Egzamin z ekonometrii dla IiE 11.06.2007
I semestr

Pytania teoretyczne

1. Udowodnić, że w modelu ze stałą $TSS = ESS + RSS$
2. Pokazać, że s^2 jest nieobciążonym estymatorem σ^2.
3. Jak przeprowadza się test Goldfelda-Quandta?
4. Udowodnij, że estymator MNK jest estymatorem zgodnym. Wypisz założenia konieczne do tego dowodu.

ZADANIE 1

1. Pokazać, że w modelu:
$$y_i = \beta_1 + \beta_2 x_{2i} + \ldots + \beta_K x_{Ki} + \varepsilon_i$$
   (a) semielastyczności oszacowane dla średnich wartości zmiennych objaśniających są równe:
$$\widehat{e}_{y,x_k} = \frac{b_k}{\overline{y}},$$
   (b) elastyczności oszacowane dla średnich wartości zmiennych objaśniających są równe:
$$\widehat{e}_{y,x_k} = \frac{\overline{x}_k b_k}{\overline{y}},$$
   (c) udowodnić, że oszacowane w ten sposób elastyczności sumują się do $1 - \frac{b_1}{\overline{y}}$.
2. Pokazać, że w modelu:
$$y_i = \beta_1 + \beta_2 \ln x_{2i} + \ldots + \ln x_{Ki} + \varepsilon_i$$
   (a) semielastyczności oszacowane dla średnich wartości zmiennych objaśniających są równe:
$$\widehat{e}_{y,x_k} = \frac{b_k}{\overline{y}\overline{x}_k},$$
   (b) elastyczności oszacowane dla średnich wartości zmiennych objaśniających są równe:
$$\widehat{e}_{y,x_k} = \frac{b_k}{\overline{y}}.$$

Rozwiązanie:

1.

   (a) Z definicji semielastyczności:
$$e_{y,x_k} = \frac{\partial \ln \mathrm{E}\left(y_i | \boldsymbol{x}_i\right)}{\partial x_{ki}} = \frac{\partial \ln\left(\beta_1 + \beta_2 x_{2i} + \ldots + \beta_K x_{Ki}\right)}{\partial x_{ki}} = \frac{\beta_k}{\beta_1 + \beta_2 x_{2i} + \ldots + \beta_K x_{Ki}}$$
   dla oszacowanych $\beta_1, \ldots, \beta_K$ i średnich wartości $x_2, \ldots, x_K$ mianownik jest równy
$$b_1 + b_2 \overline{x}_{2i} + \ldots + b_K \overline{x}_{Ki} = \overline{\widehat{y}} = \overline{y}$$
   a więc
$$\widehat{e}_{y,x_k} = \frac{b_k}{\overline{y}}$$

(b) Z definicji elastyczności:

$$e_{y,x_k} = \frac{\partial \ln \mathrm{E}\left(y_i | \boldsymbol{x}_i\right)}{\partial \ln\left(x_{ki}\right)} = \frac{\partial \ln\left(\beta_1 + \beta_2 x_{2i} + \ldots + \beta_K x_{Ki}\right)}{\partial \ln\left(x_{ki}\right)}$$
$$= \frac{\partial \ln\left(\beta_1 + \beta_2 x_{2i} + \ldots + \beta_K x_{Ki}\right)}{\partial x_{ki}} \left(\frac{\partial \ln\left(x_{ki}\right)}{\partial x_{ki}}\right)^{-1}$$
$$= \frac{x_{ki}\beta_k}{\beta_1 + \beta_2 x_{2i} + \ldots + \beta_K x_{Ki}}$$

a więc

$$\widehat{e}_{y,x_k} = \frac{\overline{x}_k b_k}{\overline{y}}$$

(c) Sumując wszystkie oszacowane elastyczności otrzymujemy:

$$\sum_{k=2}^{K} \widehat{e}_{y,x_k} = \frac{\sum_{k=2}^{K} \overline{x}_k b_k}{\overline{y}} = \frac{\overline{y} - b_1}{\overline{y}} = 1 - \frac{b_1}{\overline{y}}$$

