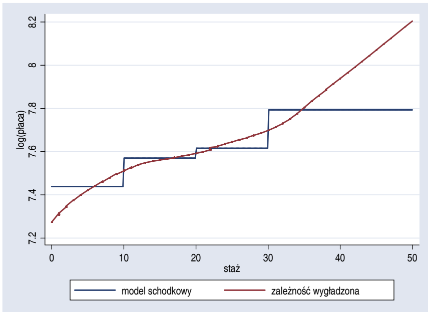
Egzamin z ekonometrii 13.06.2006
Informatyka i Ekonometria
I semestr

1. Opisać sposoby przybliżania za pomocą modelu liniowego zależności nieliniowej.

   Nieliniową zależność między y_i a x_i można przybliżyć za pomocą modelu liniowego stosując model:

   (a) Wielomianowy:
   $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \ldots + \beta_k x_i^k + \varepsilon_i$
   Przy większej liczbie zmiennych objaśniających wstawia się do modelu ich kwadraty i iloczyny.

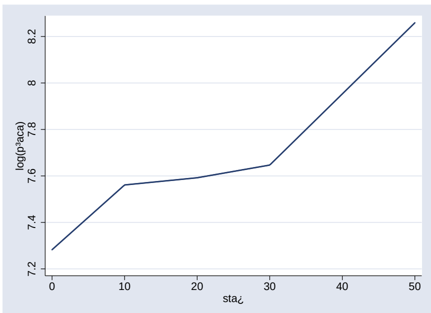
   (b) schodkowy
   W tym przypadku definiujemy zmienne zerojedynkowe $D_{s,i} = \mathbb{I}(x_s^* < x_i \leq x_{s+1}^*)$ związane z przedziałami x_i i przeprowadzamy regresję na tych zmiennych zamiast na x_i. Wyestymowany model można zilustrować rysunkiem:



   (c) krzywej łamanej (odcinakami liniowej) o postaci

$$y = \begin{cases} \alpha + \beta_1 x + \varepsilon & \text{dla} \quad x_i \leq x_1^* \\ \alpha + \beta_1 x_1^* + \beta_2 (x - x_1^*) + \varepsilon & \text{dla} \quad x_1^* < x_i \leq x_2^* \\ \vdots & \qquad \vdots \\ \alpha + \beta_1 x_1^* + \sum_{j=2}^{s-1} \beta_j \left(x_j^* - x_{j-1}^* \right) + \beta_s x + \varepsilon & \text{dla} \quad x_i > x_s^* \end{cases}$$

   Zależność nieliniowa przybliżona jest w tym przypadku krzywą, którą można zilustrować rysunkiem:



2. Udowodnić twierdzenie Gaussa-Markowa.

   Twierdzenie Gaussa-Markowa: dla spełnionych założeń $KMRL$ estymator MNK jest najlepszym estymatorem wektora parametrów $\boldsymbol{\beta}$ w klasie liniowych i nieobciążonych estymatorów tego parametru.

   Estymator liniowy wektora parametrów $\boldsymbol{\beta}$

$$\widetilde{\boldsymbol{\beta}} = \boldsymbol{C}\boldsymbol{y} = \boldsymbol{C}\left(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\right)$$

Wartość oczekiwana

$$\begin{aligned} \mathrm{E}\left(\widetilde{\boldsymbol{\beta}}\right) &= \mathrm{E}\left(\boldsymbol{C}\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{C}\boldsymbol{\varepsilon}\right) \\ &= \boldsymbol{C}\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{C}\underbrace{\mathrm{E}\left(\boldsymbol{\varepsilon}\right)}_{\mathbf{0}} = \boldsymbol{C}\boldsymbol{X}\boldsymbol{\beta} \end{aligned}$$

Estymator $\widetilde{\boldsymbol{\beta}}$ jest nieobciążony, a więc $\mathrm{E}\left(\widetilde{\boldsymbol{\beta}}\right) = \boldsymbol{\beta}$, co implikuje, że $\boldsymbol{CX} = \boldsymbol{I}$. Wariancja estymatora liniowego jest równa

$$\begin{aligned}\mathrm{Var}\left(\widetilde{\boldsymbol{\beta}}\right) &= \mathrm{Var}\left(\boldsymbol{Cy}\right) = \mathrm{Var}\left(\boldsymbol{CX\beta} + \boldsymbol{C\varepsilon}\right) \\ &= \mathrm{Var}\left(\boldsymbol{C\varepsilon}\right) = \boldsymbol{C}\underbrace{\mathrm{Var}\left(\boldsymbol{\varepsilon}\right)}_{\sigma^2\boldsymbol{I}}\boldsymbol{C}' = \sigma^2\boldsymbol{CC}'\end{aligned}$$

Zdefiniujmy taką macierz $\boldsymbol{D}$, że

$$\boldsymbol{D} = \boldsymbol{C}' - \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$$

$\boldsymbol{C}$ można zapisać jako:

$$\boldsymbol{C} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}' + \boldsymbol{D}'$$

Mnożąc obie strony prawostronnie przez $\boldsymbol{X}$ uzyskujemy

$$\underbrace{\boldsymbol{CX}}_{\boldsymbol{I}} = \underbrace{\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{X}}_{\boldsymbol{I}} + \boldsymbol{D}'\boldsymbol{X}$$

a więc

$$\boldsymbol{D}'\boldsymbol{X} = \boldsymbol{0}$$

Wielkość wariancji $\widetilde{\boldsymbol{\beta}}$ jest równa:

$$\begin{aligned}\mathrm{Var}\left(\widetilde{\boldsymbol{\beta}}\right) &= \sigma^2\boldsymbol{CC}' = \sigma^2\left[\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}' + \boldsymbol{D}'\right]\left[\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}' + \boldsymbol{D}'\right]' \\ &= \sigma^2\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\underbrace{\boldsymbol{X}'\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}}_{\boldsymbol{I}} + \sigma^2\boldsymbol{D}'\boldsymbol{D} \\ &+ \sigma^2\underbrace{\boldsymbol{D}'\boldsymbol{X}}_{\boldsymbol{0}}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} + \sigma^2\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\underbrace{\boldsymbol{X}'\boldsymbol{D}}_{\boldsymbol{0}} \\ &= \sigma\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} + \sigma\boldsymbol{D}'\boldsymbol{D}\end{aligned}$$

Estymator MNK ma wariancję $\mathrm{Var}\left(\boldsymbol{b}\right) = \sigma\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$. Różnica między macierzą wariancji dowolnego estymatora liniowego i niobciążonego i macierzą wariancji estymatora MNK jest równa

$$\mathrm{Var}\left(\widetilde{\boldsymbol{\beta}}\right) - \mathrm{Var}\left(\boldsymbol{b}\right) = \sigma^2\boldsymbol{D}'\boldsymbol{D}$$

i jest dodatnio określona.

1. Co to jest prognoza? Udowodnić, że prognoza postaci $\boldsymbol{x}_f\boldsymbol{b}$ jest nieobciążona.

   Prognoza jest przewidywaną na podstawie modelu wielkością y_f dla pewnej znanej wielkości $\boldsymbol{x}_f$. Dla modelu spełniającego założenia $KMRL$ prognoza $\widehat{y}_f = \boldsymbol{x}_f\boldsymbol{b}$ jest nieobciążona, ponieważ oczekiwany błąd prognozy $e_f = y_f - y_f$ jest równy zeru:

   $$\mathrm{E}\left(e_f\right) = \mathrm{E}\left(\widehat{y}_f - y_f\right) = \mathrm{E}\left[\boldsymbol{x}_f\left(\boldsymbol{b} - \boldsymbol{\beta}\right) - \varepsilon_f\right] = \boldsymbol{x}_f\,\mathrm{E}\left(\boldsymbol{b} - \boldsymbol{\beta}\right) + \mathrm{E}\left(\varepsilon_f\right) = 0$$

   Ponieważ estymator MNK jest nieobciążony a wartość oczekiwana ε_i jest z założeń $KMRL$ równa zero.

2. Jak testujemy hipotezę postaci, $H_0 : \boldsymbol{H\beta} = \boldsymbol{h}$ używając do tego sum kwadratów reszt z modelu bez ograniczeń i z ograniczeniami?

