Egzamin z ekonometrii 17.06.2010

1. Opisać procedurę od ogólnego do szczegółowego na przykładzie doboru liczby opóźnień w modelu.
2. Na czym polega najważniejsza różnica między testowaniem stacjonarności za pomocą testu ADF i $KPSS$?
3. Co modelowane jest w przypadku, kiedy zmienna zależna jest zmienną binarną (zerojedynkową)? Wyjaśnić, jaka jest relacja między zmienną obserwowalną a ukrytą w przypadku modeli dla zmiennych binarnych i jak tę relację można uzasadnić w oparciu o teorię ekonomii.
4. Wyjaśnić, dlaczego za pomocą estymatora efektów stałych nie można oszacować parametrów przy zmiennych nie zmieniających się w czasie.

ZADANIE 1 Dla danych miesięcznych z lat 2000.4 do 2008.7 wyestymowano następujący model dla dynamiki płacy realnej brutto w sektorze przedsiębiorstw:
$\text{płaca}_t = \beta_1 + \beta_2 \text{płaca}_{t-1} + \beta_3 \text{inf}_t + \beta_4 \text{inf}_{t-1} + \beta_5 \text{prod}_t + \beta_6 \text{rejest}_t + \varepsilon_t$,
gdzie inf_t jest stopą dynamiki inflacji, prod_t jest stopą dynamiki produkcji przemysłowej w wyrażeniu realnym a rejest_t jest stopą bezrobocia rejestrowanego. Wszystkie dynamiki są dynamikami policzonymi w stosunku do analogicznego miesiąca poprzedniego roku. Uzyskano następujące wyniki regresji:

```
      Source |       SS       df       MS              Number of obs =     100
-------------+------------------------------           F(  5,    94) =   52.57
       Model |  558.628608     5  111.725722           Prob > F      =  0.0000
    Residual |  199.772594    94  2.12524036           R-squared     =  0.7366
-------------+------------------------------           Adj R-squared =  0.7226
       Total |  758.401202    99   7.6606182           Root MSE      =  1.4578

------------------------------------------------------------------------------
       placa |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       placa |
         L1. |   .2219342   .0950748     2.33   0.022     .0331609    .4107075
         inf |
         --. |  -1.431173   .3629814    -3.94   0.000    -2.151881   -.7104653
         L1. |    .964093   .3642931     2.65   0.010     .2407805    1.687405
        prod |   .0962173   .0271911     3.54   0.001     .0422288    .1502058
      rejest |  -.5947121   .0856992    -6.94   0.000      -.76487   -.4245542
       _cons |    127.748    17.3643     7.36   0.000      93.2708    162.2252
------------------------------------------------------------------------------
```

```
Durbin-Watson test statistic:    (6,100)    = 2.033021
Breusch-Godfrey LM statistic:    Chi-sq(1)  = 1.167     [.2800]
ARCH test statistic:             Chi-sq(1)  = 5.094     [.0240]
```

Wartości krytyczne testu DW: $d_L = 1.57$, $d_U = 1.78$

1. Zbadać występowanie autokorelacji błędu losowego w tym modelu. Jaki test powinno się zastosować w tym celu w analizowanym modelu? Dlaczego testowanie występowania autokorelacji jest ważne w modelach ADL?
2. Zintepretować wynik testu ARCH.
3. Zintrpretować wielkość mnożnika bezpośredniego dla wielkości produkcji.
4. Policzyć i zintrpretować mnożnik długookresowy dla inflacji.
5. Policzyć rowiązanie długookresowe dla tego modelu.

Rozwiązanie:

1. Wartość p dla testu Breuscha-Godfreya $[0.28 > 0.05]$ oznacza, że nie można odrzucić H_0 o braku autokorelacji. W przypadku modeli ze opóźnioną zmienną zależną nie można stosować do testowania autokorelacji testu DW.

2. Wartość p dla testu ARCH $[0.024 < 0.05]$ oznacza, że musimy odrzucić H_0 o braku heteroskedastyczności warunkowej i przyjąć H_1, że heteroskedastyczność warunkowa występuje w modelu.

3. Wielkość mnożnika bezpośredniego dla wielkości produkcji $[3.54]$ oznacza, że wzrost produkcji o jeden punkt procentowy wiąże się ze zwiększeniem wysokości płacy realnej w tym samym okresie o 3.54 punktu procentowego.

4. Wielkość mnożnika długookresowego dla inflacji jest równa:

$$\frac{\beta_3 + \beta_4}{1 - \beta_2} = \frac{-1.431173 + .964093}{1 - .2219342} = -0.600\,31$$

   Oznacza to, że trwałe zwiększenie inflacji o jedne punkt procentowy powoduje spadek płacy realnej o 0.6 punktu procentowego

5. Rozwiązanie długookresowe

$$\text{płaca} = \frac{\beta_1}{1-\beta_2} + \frac{\beta_3+\beta_4}{1-\beta_2}\text{inf} + \frac{\beta_5}{1-\beta_2}\text{prod} + \frac{\beta_6}{1-\beta_2}\text{rejest}$$

$$\begin{aligned}\text{płaca} &= \frac{127.748}{1-.2219342} + \frac{-1.431173+.964093}{1-.2219342}\text{inf} + \frac{.0271911}{1-.2219342}\text{prod} + \frac{.0856992}{1-.2219342}\text{rejest} \\ &= 164.\,19 - 0.6\text{inf} + 0.035\text{prod} + 0.11\text{rejest}\end{aligned}$$

ZADANIE 2 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

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

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

1	2	3	4	5	$\sum$

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.
$$\begin{array}{rc} \textit{zmienne egzogeniczne} & 1, I_t, G_t \\ \textit{zmienne endogeniczne} & C_t, Y_t \end{array}$$
$$K = 3 \quad G_1 = 2 \quad K_1 = 1 \quad \text{zidentyfikowane } 3 = K \geq G_1 + K_1 - 1 = 2$$
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