(d) Z definicji semielastyczności:

$$e_{y,x_k} = \frac{\partial \ln \mathrm{E}\left(y_i | \boldsymbol{x}_i\right)}{\partial x_{ki}} = \frac{\partial \ln\left(\beta_1 + \beta_2 \ln x_{2i} + \ldots + \beta_K \ln x_{Ki}\right)}{\partial x_{ki}} = \frac{\beta_k}{\left(\beta_1 + \beta_2 \ln x_{2i} + \ldots + \beta_K \ln x_{Ki}\right)} \frac{1}{x_{ki}}$$

a więc

$$\widehat{e}_{y,x_k} = \frac{b_k}{\overline{x}_k \overline{y}}$$

(e) Z definicji elastyczności:

$$e_{y,x_k} = \frac{\partial \ln \mathrm{E}\left(y_i | \boldsymbol{x}_i\right)}{\partial \ln\left(x_{ki}\right)} = \frac{\partial \ln\left(\beta_1 + \beta_2 \ln x_{2i} + \ldots + \beta_K \ln x_{Ki}\right)}{\partial \ln\left(x_{ki}\right)}$$
$$= \frac{\partial \ln\left(\beta_1 + \beta_2 \ln x_{2i} + \ldots + \beta_K \ln x_{Ki}\right)}{\partial x_{ki}} \left(\frac{\partial \ln\left(x_{ki}\right)}{\partial x_{ki}}\right)^{-1}$$
$$= \frac{x_{ki}\beta_k}{\beta_1 + \beta_2 x_{2i} + \ldots + \beta_K x_{Ki}} \frac{1}{x_{ki}} = \frac{\beta_k}{\beta_1 + \beta_2 x_{2i} + \ldots + \beta_K x_{Ki}}$$

a więc

$$\widehat{e}_{y,x_k} = \frac{b_k}{\overline{y}}$$

ZADANIE 2 Oszacowano regresję logarytmu zagregowanej konsumpcji w wyrażeniu *nominalnym* (zmienna y) na logarytmie PKB w wyrażeniu *nominalnym* (zmienna x_1) i logarytmie deflatora (zmienna x_2). Regresję przeprowadzono dla danych kwartalnych dla Polski z okresu 1995-2002. Wprowadzono także do modelu zmienne sezonowe przyjmujące wartość 1 dla odpowiedniego kwartału a 0 dla pozostałych. Wyniki regresji znajdują się w tabeli poniżej.

```
y          |   Coef. Std.Err.   t      P>|t|
-----------+------------------------------
x1         |    .80     .081     9.72   0.000
x2         |   -.59     .117    -5.00   0.000
seasonal2  |   -.05     .011    -4.38   0.000
seasonal3  |   -.07     .011    -6.25   0.000
seasonal4  |   -.17     .012   -14.36   0.000
cons       |   1.87     .926     2.02   0.054
------------------------------------------
Number of obs = 32, F( 5, 26) = 229.88 [0.0000]
R-squared = 0.9779, s = .02027
------------------------------------------
Durbin-Watson test statistic:    d( 6, 32)  =2.261303
```

```
Breusch-Godfrey LM statistic:    Chi-sq( 2) =9.894179 [.0071]
White's general test statistic:  Chi-sq(14) =25.15268 [.0331]
Jarque-Berra test statistic:     Chi-sq(2)  =3.34     [.1886]
Ramsey RESET test statistic:     F(3, 23)   =1.66     [.2030]
Chow test statistic (t=1999.1):  F( 3, 23)  =0.28     [.8359]
```

Przy założonym poziomie istotności $\alpha = 0.01$ przeprowadzić analizę wyników. Każdą z odpowiedzi należy uzasadnić za pomocą odpowiedniego testu.
Podpowiedź: wartości d_L i d_U dla testu DW przy 32 obserwacjach, 5 zmiennych i stałej oraz $\alpha = 0.01$ wynoszą $d_L = 0.917, d_U = 1.597$.

1. Określić, czy model jest dobrze dopasowany, czy zbiór zmiennych niezależnych istotnie objaśnia zmienną zależną.
2. Podać, które zmienne w modelu są istotne.
3. Zbadać, czy w modelu występuje autokorelacja.
4. Zbadać, czy w modelu występuje heteroskedastyczność.
5. Sprawdzić, czy forma funkcyjna modelu jest prawidłowa.
6. Przetestować, czy błąd losowy w modelu ma rozkład normalny.
7. Sprawdzić, czy parametry modelu są stabilne.
8. Zinterpretować współczynnik przy zmiennej x_1.
9. W jaki sposób należałoby weryfikować hipotezę o neutralności pieniądza w gospodarce (tzn., że ważne są jedynie wartości realne a nie nominalne)?
10. W jaki sposób należałoby weryfikować hipotezę mówiącą o tym, że zmienne sezonowe są nieistotne w modelu?
11. Jeśli model nie spełnia założeń KMRL określić:
    (a) które założenie nie jest spełnione?
    (b) jakie ma to konsekwencje dla wnioskowania statystycznego?
    (c) jakie są metody radzenia sobie z tym problemem?

