# Egzamin z ekonometrii 12.09.2006
# I semestr

**Pytania teoretyczne**

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

   Parametry są nielosowe ale nieobserwowalne. Oszacowanie parametrów są funkcjami obserwowalnych $y_i$, $\boldsymbol{x}_i$ a więc są obserwowalne ale z reguły są losowe. Na przykład, w $KMRL$ o wektorze parametrów $\boldsymbol{\beta}$ zakładamy, że jest nielosowy i nieobserwowalny ale oszacowanie tego parametru

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

   jest losowe, ponieważ jest funkcją błędów losowych $\boldsymbol{\varepsilon}$.

   Błędy losowe w $KMRL$ odpowiadają, za losową niewyjaśnioną część zmienności $y_i$

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

   Reszty stanowią oszacowanie błędów losowych i liczymy je jako różnice między dopasowanymi $y_i$ i zaobserwowanymi $y_i$:

   $$e_i = y_i - \widehat{y}_i = y_i - \boldsymbol{x}_i\boldsymbol{b}$$

2. Udowodnić, że w $KMRL$ estymator $\boldsymbol{b}$ jest nieobciążony.

   $$\mathrm{E}\left[\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{y}\right] = \mathrm{E}\left[\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\left(\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}\right)\right] = \boldsymbol{\beta} + \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\underbrace{\mathrm{E}\left(\boldsymbol{\varepsilon}\right)}_{\boldsymbol{0}} = \boldsymbol{\beta}$$

3. Dlaczego w modelu nie powinno się umieszczać jednocześnie stałej i wszystkich zmiennych zerojedynkowych związanych z poziomami zmiennej dyskretnej? Odpowiedź uzasadnić.

   Jeśli zmienna dyskretna ma $s$ poziomów, to jedna i tylko jednak z $s$ utworzonych na jej podstawie zmiennych zerojedynkowych $D_i$ przyjmuje wartość 1 a pozostałe przyjmują wartość 0. Wynika z tego, że $\sum_i^s D_i = 1$. Istnieje więc taka kombinacja liniowa $D_i$, która daje kolumnę jedynek. Oznacza to, że zbiór zawierający wszystkie zmienne dyskretne oraz stałą jest współliniowy.

4. Do czego służą testy diagnostyczne?

   Testy diagnostyczne służą do testowania prawdziwości założeń modelu. Przykładowo w przypadku $KMRL$ testy diagnostyczne mogą być wykorzystane do tetowania prawdziwości założeń o homoskedastyczności, braku autokorelacji, czy normalności rozkładu błędu losowego.

ZADANIE 1

1. Jakie wymiary mają macierze $(\boldsymbol{X'X})^{-1}$ i $\boldsymbol{X'y}$, jeżeli szacowany jest KMRL ze stałą i z dziesięcioma zmiennymi objaśniającymi, a liczba obserwacji wynosi 120?

2. Zapisać pierwszy wiersz macierzy $\boldsymbol{X'X}$ dla modelu ze stałą?

3. Na podstawie próby zawierającej $s$ obserwacji wyestymowano model:

   $$y = \alpha_1 + \alpha_2 x_2 + \alpha_3 x_3 + ... + \alpha_s x_s + \varepsilon$$

   Ile wyniesie współczynnik $R^2$ dla tego modelu?

*Rozwiązanie*:

1. Macierze $(\boldsymbol{X'X})^{-1}$ i $\boldsymbol{X'y}$ będą wymiarów, odpowiednio: $11 \times 11$ i $11 \times 1$.

2. Pierwsza kolumna macierzy $\boldsymbol{X}$ składa się z 1, więc pierwszy wiersz macierzy $\boldsymbol{X'X}$ będzie miał elementy równe sumom elementów kolumn macierzy $\boldsymbol{X}$.

3. Współczynnik $R^2$ tego modelu jest równy 1, ponieważ wszystkie punkty/obserwacje leżą na hiperpłaszczyźnie regresji (np. dla jednego regresora liczba obserwacji wynosi 2). Ponadto, ponieważ liczba obserwacji ($s$) jest równa liczbie nieznanych parametrów modelu, liczba stopni swobody wynosi 0. Powoduje to że nie można oszacować błędów standardowych szukanych parametrów.