   Liczymy model bez ograniczeń. Następnie wstawiamy ograniczenia wynikające H_0 do modelu i estymujemy model z ograniczeniami. Do testowania H_0 wykorzystujemy statystykę

   $$F = \frac{\left(S_R - S\right)/\,g}{S/\left(N - K\right)} \sim F\left(g, N - K\right)$$

   gdzie S_R jest sumą kwadratów reszt z modelu z ograniczeniami a S z modelu bez ograniczeń

ZADANIE 1 Oszacowano model

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$$

za pomocą MNK i uzyskano oszacowania parametrów b_0 i b_1. Pomiar zmiennej niezależnej x_i był jednak systematycznie niedokładny. Oznaczmy jako x_i^* prawdziwą wartość zmiennej niezależnej a jako c identyczny dla wszystkich obserwacji błąd pomiaru. Załóżmy, że założenia $KMRL$ są spełnione dla modelu:

$$y_i = \beta_1 + \beta_2 x_i^* + \varepsilon_i$$

1. Pokazać, że jeśli dla każdej obserwacji $x_i = c x_i^*$, to estymator MNK parametru β_2 będzie obciążony a parametru β_1 nieobciążony.

   Podpowiedź: Skorzystaj z tego, że w modelu z ze stała i jedną zmienną objaśniającą $b_2 = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}$ a $b_1 = \overline{y} - b_2\overline{x}$ i zastanów się czemu jest równa $\mathrm{E}(y_i)$

Rozwiązanie:

1. Wartość oczekiwana y_i

$$\mathrm{E}(y_i) = \mathrm{E}(\beta_1 + x_i^*\beta_2 + \varepsilon_i) = \beta_1 + x_i^*\beta_2 = \beta_1 + c^{-1}x_i\beta_2$$

   Wartość oczekiwana b_2

$$\begin{aligned}\mathrm{E}(b_2) &= \mathrm{E}\left[\frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}\right] \\ &= \frac{1}{n\sum x_i^2 - (\sum x_i)^2}\,\mathrm{E}\left(n\sum x_i y_i - \sum x_i \sum y_i\right)\end{aligned}$$

$$\begin{aligned}\mathrm{E}\left(n\sum x_i y_i - \sum x_i \sum y_i\right) &= n\sum x_i\,\mathrm{E}(y_i) - \sum x_i \sum \mathrm{E}(y_i) \\ &= c^{-1}n\sum x_i^2\beta_2 - c^{-1}\left(\sum x_i\right)^2\beta_2 \\ &= c^{-1}\left\{n\sum x_i^2 - \left(\sum x_i\right)^2\right\}\end{aligned}$$

$$\mathrm{E}(b_2) = \frac{1}{n\sum x_i^2 - (\sum x_i)^2}c^{-1}\left\{n\sum x_i^2 - \left(\sum x_i\right)^2\right\} = c\beta_2$$

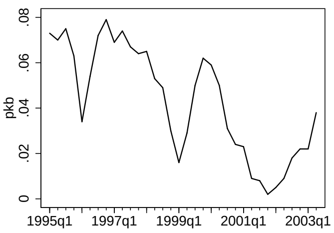
   Wartość oczekiwana b_1

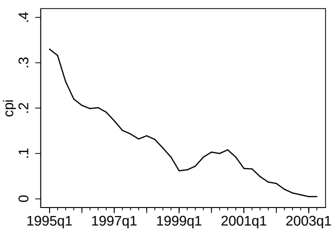
$$\begin{aligned}\mathrm{E}(b_1) = \mathrm{E}(\overline{y} - b_2\overline{x}) &= \mathrm{E}\left(\frac{\sum y_i}{n} - b_2\frac{\sum x_i}{n}\right) \\ &= \frac{\sum \mathrm{E}(y_i)}{n} - \mathrm{E}(b_2)\frac{\sum x_i}{n} \\ &= \frac{\sum\left(\beta_1 + c^{-1}x_i\beta_2\right)}{n} - c^{-1}\beta_2\frac{\sum x_i}{n} \\ &= \beta_1 + c^{-1}\frac{\sum x_i}{n}\beta_2 - c^{-1}\beta_2\frac{\sum x_i}{n} = \beta_1\end{aligned}$$

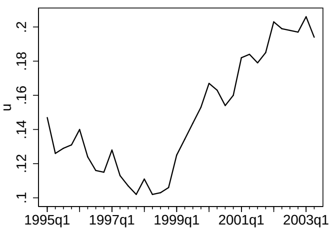
ZADANIE 2 Oszacowano regresję stopy bezrobocia (zmienna u) na wzroście realnego PKB (zmienna pkb) i stopie inflacji (zmienna cpi) oraz na zmiennych zero-jedynkowych związanymi z kwartałami. Regresję przeprowadzono na danych kwartalnych dla Polski z okresu 1995-2003. Wyniki regresji oraz wykresy zmiennych znajdują się poniżej:

```
      Source |       SS       df       MS              Number of obs =      34
-------------+------------------------------           F(  5,    28) =    6.52
       Model |  .020931923     5  .004186385           Prob > F      =  0.0004
    Residual |   .01796652    28  .000641661           R-squared     =  0.5381
-------------+------------------------------           Adj R-squared =  0.4556
```

```
       Total |  .038898443    33  .001178741           Root MSE      =  .02533
------------------------------------------------------------------------------
           u |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         pkb |    -.72455   .2843119    -2.55   0.017    -1.306937   -.1421634
         cpi |  -.0966945    .081363    -1.19   0.245     -.263359      .06997
       _Iq_2 |  -.0068327   .0119961    -0.57   0.574    -.0314056    .0177403
       _Iq_3 |  -.0117543   .0124543    -0.94   0.353    -.0372657    .0137572
       _Iq_4 |  -.0126153   .0124595    -1.01   0.320    -.0381374    .0129067
       _cons |   .1980322   .0112852    17.55   0.000     .1749155    .2211489
------------------------------------------------------------------------------
Durbin-Watson(6,34)      =   .24
Breusch-Godfrey F(2,26)  = 12.62 [0.0001]
Breusch-Pagan   chi2(2)  = 10.41 [0.0054]
White           chi2(14) = 25.80 [0.0274]
RESET            F(3,25) =  2.01 [0.1379]
Jarque-Bera       chi(2) =  2.18 [.33723]
Chow(k=1998.1)   F(6,22) =  3.87 [0.0087]
```







Przy założonym poziomie istotności $\alpha = 0.05$ przeprowadzić analizę wyników. Każdą z odpowiedzi należy uzasadnić wielkościami odpowiednich statystyk testowych.

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 w modelu są stabilne.
8. Zinterpretować współczynnik przy pkb.
9. Wyjaśnić na podstawie wykresów zmiennych dlaczego w modelu zawarto kwartalne zmienne sezonowe.
10. W jaki sposób należałoby weryfikować hipotezę mówiącą o tym, że jedyną zmienną wpływającą na bezrobocie jest pkb?
11. Jeśli model nie spełnia założeń KMRL określić:

(a) które założenie nie jest spełnione?

(b) jakie ma to konsekwencje dla wnioskowania statystycznego?

(c) za pomocą jakich metod estymacji można rozawiązać problemy sygnalizowane przez wyniki testów?

Rozwiązanie:

1. Oszacowany model objaśnia prawie 54% zmienności zmiennej zależnej (współczynnik $R^2 = 0.538$). Zbiór zmiennych objaśniających (poza stałą) jest łącznie istotny ($F_{5,28} = 6.52$).

2. Tylko zmienna pkb i stała są istotne na zadanym poziomie istotności (*p-value* dla statystyk t są mniejsze od 0.05).

3. Na podstawie testu Breuscha-Godfreya ($0.0001 < 0.05$) i Durbina-Watsona ($0.24 < d_L = 1.15$) należy odrzucić H_0 o braku autokorelacji.

4. Na podstawie testu Breuscha-Pagana ($0.8153 > 0.05$) i White'a ($0.8436 > 0.05$) nie można odrzucić hipotezy o homoskedastyczności.

5. Na podstawie testu RESET ($0.1379 > 0.05$) nie można odrzucić hipotezy o poprawnej formie funkcyjnej.

6. Na podstawie testu Jarque'a-Bery ($0.3372 > 0.05$) nie można odrzucić hipotezy o normalności rozkładu błędów losowych.

7. Na podstawie testu Chowa ($0.0087 < 0.05$) należy odrzucić hipotezę o stabilności parametrów modelu w okresach $t < 1998Q1$ i $t \geq 1998Q1$.

8. Współczynnik przy zmiennej oznacza, że wzrost PKB o 1% spowoduje spadek stopy bezrobocia o 0.72%.

9. Na wykresie widać, że w pierwszym kwartale następuje sezonowy wzrost bezrobocia. Aby to uwzględnić, zawarto w modelu zmienne sezonowe.

10. Hipoteze o łącznej nieistotności stałej, cpi i zmiennych kwartalnych należy testestować za pomocą testu F, nakładając łączne ograniczenie na parametry przy tych zmiennych - przyrównując je jednocześnie do zera.

11. (a) Niespełnione jest założenie o braku autokorelacji składnika losowego i o stałości parametrów w całej próbie.

    (b) W przypadku autokorelacji składnika losowego estymator MNK $\boldsymbol{b}$ jest nadal nieobciążony, ale jest nieefektywny. Obciążone oszacowanie macierzy wariancji-kowariancji uniemożliwia prawidłowe przeprowadzenie wnioskowania statystycznego. Z racji na niestabilność parametrów w czasie wynik regresji wydaje się w całości wątpliwy.

    (c) Model można szacować oddzielnie na podpróbach, co umożliwi obliczenie różnych wielkości współczynników regresji w poszczególnych podpróbkach. W celu uzyskania prawidłowego oszacowania macierzy waraincji kowariancji można zastosować odporną macierz Newey'ego-Westa. Można też podjąć próbę usunięcia autokorelacji za pomocą Stosowalnej UMNK.

ZADANIE 3 Uzyskano na podstawie dwóch niezależnych próbek o liczebnościach n_1 i n_2 oszacowania $\boldsymbol{b}_1$ i $\boldsymbol{b}_2$ parametru $\boldsymbol{\beta}$ w modelu:

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

Oznaczmy jako $\boldsymbol{X}_1, \boldsymbol{y}_1$ macierze obserwacji w pierwszej próbie i odpowiednio jako $\boldsymbol{X}_2, \boldsymbol{y}_2$ macierze obserwacji w drugiej próbie. Załóżmy, że w obu próbkach spełnione są założenia $KMRL$.
Uwaga: przez niezależność próbek rozumiemy, że błędy losowe w obu próbach są od siebie niezależne.

1. Zaproponowano, żeby β wyestymować za średniej ważonej następującej średniej ważonej $\boldsymbol{b}_1$ i $\boldsymbol{b}_2$:

$$\bar{\boldsymbol{b}} = w_1\boldsymbol{b}_1 + w_2\boldsymbol{b}_2$$

gdzie nielosowe wagi $w_1 + w_2 = 1$. Pokazać, że estymator ten jest nieobciążony i policzyć jego wariancję.

2. Pokazać, że estymator MNK policzona na połączonych próbkach będzie miał inną postać niż estymator $\overline{\boldsymbol{b}}$, że można go przedstawić jako $\boldsymbol{b} = \boldsymbol{W}_1\boldsymbol{b}_1 + \boldsymbol{W}_2\boldsymbol{b}_2$ i że $\boldsymbol{W}_1 + \boldsymbol{W}_2 = \boldsymbol{I}$.

   Podpowiedź: Macierze obserwacji dla całej próby można zapisać jako $\boldsymbol{y} = \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix}$ i $\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix}$.

3. Dla szczególnego przypadku, kiedy w modelu mamy jedynie stałą znaleźć takie wagi w_1 i w_2, dla których estymator $\overline{\boldsymbol{b}}$ będzie równoważny estymatorowi $\boldsymbol{b}$.

4. Przy założeniu, że β i σ^2 są w obu podpróbkach takie same, który z estymatorów (MNK czy $\overline{\boldsymbol{b}}$) będzie lepszy? (odpowiedź uzasadnić)

Rozwiązanie:

1. Wartość oczekiwana estymatora $\overline{\boldsymbol{b}}$

$$\begin{aligned} \mathrm{E}\left(\overline{\boldsymbol{b}}\right) &= \mathrm{E}\left(w_1\boldsymbol{b}_1 + w_2\boldsymbol{b}_2\right) = w_1\,\mathrm{E}\left(\boldsymbol{b}_1\right) + w_2\,\mathrm{E}\left(\boldsymbol{b}_2\right) \\ &= w_1\,\boldsymbol{\beta} + w_2\boldsymbol{\beta} = \boldsymbol{\beta} \end{aligned}$$

   przy czym skorzystaliśmy z nieobciążoności estymatora MNK w $KMRL$. Wariancja $\overline{\boldsymbol{b}}$

$$\begin{aligned} \mathrm{Var}\left(w_1\boldsymbol{b}_1 + w_2\boldsymbol{b}_2\right) &= w_1^2\,\mathrm{Var}\left(\boldsymbol{b}_1\right) + w_2^2\,\mathrm{Var}\left(\boldsymbol{b}_2\right) + w_1w_2\,\mathrm{Cov}\left(\boldsymbol{b}_1, \boldsymbol{b}_2\right) \\ &= w_1^2\sigma_1^2\left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1} + w_2^2\sigma_2^2\left(\boldsymbol{X}_2'\boldsymbol{X}_2\right)^{-1} \end{aligned}$$

   Przy czym skorzystaliśmy z tego, że estymatory policzone na podstawie niezależnych próbek są od siebie niezależne a więc $\mathrm{Cov}\left(\boldsymbol{b}_1, \boldsymbol{b}_2\right) = 0$

2. Estymator MNK ma postać

$$\boldsymbol{X}'\,\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1' & \boldsymbol{X}_2' \end{bmatrix}\begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix} = \boldsymbol{X}_1'\boldsymbol{X}_1 + \boldsymbol{X}_2'\boldsymbol{X}_2$$

$$\boldsymbol{X}'\,\boldsymbol{y} = \begin{bmatrix} \boldsymbol{X}_1' & \boldsymbol{X}_2' \end{bmatrix}\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix} = \boldsymbol{X}_1'\boldsymbol{y}_1 + \boldsymbol{X}_2'\boldsymbol{y}_2$$

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

   i rzeczywiście

$$\begin{aligned} \boldsymbol{W}_1 + \boldsymbol{W}_2 &= \left(\boldsymbol{X}_1'\boldsymbol{X}_1 + \boldsymbol{X}_2'\boldsymbol{X}_2\right)^{-1}\left(\boldsymbol{X}_1'\boldsymbol{X}_1\right) + \left(\boldsymbol{X}_1'\boldsymbol{X}_1 + \boldsymbol{X}_2'\boldsymbol{X}_2\right)^{-1}\left(\boldsymbol{X}_2'\boldsymbol{X}_2\right) \\ &= \left(\boldsymbol{X}_1'\boldsymbol{X}_1 + \boldsymbol{X}_2'\boldsymbol{X}_2\right)^{-1}\left(\boldsymbol{X}_1'\boldsymbol{X}_1 + \boldsymbol{X}_2'\boldsymbol{X}_2\right) = \boldsymbol{I} \end{aligned}$$

3. Jeśli $\boldsymbol{X}_1 = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$ i $\boldsymbol{X}_2 = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$, to $\boldsymbol{X}_1'\boldsymbol{X}_1 = n_1$, $\boldsymbol{X}_2'\boldsymbol{X}_2 = n_2$ a $\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{X}_1'\boldsymbol{X}_1 + \boldsymbol{X}_2'\boldsymbol{X}_2 = n_1 + n_2$. Macierz $\boldsymbol{W}_1 = \frac{n_1}{n_1+n_2} = w_1$ a $\boldsymbol{W}_2 = \frac{n_2}{n_1+n_2} = w_2$.

4. Jeśli β i σ^2 są takie same w obu podpróbkach, to dla próby złożonej ze wszystkich obserwacji spełnione są założenia $KMRL$. Z twierdzenia Gaussa-Markowa wnioskujemy, że najlepszym liniowym i niebciążonym estymatorem bazującym na wszystkich obserwacjach jest estymator MNK. Estymator $\overline{\boldsymbol{b}}$ jest liniowy i nieobciążony ale ma inną postać niż estymator MNK - musi więc być gorszy niż estymator MNK (mieć większą wariancję)