Rozwiązanie:

1. Oszacowany model objaśnia prawie 98% zmienności zmiennej zależnej (współczynnik $R^2 = 0.9779$). Zbiór zmiennych objaśniających (bez stałej) jest łącznie istotny ($F_{5,26} = 229.88$) $[0.0000 < 0.01]$.
2. Wszystkie zmienne objaśniające poza stałą są istotne na zadanym poziomie istotności (*p-value* dla statystyk t są mniejsze od 0.01).
3. W modelu nie występuje autokorelacja I rzędu ponieważ statystyka $DW \approx 2.26 \in (1.597, 2.403)$ przyjmuje wartość, która nie pozwala odrzucić hipotezy zerowej. Odpowiednie wartości krytyczne dla tego testu wynoszą ($d_L = 0.917, d_U = 1.597$). W modelu może jednak występować autokorelacją wyższego rzędu, o czym świadczy *p-value* dla testu Breuscha-Godfreya $[.0071 < 0.01]$.
4. Nie ma podstaw do odrzucenia hipotezy o homoskedastyczności składnika losowego w związku z wartością *p-value* dla testu White'a $[.0331 > 0.01]$.
5. Forma specyfikacji modelu jest prawidłowa, o czym świadczy *p-value* dla testu RESET $[.2030 > 0.01]$.
6. Błąd losowy ma rozkład normalny, o czym świadczy *p-value* dla testu Jarque'a-Bera $[.1886 > 0.01]$.
7. Parametry modelu są stabilne, o czym świadczy *p-value* dla testu Chowa $[.8359 > 0.01]$.
8. Współczynnik przy zmiennej x_1 jest elastycznością zagregowanej konsumpcji względem PKB, zatem wzrost PKB o 1% spowoduje wzrost zagregowanej konsumpcji o 0.8%.

9. Model można zapisać jako

$$\log(cons) = \alpha_1 + \alpha_2 \log(pkb) + \alpha_3 \log(defl) + \varepsilon$$

Konsumpcja i PKB w wyrażeniu realnym są równe odpowiednio $cons^* = \frac{cons}{defl}$, $pkb^* = \frac{pkb}{defl}$. Odejmując od obu stron $\log(defl)$ uzyskujemy

$$\log(cons^*) = \alpha_1 + \alpha_2 \log(pkb^*) + (\alpha_2 + \alpha_3 - 1)\log(defl) + \varepsilon$$

a więc na konsumpcję w wyrażeniu realnym wpływa jedynie dochód w wyrażeniu realnym jeśli prawdziwe jest $H_0 : \alpha_2 + \alpha_3 = 1$.

10. Hipotezę o łącznej nieistotności zmiennych sezonowych należy testować za pomocą testu F, nakładając łączne ograniczenie na parametry przy zmiennych $seasonal_2$, $seasonal_3$, $seasonal_4$, - przyrównując je jednocześnie do zera.

11. (a) Model nie spełnia założenia braku autokorelacji składnika losowego.
    (b) Oszacowanie macierzy wariancji-kowariancji MNK jest obciążone, co powoduje, że niewłaściwe są błędy standardowe parametrów i tym samym niewłaściwe są statystyki t.
    (c) Jeśli nie chcemy usunąć autokorelacji, należy posłużyć się tzw. odporną macierzą wariancji-kowariancji Newey'a-Westa. Można też podjąć próbę usunięcia autokorelacji za pomocą Stosowalnej UMNK.

ZADANIE 3 Na podstawie dwóch niezależnych próbek uzyskano MNK oszacowania $\boldsymbol{b}_1$ i $\boldsymbol{b}_2$ parametru $\boldsymbol{\beta}$ w modelu:

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

Zakładamy, że w obu próbkach spełnione są założenia $KMRL$.
Uwaga: przez niezależność próbek rozumiemy, że błędy losowe w obu próbach są od siebie niezależne.