ZADANIE 2 Na podstawie danych pochodzących ze strony `www.economagic.com` zbudowano KMRL wyjaśniający kurs dolara amerykańskiego w stosunku do euro za pomocą podaży pieniądza w USA i indeksu cen dóbr konsumpcyjnych w USA (inflacji): $kurs = stala + \beta_1 podaz + \beta_2 cpi + \varepsilon$. Model bazuje na 108 obserwacjach od stycznia 1990 do stycznia 1999. Otrzymano następujące wyniki:

```
      Source |       SS       df       MS              Number of obs =     108
-------------+------------------------------           F(  2,   105) =   13.06
       Model |  .136566483     2  .068283242           Prob > F      =  0.0000
    Residual |  .548871572   105  .005227348           R-squared     =  0.1992
-------------+------------------------------           Adj R-squared =  0.1840
       Total |  .685438055   107  .006405963           Root MSE      =   .0723
------------------------------------------------------------------------------
        kurs |      Coef.   Std. Err.      t    P>|t|  [95% Conf. Interval]
-------------+----------------------------------------------------------------
       podaz |   .0000863   .0000958     0.90   0.370   -.0001037   .0002764
         cpi |  -.0040808   .0010651    -3.83   0.000   -.0061927   -.001969
       _cons |    1.73358   .1012719    17.12   0.000    1.532776   1.934383
------------------------------------------------------------------------------
Durbin-Watson d-statistic(  3,   108) =  .1814083
Breusch-Pagan chi2(1)                 =   1.61  Prob > chi2   = 0.2049
Ramsey RESET  F(3, 102)               = 15.09  Prob > F      = 0.0000
Jarque-Bera chi2(2)                   =   6.91  Prob > chi2   = 0.0315
```

oraz macierz wariancji i kowariancji estymatorów:

```
             |   podaz      cpi    _cons
-------------+---------------------------
       podaz |  9.2e-09
         cpi | -8.0e-08  1.1e-06
       _cons |  2.4e-06 -.000086  .010256
```

Na podstawie powyższych wyników odpowiedzieć na następujące pytania. Każdą odpowiedź uzasadnić wielkościami odpowiednich statystyk testowych. Testy przeprowadzić na poziomie istotności $\alpha = 0.05$.
**Podpowiedź**: wartości $d_L$ i $d_U$ dla testu $DW$ przy 108 obserwacjach, 2 zmiennych i stałej oraz $\alpha = 0.05$ wynoszą $d_L = 1.63, d_U = 1.72$

1. Sprawdzić istotność oraz łączną istotność uzyskanych wyników.
2. Zinterpretować parametry modelu.
3. Sprawdzić, czy w modelu występuje heteroskedastyczność.
4. Sprawdzić, czy w modelu występuje autokorelacja.
5. Sprawdzić poprawność formy funkcyjnej modelu.
6. Sprawdzić, czy składnik losowy ma rozkład normalny.
7. Korzystając z wyników poprzednich podpunktów napisać, jakiej procedury estymacyjnej należy użyć, by otrzymać zgodny estymator macierzy wariancji?

*Rozwiązanie*:

1. Statystyka $F(3, 105) = 13.06$ oraz jej *p-value*≈0 wskazuje na łączną istotność zmiennych w modelu. Pojedynczo istotna jest $stala$, ponieważ statystyka $|t_{stala}| = 17.12 > 1,96$ $[0.000 < 0.05]$ oraz $cpi$ ponieważ $|t_{CPI}| = 3,83 > 1,96$ $[0.000 < 0.05]$, natomiast nieistotna jest $podaz$ ponieważ *p-value* statytyki $t$ dla tej zmiennej jest większe od 0.05 $[0.370 > 0,05]$ .

2. Wzrost podaży pieniądza o jednostkę powoduje wzrost kursu dolara w stosunku do euro o $0.00008$. Wzrost wskaźnika cen konsumpcyjnych o jednostkę powoduje spadek kursu o $0.004$.

3. Na podstawie testu Breuscha-Pagana nie ma podstaw do odrzucenia hipotezy o homoscedastyczności ponieważ *p-value* statystyki testowej wynosi $0.20$ i jest większe od $0.05$.

4. W modelu występuje autokorelacja, ponieważ statystyka Durbina-Watsona jest bliska 0. Dolna wartość krytyczna odczytana z tablic dla 2 zmiennych i 100 obserwacji wynosi $1.63$. Wobec tego $DW < d_L$.

5. Na podstawie testu RESET $F(3, 105) = 15,09$, widzimy, że statystyka testowa jest duża, więc odrzucamy hipotezę zerową o poprawności formy funkcyjnej modelu.

6. Na podstawie statystyki Jarque-Berra równej $6,91$ odrzucamy hipotezę zerową o normalności składnika losowego $[0.0315 < 0.05]$.

7. Ponieważ w modelu występuje autokorelacja pierwszego rzędu i nie ma hetroscedastyczności zastosować estymator odporny Newey'a-Westa.Można też podjąć próbę usunięcia autokorelacji za pomocą Stosowalnej UMNK.

ZADANIE 3 Wyestymowano MNK następujący model dla wysokości płacy:

$$\ln\left(\text{płaca}_i\right) = \beta_0 + \beta_1 \text{staż}_i + \beta_2 \text{płeć}_i + \beta_3 \text{płeć}_i \times \text{staż}_i + \varepsilon_i \qquad (*)$$

gdzie zmienna płeć koduje płeć respondenta (0 mężczyzna, 1 kobieta), a zmienna staż oznacza staż pracy ogółem. Uzyskano następujące oszacowania parametrów:

| | **Współczynnik** | **Błąd Std.** | **t** | $\Pr(\lvert t\rvert > t^*)$ |
|---|---|---|---|---|
| **płeć** | -.168465 | .0265421 | -6.347 | 0.000 |
| **staż** | .010842 | .0008269 | 13.111 | 0.000 |
| **płeć×staż** | .000430 | .0012526 | .343 | 0.731 |
| **stała** | 7.457208 | .0180958 | 412.096 | 0.000 |

1. Zinterpretować wielkości oszacowań parametrów w tym modelu (także nieistotnych).

2. Wyjaśnić z jakiego powodu w modelu umieszczono zmienną płeć×staż.

3. Pokazać, że otrzymane w tej regresji oszacowania parametrów $\beta_0$, $\beta_1$, $\beta_2$,$\beta_3$ można również uzyskać przeprowadzając osobno regresję dla kobiet i dla mężczyzn.

   **Podpowiedź:** skorzystaj z tego, że model (*) można równoważnie zapisać jako:

   $$\ln\left(\text{płaca}_i\right) = \beta_0\left(1 - \text{płeć}_i\right) + \beta_2\left(1 - \text{płeć}_i\right) \times \text{staż}_i + \left(\beta_0 + \beta_1\right)\text{płeć}_i + \left(\beta_2 + \beta_3\right)\text{płeć}_i \times \text{staż}_i + \varepsilon_i \quad (**)$$

   oraz z tego, że dla modelu (**), przy odpowiednim uporządkowaniu obserwacji, macierzy $\boldsymbol{X}$ jest macierzą blokowo diagonalną.

*Rozwiązanie*:

1. Współczynnik przy płci ozacza, że kobiety zarabiają średnio o $16.8\%$ mniej niż mężczyźni, współczynnik przy stażu oznacza, że mężczyźni zarabiają o $1\%$ więcej z każdym dodatkowym rokiem stażu, współczynnik przy płeć×staż oznacza, że kobiety zyskują o $0.043\%$ punktu procentowego więcej na dodatkowym roku stażu niż mężczyźni

2. Zmienną płeć×staż umieszczono po to, by uwzględnić w modelu różnicę między wzrostem płacy związanej ze wzrostem stażu w przypadku mężczyzn i kobiet.

3. Zmienna płeć jest zmienną zerojedynkową. Oznacza to, że jeśli płeć$_i$ przyjmuje wartość $1$, to $1-$płeć$_i$ przyjmuje wartość $0$. Jeśli uporządkujemy obserwacje według płci to macierz obserwacji $\boldsymbol{X}$ dla modelu

postaci (**) przyjmie postać:

$$
\boldsymbol{X} = \begin{array}{cccc}
\hline
\text{płeć}_i & \text{staż}_i & {}_i1-\text{płeć} & (1-\text{płeć}_i)\times\text{staż}_i \\
\hline
1 & \text{staż} & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots \\
1 & \text{staż} & 0 & 0 \\
0 & 0 & 1 & \text{staż} \\
\vdots & \vdots & \vdots & \vdots \\
0 & 0 & 1 & \text{staż} \\
\hline
\end{array} = \begin{bmatrix} \boldsymbol{X}_1 & 0 \\ 0 & \boldsymbol{X}_2 \end{bmatrix}
$$

a więc estymator $MNK$ w modelu (*) można zapisać jako:

$$
\begin{aligned}
\boldsymbol{b} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\,\boldsymbol{y} &= \begin{bmatrix} \left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1} & 0 \\ 0 & \left(\boldsymbol{X}_2\boldsymbol{X}_2\right)^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_1'\boldsymbol{y}_1 \\ \boldsymbol{X}_1'\boldsymbol{y}_2 \end{bmatrix} \\
&= \begin{bmatrix} \left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'\boldsymbol{y}_1 \\ \left(\boldsymbol{X}_2'\boldsymbol{X}_2\right)^{-1}\boldsymbol{X}_2'\boldsymbol{y}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \end{bmatrix}
\end{aligned}
$$

Estymatory $\boldsymbol{b}_1$ i $\boldsymbol{b}_2$ uzyskujemy z regresji logarytmu płacy na stałej i stażu osobno na podpróbce dla kobiet i mężczyzn. Oszacowania równe oszacowaniom w orginalnym modelu można uzyskać wykorzystując relacja między parametrami w modelu (*) i (**)