Egzamin z ekonometrii 13.06.2006
I semestr

Pytania teoretyczne

1. Zapisać model liniowy. Podać interpretację poszczególnych elementów tego modelu.

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

gdzie y_i jest zmienną objaśnianą, $\boldsymbol{x}_i$ wektorem zmiennych objaśniających, $\boldsymbol{\beta}$ nieznanym wektorem parametrów, ε_i zaburzeniem losowym

2. Dlaczego nie da się uzyskać oszacowań MNK jeśli liczba zmiennych objaśniających w modelu jest większa od liczby obserwacji?

Ponieważ w takim przypadku macierz $\boldsymbol{X}'\boldsymbol{X}$ jest macierzą osobliwą i nie da się znaleźć wielkości $\boldsymbol{b} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{y}$

3. Wyprowadzić postać macierzy wariancji kowariancji $\boldsymbol{b}$ i podać interpretację jej elementów

Wyprowadzenie:

$$\begin{aligned}\operatorname{Var}(\boldsymbol{b}) &= \operatorname{Var}\left[\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{y}\right] = \operatorname{Var}\left[\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\left(\boldsymbol{X}'\boldsymbol{\beta}+\boldsymbol{\varepsilon}\right)\right] = \operatorname{Var}\left[\boldsymbol{\beta}+\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon}\right] \\ &= \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\underbrace{\operatorname{Var}(\boldsymbol{\varepsilon})}_{\sigma^2\boldsymbol{I}}\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} = \sigma^2\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\end{aligned}$$

Elementy na przekątnej macierzy $\operatorname{Var}(\boldsymbol{b})$ to wariancje elementów wektora $\boldsymbol{b}$ - $\operatorname{Var}(b_r)$, a elementy poza przekątną to kowariancje między elementami wektora $\boldsymbol{b}$ - $\operatorname{Cov}(b_r, b_s)$.

4. Pokazać, w jaki sposób można, dla znanej macierzy $\boldsymbol{\Omega}$, sprowadzić model z niesferycznymi błędami losowymi do modelu spełniającego założenia $KMRL$.

Model jest liniowy

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u},$$

$\operatorname{Var}(\boldsymbol{u}) = \boldsymbol{\Omega} = \sigma^2\boldsymbol{V}$, $\boldsymbol{V}$ jest znana. Definiujemy $\boldsymbol{L}$ takie, że $\boldsymbol{L}\boldsymbol{V}\boldsymbol{L}' = \boldsymbol{I}$

$$\boldsymbol{L}\boldsymbol{y} = \boldsymbol{L}\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{L}\boldsymbol{u}$$

lub

$$\boldsymbol{y}^* = \boldsymbol{X}^*\boldsymbol{\beta} + \boldsymbol{u}^*$$

Wariancję błędu losowego $\boldsymbol{u}^*$ dla modelu przekształconego:

$$\operatorname{Var}(\boldsymbol{u}^*) = \operatorname{Var}(\boldsymbol{L}\boldsymbol{u}) = \sigma^2\boldsymbol{L}\boldsymbol{V}\boldsymbol{L}' = \sigma^2\boldsymbol{I}$$

ZADANIE 1 Rozwiązać poniższe problemy:

1. Oszacowano model: $y_i = \beta_1 + \beta_2 x_{1i} + \epsilon_i$. Obliczyć nieznane wartości reszt (e_3, e_4), jeżeli wiadomo, że $\boldsymbol{x}_1' = [2,3,1,2]$, a $\boldsymbol{e}' = [2,1,e_3,e_4]$.

2. Oszacowano model $y_i = \beta_0 + \beta_1 x_{1i} + \varepsilon_i$ dla $\boldsymbol{y}' = [2,3,5,3,2,4]$ i uzyskano wartości dopasowane $\widehat{\boldsymbol{y}}' = [2,2,6,2,1,\widehat{y}_6]$. Znaleźć $\widehat{y}_6$.

