Egzamin z ekonometrii IiE 07.02.2006
I semestr

Pytania teoretyczne

1. Wyjaśnić różnicę między parametrami i oszacowaniami parametrów oraz między odchyleniami losowymi i resztami

   Parametry są wielkościami deterministycznymi, które charakteryzują model. Oszacowania parametrów są fukcjami losowych obserwacji i w rezultacie same też są losowe. Uzyskane z różnych prób oszacowania parametrów będą z reguły różne. Na przykład KMRL $\boldsymbol{\beta}$ to parametr a estymator $\boldsymbol{b}$ to oszacowanie. Błędy losowe odpowiadają w modelu liniowym z część zmienności, której nie zależy od zmiennych niezależnych $\varepsilon_i = y_i - \boldsymbol{x}_i\boldsymbol{\beta}$. Reszty w modelu są różnicą między wartością zaobserwowaną a wartością dopasowaną z oszacowanego modelu $e_i = y_i - \boldsymbol{x}_i\boldsymbol{b}$. Reszty można traktować jako oszacowania błędów losowych.

2. Wyprowadzić estymator MNK dla modelu z wieloma zmiennymi objaśniającymi.

   Minimalizujemy sumę kwadrtów reszt:

$$\begin{aligned} S(\boldsymbol{b}) &= \boldsymbol{e}'\boldsymbol{e} = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}) \\ &= \boldsymbol{y}'\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{X}\boldsymbol{b} - \boldsymbol{b}'\boldsymbol{X}'\boldsymbol{y} + \boldsymbol{b}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b} \\ &= \boldsymbol{y}'\boldsymbol{y} - 2\boldsymbol{y}'\boldsymbol{X}\boldsymbol{b} + \boldsymbol{b}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b} \end{aligned}$$

$$\begin{aligned} \frac{\partial S(\boldsymbol{b})}{\partial \boldsymbol{b}} &= \frac{\partial \boldsymbol{y}'\boldsymbol{y}}{\partial \boldsymbol{b}} - 2\frac{\partial \boldsymbol{y}'\boldsymbol{X}\boldsymbol{b}}{\partial \boldsymbol{b}} + \frac{\partial \boldsymbol{b}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b}}{\partial \boldsymbol{b}} \\ &= -2\boldsymbol{X}'\boldsymbol{y} + 2\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b} \end{aligned}$$

   Warunki pierwszego rzędu na minimalizację sumy kwadratów reszt $S(\boldsymbol{b})$ uzyskujemy, przyrównując $\frac{\partial S(\boldsymbol{b})}{\partial \boldsymbol{b}}$ do zera:

$$-2\boldsymbol{X}'\boldsymbol{y} + 2\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b} = \boldsymbol{0}$$

   Rozwiązując dla $\boldsymbol{b}$

$$\boldsymbol{b} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{y}$$

3. Opisz sposób przeprowadzania testu $RESET$.

   Estymujemy model. Z modelu generujemy wartości dopasowane. Do modelu dodajemy potęgi wartości dopasowanych $\widehat{y}_i^2, \widehat{y}_i^3, \ldots$. Testujemy łączną nieistotność dodanych potęg wartości dopasowanych. W razie odrzucenia H_0 wnioskujemy, że forma funkcyjna jest niepoprawna.

   Alternatywnie do modelu dodajemy wszystkie kwadraty i iloczyny krzyżowe między zmiennymi objaśniającymi i testujemy ich łączną istotność,

4. Udowodnij, że estymator MNK jest estymatorem zgodnym. Wypisz założenia konieczne do tego dowodu.

   Jeśli

   $\mathrm{E}\left(\boldsymbol{x}_i'\varepsilon_i\right) = \boldsymbol{0}$, dla $1, \ldots, N$

   $\mathrm{E}\left(\boldsymbol{x}_i'\boldsymbol{x}_i\right) = \boldsymbol{\Sigma}_X$ jest odwracalna dla $1, \ldots, N$

