(\varepsilon_i | \boldsymbol{x}_i\right) = 0$$

   gdzie $\varepsilon_i = y_i - \lambda_i$.
   Moment bezwarunkowy

$$\mathrm{E}\left(\varepsilon_i \boldsymbol{x}_i\right) = 0$$

   Odpowiednik próbkowy

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

3. Nieliniowa Metoda Najmniejszych Kwadratów:

$$y_i = \lambda_i + \varepsilon_i$$

   Estymator liczymy minimalizując następującą sumę kwadratów reszt:

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{N}\left(y_i - \lambda_i\right)^2 = \sum_{i=1}^{N} e_i^2$$

   Warunek pierwszego rzędu

$$\frac{\partial \sum_{i=1}^{N}\left(y_i - \lambda_i\right)^2}{\partial \boldsymbol{\beta}} = 2\sum_{i=1}^{N}\left(y_i - \lambda_i\right)\lambda_i \boldsymbol{x}_i = 2\sum_{i=1}^{N} e_i \lambda_i \boldsymbol{x}_i = 0$$

4. Zauważmy, że

$$\mathrm{Var}\left(\varepsilon_i \mid \boldsymbol{x}_i\right) = \mathrm{Var}\left(y_i - \lambda_i \mid \boldsymbol{x}_i\right) = \mathrm{Var}\left(y_i \mid \boldsymbol{x}_i\right) = \lambda_i$$

Metoda ważona

$$\boldsymbol{x}_i^* = \frac{\boldsymbol{x}_i}{\sqrt{\lambda_i}},\ y_i^* = \frac{y_i}{\sqrt{\lambda_i}}$$

w konsekwencji $e_i^* = \frac{e_i}{\sqrt{\lambda_i}}$ i rozwiązujemy układ

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

5. Skoro postać estymatora MNW w modelu Poissona jest taka sama jak estymatora UMM przy założeniu, że $\mathrm{E}\left(y_i \mid \boldsymbol{x}_i\right) = \lambda_i$, więc estymator ten jest zgodny jeśli tylko zmienna zależna spełnia to założenie.

ZADANIE 3 Na potrzeby pewnej instytucji kredytowej oszacowano model dla prawdopodobieństwa, że osoba, której udzielono kredytu, spłaci go w terminie. Zmienna objaśniana: kredyt – 1 dla osób, które spłaciły kredyt w terminie, 0 w przeciwnym przypadku. Zmienne objaśniające: *edu* (1 – wykształcenie podstawowe lub brak, 2 – wykształcenie średnie, 3 – wykształcenie wyższe); *wiek* (wiek w latach); *dług* (stosunek wartości wnioskowanego kredytu do dochodu kredytobiorcy); *płeć* (0 – mężczyzna, 1 – kobieta). W modelu umieszczono, również zmienną *płeć*×*dług*. Oszacowania modelu logitowego były następujące:

Zmienne	**Współczynniki**	**Ilorazy szans**
stała	0,23	–
płeć	-0,11	0,78
edu2	1,29	1,23
edu3	1,78	1,59
(1-płeć)×dług	1,52	1,06
płeć×dług	-1,78	0,96
wiek	3,45	X

Poniżej znajdują się tabele trafności prognoz uzyskanych na podstawie oszacowanego modelu, dla trzech punktów odcięcia.

Prognozowane: $p^* = 0,75$	Zaobserwowane 1	0	Razem
1	250	48	298
0	617	85	702
Razem	867	133	1000

Prognozowane: $p^* = 0,50$	Zaobserwowane **1**	**0**	**Razem**
1	450	89	539
0	417	44	461
Razem	867	133	1000

Prognozowane: $p^* = 0,25$	Zaobserwowane 1	0	Razem
1	720	119	839
0	147	14	161
Razem	867	133	1000

1. Zinterpretować oszacowania parametrów modelu.
2. Zinterpretować wielkości ilorazów szans.
3. Co można powiedzieć o wielkości ilorazu szans dla zmiennej wiek (oznaczone literą X). Wskazać przedział w jakim może znaleźć się X z uzasadnieniem.
4. Dla $p^* = 0,50$

   wrażliwość =

   specyficzność =

5. Załóżmy, że koszt związany z nieterminową spłatą kredytu wynosi przeciętnie 10 tys. PLN oraz że oczekiwany średni przychód z terminowo spłaconego kredytu wynosi 5 tys. PLN.

   (a) Wskazać punkt odcięcia p^* (do wyboru trzy wartości $0,25; 0,50; 075$), dla którego oczekiwany zysk instytucji będzie najwyższy (lub najniższa strata). Wybór uzasadnić stosowanym obliczeniem. (Oczywiście należy założyć, że instytucja udzieli kredytu tylko osobie, dla której model wyprognozuje terminową spłatę oraz że odsetek niespłaconych kredytów jest w przybliżeniu taki, jak zaobserwowano.) Czy w tym przypadku opłaca się maksymalizować wrażliwość czy specyficzność?

(b) Czy odpowiedź zmieni się, jeżeli odsetek osób niespłacających kredyty wzrośnie do 50%. Odpowiedź uzasadnić.

Rozwiązanie:

1.

2.

3. $0 < X < 1$, bo oszacowanie parametru przy wieku jest ujemne

4. Dla $p^* = 0,50$

   wrażliwość $=450/867$

   specyficzność = $44/133$

5.

   (a) p – odsetek niespłacanych kredytów:
   korzystamy z wzoru na wartość oczekiwaną zysku uzyskanego na jednym kredycie:

   $$wrażliwość \times (1-p) * 5000 + (1 - specyficzność) \times p \times (-10000)$$

   $\widehat{p} = 133/1000$
   Alternatywne rozwiązanie: liczy się dla każdej tabelki zysk wykorzystując tablicowane ilości udzielonych kredytów (spłaconych i niespłaconych) dla $p^* = 0,25$ (najlepszy wybór):
   $720 * 5000 - 119 \times 10000 = 1220 tys.$

   (b) teraz działalności przynosi straty