3. Pokazać, że w modelu $y_i = \beta + \varepsilon_i$ suma kwadratów reszt $\boldsymbol{e}'\,\boldsymbol{e} = \sum_{i=1}^{N}(y_i - \overline{y})^2$

Rozwiązanie:

1. Korzystamy z tego, że w modelu ze stałą $\sum e_i = 0$ i $\boldsymbol{X}'\,\boldsymbol{e} = \boldsymbol{0} \Longrightarrow \sum e_i x_{1i} = 0$. Wektor reszt będzie postaci: $\boldsymbol{e}' = [2,-1,-5,2]$.

2. Do rozwiązania zadania należy wykorzystać następującą właściwość hiperpłaszczyzny regresji - $\sum y_i = \sum \hat{y}_i$. Zatem $\hat{y}_6 = 6$.

3. W modelu, w którym występuje tylko stała $\boldsymbol{X} = \boldsymbol{l}$ a $\boldsymbol{b} = \left(\boldsymbol{l}'\boldsymbol{l}\right)^{-1} \boldsymbol{l}'\boldsymbol{y} = \frac{1}{N}\sum_{i=1}^{N} y_i = \overline{y}$. Z kolei

$$e_i = y_i - x_i\, \boldsymbol{b} = y_i - \overline{y}$$

$$\boldsymbol{e}'\boldsymbol{e} = \sum_{i=1}^{N} (y_i - \overline{y})^2$$

ZADANIE 2 Procentowa różnica między średnią płacą kobiet i mężczyzn jest dana wzorem:

$$G = \frac{\overline{W}^m - \overline{W}^f}{\overline{W}^f}$$

gdzie $\overline{W}^f = \exp\left[\frac{1}{N}\sum_{i=1}^{N} \ln\left(W_i^f\right)\right] = \sqrt[N]{\prod_{i=1}^{N} W_i^f}$ jest średnią geometryczną płac dla kobiet. Analogicznie definiujemy $\overline{W}^m$.

Przekształcając wyjściowy wzór na G otrzymujemy:

$$\ln(G+1) = \ln\left(\overline{W}^m\right) - \ln\left(\overline{W}^f\right)$$

Załóżmy, że płace determinowane są w następujący sposób

$$\ln\left(W_i^f\right) = \boldsymbol{x}_i^f \boldsymbol{\beta}^f + \varepsilon_i^f$$
$$\ln\left(W_i^m\right) = \boldsymbol{x}_i^m \boldsymbol{\beta}^m + \varepsilon_i^m$$

gdzie $\boldsymbol{x}_i^f, \boldsymbol{x}_i^m$ są wierszowymi wektorami charakterystyk jednostek a ε_i^f i ε_i^m błedami losowymi. Oczekiwaną wartość $\ln(1+G)$ można wyrazić za pomocą wielkości parametrów w obu modelach i wektorów średnich charakterystyk $\overline{\boldsymbol{x}}^f$ i $\overline{\boldsymbol{x}}^m$

$$\mathrm{E}\left[\ln(1+G)\right] = \overline{\boldsymbol{x}}^f \boldsymbol{\beta}^f - \overline{\boldsymbol{x}}^m \boldsymbol{\beta}^m = \overline{\boldsymbol{x}}^f \boldsymbol{\beta}^f - \overline{\boldsymbol{x}}^m \boldsymbol{\beta}^f + \left(\overline{\boldsymbol{x}}^m \boldsymbol{\beta}^f - \overline{\boldsymbol{x}}^m \boldsymbol{\beta}^m\right) = \left(\overline{\boldsymbol{x}}^f - \overline{\boldsymbol{x}}^m\right) \boldsymbol{\beta}^f + \overline{\boldsymbol{x}}^m \left(\boldsymbol{\beta}^f - \boldsymbol{\beta}^m\right)$$

Ronald Oaxaca (1973) zaproponował, by wielkość $d_x = \left(\overline{\boldsymbol{x}}^f - \overline{\boldsymbol{x}}^m\right) \boldsymbol{\beta}^f$ interpretować jako różnicę w płacach wynikającą ze z różnic w charakterystykach kobiet i mężczyzn a wielkość $d_\beta = \overline{\boldsymbol{x}}^m \left(\boldsymbol{\beta}^f - \boldsymbol{\beta}^m\right)$ traktować jako miarę dyskryminacji.

Zlogarytmowana płaca została w poniższych modelach uzależniona od wykształcenia w latach i wieku respondentów. Dla kobiet i mężczyzn regresje MNK oszacowano osobno. Zakładamy, że próby dla mężczyzn i kobiet są niezależne.

Regresja dla mężczyzn

$\ln$ (**płaca**)	**Współczynnik**	**Bł. Std**
wykształcenie	0.313	0.008
wiek	0.060	0.002

Regresja dla kobiet

$\ln$ (**płaca**)	**Współczynnik**	**Bł. Std**
wykształcenie	0.252	0.007
wiek	0.067	0.002

Średnie charakterystki dla kobiet i mężczyzn:

płeć	wykształcenie	wiek
mężczyzna	11.191	39.040
kobieta	11.764	39.827

1. Znaleźć oszacowania $\widehat{d}_x$ i $\widehat{d}_\beta$
2. Oszacować współczynnik zróżnicowania płac $\mathrm{E}\left[\ln(G+1)\right]$
3. Zintepretować uzyskane wyniki. **Podpowiedź**: Dla małych x, $\ln(x+1) = x$

   Załóżmy, że znamy oszacowanie wariancji oszacowań wektorów parametrów uzyskanych z MNK: $\mathrm{Var}\left(\widehat{\boldsymbol{b}}^f\right) = \widehat{\boldsymbol{\Sigma}}^f$ i $\mathrm{Var}\left(\widehat{\boldsymbol{b}}^m\right) = \widehat{\boldsymbol{\Sigma}}^m$

4. Policzyć wariancję:

   (a) oszacowania $\widehat{d}_x$

   (b) oszacowania $\widehat{d}_\beta$

5. Policzyć wariancję oszacowania $\mathrm{E}\,[\widehat{\ln(G+1)}]$

Rozwiązanie:

1. Z podanego wzoru współczynnik zróżnicowania płac jest równy:

   $\widehat{d}_x = (11.764 - 11.191)\,0.252 + (39.827 - 39.040)\,0.067 = 0.197$

   $\widehat{d}_\beta = 11.191\,(0.252 - 0.313) + 39.040\,(0.067 - 0.060) = -0.409$

2. $\ln(G+1) = 0.197 - 0.409 = -0.212$

3. Całkowita różnica w płacach wynosi 21.12%. Z powodu dyskryminacji kobiety płaca kobiet jest o 40.9% niższa, ze zróżnicowania charakterystyk wynikałoby, że kobiety powinny zarabiać 19.7% więcej niż mężczyźni.

4. Wariancję oszacowań można $\widehat{d}_x$ liczymy ze wzoru na wariancję prognozy:

$$\mathrm{Var}\left(\widehat{d}_x\right) = \mathrm{Var}\left[\left(\overline{\boldsymbol{x}}^f - \overline{\boldsymbol{x}}^m\right)\boldsymbol{b}^f\right] = \left(\overline{\boldsymbol{x}}^f - \overline{\boldsymbol{x}}^m\right)\mathrm{Var}\left(\boldsymbol{b}^f\right)\left(\overline{\boldsymbol{x}}^f - \overline{\boldsymbol{x}}^m\right)'$$

   oszacowaniem tej wielkości będzie $\left(\overline{\boldsymbol{x}}^f - \overline{\boldsymbol{x}}^m\right)\widehat{\boldsymbol{\Sigma}}^f\left(\overline{\boldsymbol{x}}^f - \overline{\boldsymbol{x}}^m\right)'$

5. Wariancja oszacowania $\widehat{d}_\beta$ znajdujemy z ogólych własności wariancji

$$\mathrm{Var}\left(\widehat{d}_\beta\right) = \mathrm{Var}\left[\overline{\boldsymbol{x}}^m\left(\boldsymbol{b}^f - \boldsymbol{b}^m\right)\right] = \overline{\boldsymbol{x}}^m\,\mathrm{Var}\left(\boldsymbol{b}^f - \boldsymbol{b}^m\right)\overline{\boldsymbol{x}}^{m\prime}$$

   Z kolei

$$\mathrm{Var}\left(\boldsymbol{b}^f - \boldsymbol{b}^m\right) = \mathrm{Var}\left(\boldsymbol{b}^f\right) + \mathrm{Var}\left(\boldsymbol{b}^m\right) - \underbrace{2\mathrm{Cov}\left(\boldsymbol{b}^f, \boldsymbol{b}^m\right)}_{0} = \mathrm{Var}\left(\boldsymbol{b}^f\right) + \mathrm{Var}\left(\boldsymbol{b}^m\right)$$

   $\mathrm{Cov}\left(\boldsymbol{b}^f, \boldsymbol{b}^m\right) = 0$, ponieważ błędy oszacowań uzyskanych z dwóch niezależnych prób są też niezależne

   W rezultacie oszacowaniem wariancji jest $\overline{\boldsymbol{x}}^m\left(\widehat{\boldsymbol{\Sigma}}^f + \widehat{\boldsymbol{\Sigma}}^m\right)\overline{\boldsymbol{x}}^{m\prime}$

6. Wariancja oszacowania jest równa:

$$\begin{aligned}
\mathrm{Var}\left(\mathrm{E}\,[\widehat{\ln(G+1)}]\right) &= \mathrm{Var}\left(\overline{\boldsymbol{x}}^f\boldsymbol{b}^f - \overline{\boldsymbol{x}}^m\boldsymbol{b}^m\right)\\
&= \overline{\boldsymbol{x}}^f\,\mathrm{Var}\left(\boldsymbol{b}^f\right)\overline{\boldsymbol{x}}^{f\prime} + \overline{\boldsymbol{x}}^m\,\mathrm{Var}\left(\boldsymbol{b}^m\right)\overline{\boldsymbol{x}}^{m\prime} - \overline{\boldsymbol{x}}^f\underbrace{\mathrm{Cov}\left(\boldsymbol{b}^f, \boldsymbol{b}^m\right)}_{0}\overline{\boldsymbol{x}}^{m\prime}\\
&= \overline{\boldsymbol{x}}^f\,\mathrm{Var}\left(\boldsymbol{b}^f\right)\overline{\boldsymbol{x}}^{f\prime} + \overline{\boldsymbol{x}}^m\,\mathrm{Var}\left(\boldsymbol{b}^m\right)\overline{\boldsymbol{x}}^{m\prime}
\end{aligned}$$

   Oszacowaniem wariancji estymatora $\mathrm{E}\,[\widehat{\ln(G+1)}]$ będzie więc $\overline{\boldsymbol{x}}^f\widehat{\boldsymbol{\Sigma}}^f\overline{\boldsymbol{x}}^{f\prime} + \overline{\boldsymbol{x}}^m\widehat{\boldsymbol{\Sigma}}^m\overline{\boldsymbol{x}}^{m\prime}$

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    .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) =4.894179 [.0867]
Breusch-Pagan test statistic:    Chi-sq(2)  =10.41    [.0054]
White's general test statistic:  Chi-sq(14) =35.15268 [.0013]
Jarque-Berra test statistic:     Chi-sq(2)  =8.34     [.0154]
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_2.
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 ($d_L = 0.917, d_U = 1.597$).Także *p-value* dla testu Breuscha-Godfreya świadczy o tym, że w modelu nie ma autokorelacji $[.0867 > 0.01]$.

4. Odrzucamy hipotezę o homoskedastyczności składnika losowego w związku z wartością *p-value* dla testu White'a $[.0013 < 0.01]$.

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-Berra $[.1886 > 0.01]$.

7. Parametry modelu są stabilne, o czym świadczy *p-value* dla testu Chowa $[.8359 > 0.01]$.

8. Współczynnik przy zmiennej x_2 jest elastycznością zagregowanej konsumpcji względem deflatora, zatem wzrost deflatora o 1% spowoduje spadek zagregowanej konsumpcji o -0.59%.

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 braku heteroskedastyczności składnika losowgo

    (b) Macierz wariancji-kowariancji jest niewłaściwa, co powoduje, że niewłaściwe są błędy standardowe parametrów i tym samym niewłaściwe są statystyki t.

    (c) Jeśli nie chcemy usunąć heteroskedastyczności, należy posłużyć się tzw. odporną macierzą wariancji-kowariancji White'a. Można też podjąć próbę usunięcia heteroskedastyczności za pomocą Stosowalnej UMNK.