imię, nazwisko, nr indeksu:____

Ekonometria — egzamin
wersja ogólna
29/01/08

1. Egzamin trwa 90 minut.
2. Rozwiązywanie zadań należy rozpocząć po ogłoszeniu początku egzaminu a skończyć wraz z ogłoszeniem końca egzaminu. Złamanie tej zasady skutkuje usunięciem z sali i unieważnieniem pracy.
3. W trakcie egzaminu wolno używać jedynie długopisu o innym kolorze atramentu niż czerwony oraz kalkulatora.
4. Przed przystąpieniem do pisania egzaminu należy podpisać **wszystkie kartki** pracy (na dole w przewidzianym miejscu). Złożenie podpisu pod regulaminem oznacza jego akceptację. Do egzaminu mogą przystąpić osoby, które akceptują regulamin.
5. W razie braku podpisu lub numeru zadania na kartce, kartka nie zostanie oceniona. Nie będą też oceniane rozwiązania wpisane na kartkach innych, niż te rozdawane przez prowadzących.
6. Rozwiązanie każdego zadania należy zapisać na kartce z tymże zadaniem, ewentualnie na czystych kartkach znajdujących się na końcu arkusza egzaminacyjnego lub na dodatkowych kartkach uzyskanych od prowadzących egzamin.
7. Na jednej kartce może znajdować się rozwiązanie tylko jednego zadania. Oceniane jest rozwiązanie tylko tego zadania, którego numer widnieje na kartce.
8. Egzamin składa się z czterech pytań teoretycznych i 3 zadań.
9. Posiadanie przy sobie wszelkich materiałów drukowanych (w tym książek) oraz innych np. wykonanych własnoręcznie materiałów zostanie uznane za ściąganie.
10. Rozmowy z innymi zdającymi będą traktowane identycznie jak ściąganie.
11. Każda zauważona próba ściągania skutkuje usunięciem z egzaminu.
12. Wszystkie pytania należy kierować bezpośrednio do osób pilnujących.
13. Warunkiem uzyskania oceny pozytywnej jest zdobycie conajmniej 50 % punktów z części teoretycznej egzaminu oraz min. 40 % punktów z cześci zadaniowej.

Warszawa, 29/01/2008, ____________________________
podpis

Powodzenia :-)

Pytania teoretyczne

1	2	$\sum$

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

   Budując model zjawiska zachodzącego w populacji posługujemy się informacjami pochodzącymi z próby. Zakładamy że w populacji generalnej występuje zależność

   $$y = X\beta + \varepsilon \tag{1}$$

   a na podstawie informacji z próby budujemy model tej zależności

   $$y = Xb + e$$

   β jest wektorem nieznanych parametrów, b jest wektorem oszacowań nieznanych parametrów, ε to wektor tzw. składnika losowego równania, a e jest wektorem oszacowań tego składnika na podstawie próby. Podsumowując. Parametry β oraz ε są nieznany i pochodzą z populacji, a b oraz e są ich oszacowaniami na podstawie próby empirycznej.

2. Dlaczego zmienną nominalną rozkodowuje się na odpowiednią liczbę zmiennych dyskretnych. Ile powinno być tych zmiennych i dlaczego?

   Zmienną dyskretną przyjmującą s wartości należy rozkodować na $s-1$ zmiennych zero-jedynkowych. Rozkodowanie umożliwia zarówno określenie kierunku zależności, jak i interpretację zależności ilościowej, której nie można określić bez rozkodowania. Przy rozkodowaniu pomija się tzw. poziom bazowy zjawiska (najczęściej jest to najbardziej liczna kategoria) aby uniknąć problemu współliniowości grupy zmiennych 0-1 ze stałą modelu.

Pytania teoretyczne c.d.

3	4	$\sum$

3. Podaj dwa źródła błędu prognozy, oraz wzór na jej wariancję (bez wyprowadzenia).

   Wariancja prognozy wynosi:

   $$var[e^0 \mid X, x^0] = \sigma^2 + var[(b-\beta)'x^0 \mid X, x^0] = \sigma^2 + x^{0,}(X'X)^{-1}x^0$$

   Jest ona sumą niedokładności oszacowań parametrów oraz błędu losowego σ^2.

4. Wyjaśnij jakie korzyści i niebezpieczeństwa wiążą się z narzucaniem ograniczeń na model.

   Korzyści: uzyskiwane wartości oszacowań parametrów maja niższą wariancję w przypadku prawidłowości ograniczeń. Niebezpieczeństwo: w przypadku, gdy nałożone ograniczenia są nieprawdziwe uzyskane oszacowanie będzie obciążone. W pewnych przypadkach (np. pominięcie zmiennej istotnej) uzyskany estymator może nie być zgodny.

1	2	3	4	5	$\sum$

Zadanie 1 Dysponujesz następującymi modelami regresji

- ($\star$) y_i na stałej oraz x_{2i} i x_{3i}
- ($\star\star$) $y_i^\star$ na stałej oraz x_{2i} i x_{3i}, gdzie $y^\star = y_i - 3{,}14$

Proszę odpowiedzieć na następujące pytania (każdą odpowiedź należy dokładnie uzasadnić)

1. Jaka zachodzi relacja między RSS w tych regresjach?
2. Jaka zachodzi relacja między R^2 w tych regresjach?
3. Załóżmy, że regresję ($\star$) oszacowano na próbce liczącej N obserwacji (oczywiście zakładamy, że N > 4) i uzyskano oszacowania $\hat{b_1}, \hat{b_2}, \hat{b_3}$. Następnie dodano obserwację o indeksie $N+1$ dla której zachodzi:
$$y_{N+1} = \bar{y} \text{ gdzie } \bar{y} = \frac{\sum_{i=1}^N y_i}{N}$$
$$x_{2,N+1} = \bar{x_2} \text{ gdzie } \bar{x_2} = \frac{\sum_{i=1}^N x_{2i}}{N}$$
$$x_{3,N+1} = \bar{x_3} \text{ gdzie } \bar{x_3} = \frac{\sum_{i=1}^N x_{3i}}{N}$$
Ponownie oszacowano regresję ($\star$), ale na próbie liczącej $N+1$ obserwacji. Udowodnij, iż w drugiej regresji uzyskano dokładnie te same oszacowania co za pierwszym razem.
4. Porównaj R^2 w regresjach opisanych w podpunkcie 3.
5. Dany jest następujący model $y_i = \alpha x_i + \varepsilon_i$, $E(\varepsilon_i) = 0$, $var(\varepsilon_i) = \sigma^2\mathbb{I}$. Podzielono próbę na dwie równoliczne podpróby według rosnących wartości zmiennej zależnej. Niech $\bar{y_1}$ oraz $\bar{x_1}$ oznaczają odpowiednio średnie dla zmiennej y i x w pierwszej podpróbie. Analogicznie definiujemy $\bar{y_2}$ i $\bar{x_2}$ dla drugiej podpróby.
   - Zakładając że $\bar{x_1} \neq \bar{x_2}$ pokaż, że $\bar{\alpha} = \frac{\bar{y_2} - \bar{y_1}}{\bar{x_2} - \bar{x_1}}$ jest nieobciążonym estymatorem parametru β.
   - Pokazać, że $var(\bar{\alpha}) \geq var(b)$ gdzie b jest estymatorem MNK parametru β.

1	2	3	4	5	6	$\sum$

Zadanie 2 Badacz dysponował następującymi informacjami $\bar{y}$ - średnia ocen absolwenta WNE, x_1 - iloraz inteligencji badanych osób, x_2 - przeciętna ilość godzin w tygodniu, jaką poświęcali na naukę, płeć absolwenta (1 - kobieta; 0 - mężczyzna). Następnie oszacował parametry modelu $y = \alpha_0 + \alpha_1 x_1 + \nu$. Literatura wskazuje jednak, że model jest niekompletny.

1. Prawdopodobnie zmienna x_2, która w regresji została pominięta ma istotny wpływ na kształtowanie się średniej ocen absolwentów. Wiedząc, że korelacja y z x_2 jest dodatnia, oraz korelacja x_1 z x_2 ujemna, wypowiedz się o spodziewanym kierunku obciążenia estymatora parametru. