Egzamin z ekonometrii 25.06.2007
II semestr

Pytania teoretyczne

1. Opisać procedurę "od ogólnego do szczegółowego" na przykładzie doboru liczby opóźnień w modelu.
2. Wyjaśnić jak należy rozumieć przyczynowość w sensie Grangera i jak ją testujemy.
3. Co to są ilorazy szans i dlaczego w kontekście modelu logitowego lepiej jest używać ilorazów szans niż efektów krańcowych?
4. Wyjaśnić skąd bierze się problem identyfikacji w modelach wielorównaniowych. Podać warunek konieczny identyfikowalności równania modelu.

ZADANIE 1 Mamy y_t dane procesem $AR(1)$

$$y_t = \alpha y_{t-1} + \varepsilon_t$$

ze skorelowanym błędem losowym ε_t

$$\varepsilon_t = \rho u_{t-1} + u_t$$

gdzie $\mathrm{Cov}(u_t, u_{t-s}) = 0$ dla $s \neq 0$, $\mathrm{Var}(u_t) = \sigma_u^2$, $|\alpha| < 1$, $|\rho| < 1$.

1. Pokazać, że $y_t = \sum_{i=0}^{\infty} \alpha^i u_{t-i} + \rho \sum_{i=0}^{\infty} \alpha^i u_{t-i-1}$, policz wariancję y_t.
2. Pokazać, że w modelu tym występuje problem równoczesności.

   Podpowiedź: skorzystać z tego (bez dowodzenia), że $\mathrm{Cov}(y_{t-k-h}, u_{t-k}) = 0$ dla każdego $h > 0$
3. Pokazać, że y_{t-2} można użyć jako zmiennej instrumentalnej w MZI celem zgodnego oszacowania parametru α.

Rozwiązanie:

1. Ponieważ $y_{t-1} = \alpha y_{t-2} + \varepsilon_{t-1}$, więc podstawiając do modelu otrzymujemy $y_{t-1} = \alpha y_{t-2} + \varepsilon_{t-1}$ następnie podstawiamy $y_{t-2} = \alpha y_{t-3} + \varepsilon_{t-3}$ itd. i wreszcie otrzymujemy wzór:

$$y_t = \sum_{i=0}^{\infty} \alpha^i \varepsilon_{t-i}$$

   podstawiając $\varepsilon_t = \rho u_{t-1} + u_t$ otrzymujemy pierwszy żądany wzór. Korzystając z tego wzoru

$$\begin{aligned}
\mathrm{Var}(y_t) &= \mathrm{Var}\left(\sum_{i=0}^{\infty} \alpha^i u_{t-i} + \rho \sum_{i=0}^{\infty} \alpha^i u_{t-i-1}\right) \\
&= \mathrm{Var}\left[u_t + (\alpha+\rho) \sum_{i=0}^{\infty} \alpha^i u_{t-i-1}\right] \\
&= \sigma_u + (\alpha+\rho)^2 \sum_{i=0}^{\infty} \alpha^{2i}\, \mathrm{Var}(u_{t-i-1}) \\
&= \sigma_u + \frac{(\alpha+\rho)^2}{1-\alpha^2} \sigma_u^2 = \left[1 + \frac{(\alpha+\rho)^2}{1-\alpha^2}\right] \sigma_u^2
\end{aligned}$$

2. Problem równoczesności występuje, gdy zmienna objaśniająca skorelowana jest z błędem losowym. W naszym przypadku zmienną objaśniającą jest $y_{t-1} = \alpha y_{t-2} + \rho u_{t-2} + u_{t-1}$

$$\begin{aligned}
\mathrm{Cov}(y_{t-1}, \varepsilon_t) &= \mathrm{Cov}(\alpha y_{t-2} + \rho u_{t-2} + u_{t-1}, \rho u_{t-1} + u_t) \\
&= \alpha\rho \underbrace{\mathrm{Cov}(y_{t-2}, u_{t-1})}_{0} + \rho^2 \underbrace{\mathrm{Cov}(u_{t-2}, u_{t-1})}_{0} + \rho\, \mathrm{Var}(u_{t-1}) \\
&+ \alpha \underbrace{\mathrm{Cov}(y_{t-2}, u_t)}_{0} + \rho \underbrace{\mathrm{Cov}(u_{t-2}, u_t)}_{0} + \underbrace{\mathrm{Cov}(u_{t-1}, u_t)}_{0} \\
&= \rho\sigma_u^2
\end{aligned}$$

   przy czy skorzystaliśmy z braku atokorelacji u_t i tego, że $\mathrm{Cov}(y_{t-k-h}, u_{t-k}) = 0$ dla każdego $h > 0$. Istotnie więc kowariancja między y_{t-1} i ε_t nie jest równa zeru

3. y_{t-2} jest dobrym instrumentem jeśli jest nieskorelowane z ε_t ale skorelowane z y_t. Korelacja między y_{t-2} i ε_t jest równa:

$$\text{Cov}\left(y_{t-2},\varepsilon_t\right)=\text{Cov}\left(y_{t-2},\rho u_{t-1}+u_t\right)=\rho\,\text{Cov}\left(y_{t-2},u_{t-1}\right)+\text{Cov}\left(y_{t-2},u_t\right)=0$$

Korelacja między y_{t-2} i y_t

$$\begin{aligned}\text{Cov}\left(y_{t-2},y_t\right)=\text{Cov}\left(y_{t-2},y_t\right)&=\text{Cov}\left(y_{t-2},\alpha^2 y_{t-2}+\varepsilon_{t-1}+\varepsilon_t\right)\\&=\alpha^2\underbrace{\text{Var}\left(y_{t-2}\right)}_{\text{Var}(y_t)}+\underbrace{\text{Cov}\left(y_{t-2},\varepsilon_{t-1}\right)}_{\text{Cov}(y_{t-1},\varepsilon_t)}+\underbrace{\text{Cov}\left(y_{t-2},\varepsilon_t\right)}_{0}\\&=\alpha^2\left[1+\frac{(\alpha+\rho)^2}{1-\alpha^2}\right]\sigma_u^2+\rho\sigma_u\end{aligned}$$

Zadanie 2

Zadanie 3 Oszacowano model logitowy wyjaśniający prawdopodobieństwo, że osoba nie posiadająca pracy będzie jej poszukiwać. Zmiennymi niezależnymi są wiek oraz płeć. Otrzymano następujące oszacowania dla parametrów oraz następujące wielkości ilorazów szans:
Testy przeprowadzać na poziomie istotności $\alpha = 0.05$.

```
Logit estimates                                   Number of obs   =      27501
                                                  LR chi2(2)      =    2448.19
                                                  Prob > chi2     =     0.0000
Log likelihood = -12678.788                       Pseudo R2       =     0.0880

------------------------------------------------------------------------------
  LOOK4WORK   |      Coef.   Std. Err.       z     P>|z|      [95% Conf. Interval]
-------------+----------------------------------------------------------------
     _ISEX_1 |  -.3568389   .0313973   -11.37   0.000     -.4183765   -.2953012
         AGE |   -.034242   .0007752   -44.17   0.000     -.0357613   -.0327227
       _cons |   .2516005   .0358808     7.01   0.000      .1812755    .3219256
------------------------------------------------------------------------------
```

```
Logit estimates                                   Number of obs   =      27501
                                                  LR chi2(2)      =    2448.19
                                                  Prob > chi2     =     0.0000
Log likelihood = -12678.788                       Pseudo R2       =     0.0880

------------------------------------------------------------------------------
  LOOK4WORK   | Odds Ratio   Std. Err.       z     P>|z|      [95% Conf. Interval]
-------------+----------------------------------------------------------------
     _ISEX_1 |   .6998853   .0219745   -11.37   0.000      .6581144    .7443074
         AGE |   .9663376   .0007491   -44.17   0.000      .9648705    .9678068
------------------------------------------------------------------------------
```

```
SEX:       0 mezczyzna
           1 kobieta
LOOK4WORK:   0 nie poszukuje pracy
             1 poszukuje pracy
```

1. Wypisać założenia modelu logitowego.
2. Sprawdzić, czy zmienne w modelu są łacznie istotne. Zinterpretować Pseudo R^2 McKelveya-Zavoiny.
3. Podać, które zmienne w modelu są istotne.

4. Zintepretować wielkości ilorazów szans dla poszczególnych zmiennych. Czy relacje między ilorazami szans a wielkościami parametrów odpowiadają twoim oczekiwaniom?

5. Wartość funkcji gęstości policzona dla $\overline{\boldsymbol{x}}\boldsymbol{b}$ jest równa 0.145. Policzyć krańcowy wpływ wieku na prawdopodobieństwo poszukiwania pracy i zinterpretować obliczoną wartość.

6. Postanowiono sprawdzić hipotezę, że wykształcenie ma wpływ na prawdopodobieństwo poszukiwania pracy. Wartość funkcji wiarygodności w maksimum dla modelu logitowego ze zmiennymi objaśniającymi wiek, płeć i wykształcenie wyniosła -11052.407. Zmienna wykształcenie mogła przyjmować jeden z 9 poziomów. Zweryfikować hipotezę mówiącą o tym, że wykształcenie wpływa na prawdopodobieństwo poszukiwania pracy.

   Podpowiedź: $\chi^2_{0.95}(7) = 14.07$, $\chi^2_{0.95}(8) = 15.51$, $\chi^2_{0.95}(9) = 16.91$, $\chi^2_{0.95}(10) = 18.31$

Rozwiązanie:

1. Zmienna ukryta
$$y_i^* = \boldsymbol{x}_i\boldsymbol{\beta} + \varepsilon$$
   ε_i ma rozkład logitowy i poszczególne obserwacje są niezależne. Obserwujemy
$$\begin{array}{lll} y = 0 & \text{dla} & y^* \leq 0 \\ y = 1 & \text{dla} & y^* > 0 \end{array}$$
   Alternatywna odpowiedź: prawdopodobieństwa zajścia pojedynczego zdarzenia:
$$\Pr(y_i) = \begin{cases} 1 - \Lambda(\boldsymbol{x}\boldsymbol{\beta}) & \text{dla} \quad y_i = 0 \\ \Lambda(\boldsymbol{x}\boldsymbol{\beta}) & \text{dla} \quad y_i = 1 \end{cases}$$
   gdzie $\Lambda(\boldsymbol{x}_i\boldsymbol{\beta}) = \frac{\exp(\boldsymbol{x}_i\boldsymbol{\beta})}{1+\exp(\boldsymbol{x}_i\boldsymbol{\beta})}$ i poszczególne obserwacje są niezależne.

2. O łącznej istotności zmiennych w modelu świadczy wielkość statystyki $LR = 2448.19$. Hipotezę o łącznej nieistotności zmiennych objaśniających odrzucamy $[0.000 < 0.05]$. Wielkość statystyki Pseudo-R^2 McKelveya i Zavoiny świadczy o tym, że 8.8% zmienności zmiennej ukrytej zostałaby wyjaśniona przez model, gdyby zmienna ukryta była bezpośrednio obserwalna. .

3. W modelu istotnymi zmiennymi są: SEX $[-11.37, 0.000 < 0.05]$, AGE $[-44.17, 0.000 < 0.05]$, stała $[7.01, 0.000 < 0.05]$

4. Szansa to prawdopodobieństwa sukcesu do prawdopodobieństwa porażki. Szansa poszukiwania pracy jest dla kobiet jest równa 69% szansy poszukiwania pracy dla mężczyzn. Osoby o jeden o wieku wyższym o rok od średniej w próbie ma szansę poszukiwania pracy równą 97% szansy poszukiwania pracy dla osoby o wieku równym średniej w próbie. Dla ilorazów szans mniejszych od 1 (100%) a więc związanych ze spadkiem szansy obserwujemy ujemne oszacowania współczynników - jest to zgodne z tym, że ilorazy szans są równe $\exp(\Delta x_k \beta_k)$ a więc są mniejsze od 1 dla $\beta_k < 0$.

5. Efekt cząskowy można policzyć z ogólnego wzoru na efekt cząskowy w modelach dla zmiennej binarnej:
$$\frac{\partial E(y|\,x)}{\partial wiek} = f(\overline{\boldsymbol{x}}\boldsymbol{b})\beta_{wiek} = 0.145 \times -.0342422 = -.0049651$$
   Wielkość tego efektu cząskowego oznacza, że wraz ze wzrostem wieku o rok prawdopodobieństwo poszukiwania pracy maleje o 0.5%.

6. Statystyka testowa jest równa:
$$LR = 2(-11052.407 + 12678.788) = 3252.8 > \chi^2_{0.95}(8) = 15.51$$
   Odrzucamy hipozę zerowej o o zerowości współczynników przy zmiennych zerojedynkowych związanych z wykształceniem. Przyjęta liczba stopni swobody wynosi 8 ponieważ jeden poziom zmiennej zerojedynkowej związanej z wykształceniem wypada jako poziom bazowy.

ZADANIE 4 Badacz chce wyestymować dynamiczną wersję Keynsowskiej funkcji konsumpcji na szeregu czasowym dla Polski w latach 1994.1 do 2004.4.

1. Na początku badacz przeprowadził regresję ΔY_t na Y_{t-1} i stałej oraz regresję ΔC_t na C_{t-1} i stałej i uzyskał następujące wyniki:

Zmienna	**Współczynnik**	**(Błąd Stand.)**	**t**
Y_{t-1}	-0.067	(0.057)	-1.16
Stała	14982.179	(9454.617)	1.58

Zmienna	**Współczynnik**	**(Błąd Stand.)**	**t**
C_{t-1}	-0.033	(0.016)	-2.03
Stała	7450.293	(2211.963)	3.37

gdzie C jest konsumpcją ogółem w cenach bieżących a Y wielkością PKB w cenach bieżących.

2. Po przeprowadzeniu tych regresji badacz wyestymował Keynsowską funkcję konsumpcji z dodatkowymi zmiennymi zerojedynkowymi związanymi z sezonowością i uzyskał następujące oszacowania:

Zmienna	**Współczynnik**	**(Błąd Stand.)**	**t**
Y	0.849	(0.015)	55.06
kw_2	-5986.197	(1898.578)	-3.15
kw_3	-9853.442	(1906.045)	-5.17
kw_4	-25261.582	(1966.185)	-12.85
Stała	5738.341	(2621.667)	2.19

W kolejnym kroku badacz wygenerował reszty z tak oszacowanego modelu, następnie przeprowadził regresję pierwszych różnic reszt Δe_t na resztach opóźnionych e_{t-1} i uzyskał następujące oszacowania:

Zmienna	**Współczynnik**	**(Błąd Stand.)**	**t**
e_{t-1}	-1.061	(0.158)	-6.70
Stała	243.258	(612.030)	0.40

3. Kończąc badanie badacz przeprowadził regresję ΔY_t na opóźnionych resztach, opóźnieniach ΔY_t oraz opóźnieniach ΔC_t i uzyskał następujący wynik:

Zmienna	**Współczynnik**	**(Błąd Stand.)**	**t**
e_{t-1}	-0.109	(0.177)	0.545
ΔY_{t-1}	-0.167	(0.178)	0.359
ΔY_{t-2}	-0.181	(0.145)	0.222
ΔY_{t-3}	-0.432	(0.144)	0.006
ΔY_{t-4}	0.514	(0.167)	0.005
ΔC_{t-1}	0.891	(0.233)	0.001
ΔC_{t-2}	0.569	(0.303)	0.072
ΔC_{t-3}	0.379	(0.289)	0.202
ΔC_{t-4}	-0.670	(0.283)	0.026
Stała	1209.524	(1350.171)	0.379

Pytania:

1. Wyjaśnić jaki wniosek badacz uzyskał w punkcie 1 (odpowiedź uzasadnić) i czemu służyły te regresje.
2. Wyjaśnić jaki wniosek uzyskał badacz w punkcie 2 (odpowiedź uzasadnić).
3. Podać intepretację elementów w równaniu z punktu 3.

   Wartości krytyczne:

   - test pierwiastka jednostkowego ADF, $\alpha = 0.05$, 25 obs.: bez wyrazu wolnego: -1.95, z wyrazem wolnym: -3.00, z wyrazem wolnym i trendem: -3.60 (*Fuller, 1976)*,
   - test na kointegrację ADF, $\alpha = 0.05$: bez wyrazu wolnego: -2.76, z wyrazem wolnym: -3.37 (*Philips i Ouliaris, 1990)*.

Rozwiązanie:

1. Pierwsze dwie regresje służą przeprowadzeniu testu Dickey-Fullera (DF), na istnienie pierwiastka jednostkowego (niestacjonarność zmiennych). Tylko dla niestacjonarnych zmiennych sens ma testowanie kointegracji. Porównując dla Y_t wielkość statystyki t z wartością krytyczną $-1.16 > -3.00$ (mamy stała w regresji) dochodzimy do wniosku, że H_0 o istnieniu pierwiastka jednostkowego dla Y_t nie można odrzucić. Podobnie dla C_t statystyka $-2.03 > -3.00$ a więc także nie da się odrzucić H_0 o isnieniu pierwiastka jednostkowego. Wniosek: obie zmienne są niestacjonarne, można badać występowanie kointegracji między tymi zminnymi.

2. W tym punkcie badacz bada kointegrację między zmiennymi. Jeśli występuje kointegracja między zmiennymi Y_t i C_t, to resty z regresji C_t na Y_t powinny być stacjonarne. Wielkość statystyki t dla regresji reszt na resztach opóźnionych $-6.70 < -3.37$, że rzeczywiście reszty są stacjonanrne, zatem występuje kointegracja między zmiennymi

3. W ostatniej regresji badacz estymuje mechanizm korekty błędem (ECM) objaśniający szybkość dostosowań do równowagi długookresowej oszacowanej w regresji na poziomach. Współczynnik przy opóźnionych resztach jest współczynnikiem szybkości dostosowań do stanu równowagi, pozostałe elementy związane są z dynamiką krótkookresową. Wielkość statystyki t dla współczynnika szybkości dostosowania do stanu równowagi długookresowej sugeruje, elementy związane z dostosowaniem do rónowagi długookresowej są nieistotne w modelu.