1. Pokaż, że estymatory $\boldsymbol{b}_1$ i $\boldsymbol{b}_2$ są od siebie niezależne.

2. Przy założeniu, że parametry $\boldsymbol{\beta}$ i σ^2 są w obu próbach jest takie same, jaka będzie wartość oczekiwana i wariancja $\boldsymbol{b}_1 - \boldsymbol{b}_2$?

3. Przy założeniach z poprzedniego punktu i dodatkowo przy założeniu, że $\boldsymbol{\varepsilon}$ ma rozkład normalny, pokaż jaki rozkład ma $\boldsymbol{b}_1 - \boldsymbol{b}_2$.

4. Zaproponuj statystykę, która mogłaby posłużyć do przetestowaniu hipotezy o równości β w obu podpróbkach, jeśli znana jest wartość σ^2.

   Podpowiedź: Skorzystaj z tego, że dla wektora losowego $\boldsymbol{u}_{[g\times 1]} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma})$, $\boldsymbol{u}'\boldsymbol{\Sigma}^{-1}\,\boldsymbol{u} \sim \chi^2_g$

Rozwiązanie:

1. Estymator $\boldsymbol{b}_1 = \boldsymbol{\beta}_1 + \left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'\boldsymbol{\varepsilon}_1$ a $\boldsymbol{b}_2 = \boldsymbol{\beta}_2 + \left(\boldsymbol{X}_2'\boldsymbol{X}_2\right)^{-1}\boldsymbol{X}_2'\boldsymbol{\varepsilon}_2$, z $KMRL$ macierze $\boldsymbol{X}_1$ i $\boldsymbol{X}_2$ są deterministyczne a więc $\boldsymbol{b}_1$ i $\boldsymbol{b}_2$ są deterministycznymi funkcjami niezależnych zmiennych losowych $\boldsymbol{\varepsilon}_1$ i $\boldsymbol{\varepsilon}_2$ i w związku z tym też muszą być niezależne.

2. Na mocy nieobciążoności estymatora MNK przy spełnionych założeniach $KMRL$ wartość oczekiwana jest równa

$$\mathrm{E}(\boldsymbol{b}_1 - \boldsymbol{b}_2) = \mathrm{E}(\boldsymbol{b}_1) - \mathrm{E}(\boldsymbol{b}_2) = \boldsymbol{\beta} - \boldsymbol{\beta} = 0$$

   Wariancja jest równa

$$\begin{aligned}\mathrm{Var}(\boldsymbol{b}_1 - \boldsymbol{b}_2) &= \mathrm{Var}(\boldsymbol{b}_1) + \mathrm{Var}(\boldsymbol{b}_2) - 2\,\mathrm{Cov}(\boldsymbol{b}_1, \boldsymbol{b}_2)\\ &= \mathrm{Var}(\boldsymbol{b}_1) + \mathrm{Var}(\boldsymbol{b}_2) = \sigma^2\left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1} + \sigma^2\left(\boldsymbol{X}_2'\boldsymbol{X}_2\right)^{-1}\end{aligned}$$

   Przy czym skorzystaliśmy z tego, że kowariancja między zmiennymi niezależnymi jest równa zeru.

3. Jeśli $\boldsymbol{\varepsilon}$ ma rozkład normalny to $\boldsymbol{b}_1$ i $\boldsymbol{b}_2$ jako liniowe i deterministyczne funkcje $\boldsymbol{\varepsilon}$ mają także rozkład normalny. Różnica niezależnych zmiennych o rozkładzie normalnym ma także rozkład normalny:

$$\boldsymbol{b}_1 - \boldsymbol{b}_2 \sim N\left(\boldsymbol{0}, \sigma^2\left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1} + \sigma^2\left(\boldsymbol{X}_2'\boldsymbol{X}_2\right)^{-1}\right)$$

4. Podstawiając $\boldsymbol{u} = \boldsymbol{b}_1 - \boldsymbol{b}_2$ uzyskujemy statystykę

$$(\boldsymbol{b}_1 - \boldsymbol{b}_2)' \left[\sigma^2 \left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1} + \sigma^2 \left(\boldsymbol{X}_2'\boldsymbol{X}_2\right)^{-1}\right]^{-1} (\boldsymbol{b}_1 - \boldsymbol{b}_2) \sim \chi_g^2$$

gdzie g jest liczbą zmiennych w modelu.