   $\operatorname{plim}\left(\frac{1}{N}\sum_{i=1}^N \boldsymbol{x}_i'\varepsilon_i\right) = \mathrm{E}\left(\boldsymbol{x}_i'\varepsilon_i\right)$

   $\operatorname{plim}\left(\frac{1}{N}\sum_{i=1}^N \boldsymbol{x}_i'\boldsymbol{x}_i\right) = \mathrm{E}\left(\boldsymbol{x}_i'\boldsymbol{x}_i\right),$

to estymator MNK jest zgodny, ponieważ:

$$\begin{aligned} \operatorname{plim}(\boldsymbol{b}) &= \operatorname{plim}\left[\beta + \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon}\right] \\ &= \boldsymbol{\beta} + \operatorname{plim}\left[\left(\frac{1}{N}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\frac{1}{N}\boldsymbol{X}'\boldsymbol{\varepsilon}\right)\right] \\ &= \boldsymbol{\beta} + \underbrace{\left[\operatorname{plim}\left(\frac{1}{N}\sum_{i=1}^N \boldsymbol{x}_i'\boldsymbol{x}_i\right)\right]^{-1}}_{\left[\mathrm{E}\left(\boldsymbol{x}_i'\boldsymbol{x}_i\right)\right]^{-1}}\underbrace{\operatorname{plim}\left(\frac{1}{N}\sum_{i=1}^N \boldsymbol{x}_i'\boldsymbol{\varepsilon}_i\right)}_{\boldsymbol{0}} = \boldsymbol{\beta} \end{aligned}$$

ZADANIE 1 Uzyskaliśmy MNK na podstawie dwóch niezależnych próbek oszacowania $\boldsymbol{b}_1$ i $\boldsymbol{b}_2$ parametru $\boldsymbol{\beta}$ w modelu. Zakładamy, że w obu próbkach spełnione są założenia $KMRL$.

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

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 $\boldsymbol{b}_1 - \boldsymbol{b}_2$.
4. Zaproponuj statystykę, która mogłaby posłużyć do przetestowaniu hipotezy o równości β w obu podpróbkach jeśli znane jest wartość σ^2.

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

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:
$$\boldsymbol{b}_1 - \boldsymbol{b}_2 \sim N\left(\mathbf{0}, \sigma^2 \left(\boldsymbol{X}_1' \boldsymbol{X}_1\right)^{-1} + \sigma^2 \left(\boldsymbol{X}_2' \boldsymbol{X}_2\right)^{-1}\right)$$
4. Podstawiając $\boldsymbol{u} = \boldsymbol{b}_1 - \boldsymbol{b}_2$ uzyskujemy statystykę
$$(\boldsymbol{b}_1 - \boldsymbol{b}_2)' \left[\sigma^2 \left(\boldsymbol{X}_1' \boldsymbol{X}_1\right)^{-1} + \sigma^2 \left(\boldsymbol{X}_2' \boldsymbol{X}_2\right)^{-1}\right]^{-1} (\boldsymbol{b}_1 - \boldsymbol{b}_2) \sim \chi^2_g$$
   gdzie g jest liczbą zmiennych w modelu.

ZADANIE 2 Mamy następujący model
$$y_i = \exp(\boldsymbol{x}_i \boldsymbol{\alpha})\, \varepsilon_i$$
i zakładamy, że $\mathrm{E}\left(\ln \varepsilon_i^2\right) = \gamma$ oraz $\boldsymbol{x}_i$ jest nielosowe i zawiera stałą.

1. Pokazać, że estymator $\widehat{\boldsymbol{\alpha}} = \frac{1}{2}\boldsymbol{b}$, gdzie $\boldsymbol{b}$ jest wektorem oszacowań w regresji $\ln\left(y_i^2\right)$ na $\boldsymbol{x}_i$, będzie nieobciążonym estymatorem $\boldsymbol{\alpha}$ dla wszystkich parametrów poza parametrem α_0 stającym przy stałej.

   Podpowiedź: Zastanów się czemu jest równe $\mathrm{E}(y^*)$, gdzie $y^* = \left[\ln y_1^2, \ldots, \ln y_N^2\right]'$.
2. Wyjaśnić na czym polega różnica między tym modelem a modelem ze zlogarytmowaną zmienną zależną szacowananym MNK.

   Podpowiedź: Jakie założenie musimy przyjąć na temat ε_i w modelu ze zlogarytmowaną zmienną zależną aby estymator MNK był nieobciążony? Jakie wartości mogą przyjmować w tym modelu y_i i ε_i?

Rozwiązanie:

1.
$$y_i^2 = \exp(\boldsymbol{x}_i\boldsymbol{\alpha})\,\varepsilon_i^2$$
$$\ln y_i^2 = 2\boldsymbol{x}_i\boldsymbol{\alpha} + \ln\left(\varepsilon_i^2\right)$$
$$\mathrm{E}\left(\ln y_i^2\right) = 2\boldsymbol{x}_i\,\boldsymbol{\alpha} + \mathrm{E}\left[\ln\left(\varepsilon_i^2\right)\right] = 2\boldsymbol{x}_i\,\boldsymbol{\alpha} + \gamma = 2\boldsymbol{x}_i\boldsymbol{\alpha}^*$$
   gdzie $\alpha^* = \left[\alpha_0 + \frac{1}{2}\gamma, \alpha_1, \ldots, \alpha_K\right]$. Wynika z tego, że
$$\mathrm{E}\left(\boldsymbol{y}^*\right) = 2\boldsymbol{X}\boldsymbol{\alpha}^*$$
   gdzie $\boldsymbol{y}^* = \left[\ln y_1^2, \ldots, \ln y_N^2\right]'$ wartość oczekiwana $\widehat{\boldsymbol{\alpha}}$
$$\mathrm{E}\left(\widehat{\boldsymbol{\alpha}}\right) = \mathrm{E}\left(\frac{1}{2}\boldsymbol{b}\right) = \frac{1}{2}\,\mathrm{E}\left[\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{y}^*\right] = \frac{1}{2}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\,\mathrm{E}\left(\boldsymbol{y}^*\right)$$
$$= \frac{1}{2}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\left(2\boldsymbol{X}\boldsymbol{\alpha}\right) = \boldsymbol{\alpha}^*$$
$$\mathrm{E}\left(\widehat{\boldsymbol{\alpha}}\right) - \boldsymbol{\alpha} = \left[\frac{1}{2}\gamma, 0, \ldots, 0\right]$$

2. W standardowym modelu na logarytmach zakładamy, że $y_i > 0$ i $\varepsilon_i > 0$. W modelu analizowanym założenie to zostało pominięte. W modelu na logarytmach zamiast założenia $\mathrm{E}\left(\ln\varepsilon_i^2\right) = \gamma$ robimy założenie, że $\mathrm{E}\left(\ln\varepsilon_i\right) = 0$.

ZADANIE 3 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     .440    -1.34   0.192
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) =5.894179 [.0525]
White's general test statistic:  Chi-sq(14) =25.15268 [.0331]
Jarque-Berra test statistic:     Chi-sq(2)  =7.34     [.0094]
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 i $x2$ są istotne na zadanym poziomie istotności (*wartości p* 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 ($d_L = 0.917, d_U = 1.597$). Podobnie na podstawie testu Breuscha-Godfreya $[.0525 > 0.01]$ nie jesteśmy w stanie odrzucić hipotezy o braku autokorelacji.
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 nie ma rozkładu normalnego, o czym świadczy *p-value* dla testu Jarque'a-Berra $[.0094 < 0.01]$.
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. Hipoteze o łącznej nieistotności zmiennych sezonowych należy testestować 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 o normalności składnika losowego.
    (b) W przypadku małej próby może to powodować, że rozkłady testów będą niewłaściwe.
    (c) Jedynym sposobem poradzenia sobie z tym problemem jest powiększenie próby, ponieważ dla większej próby rozkłady są bliższe znanym rozkładom asymptotycznym.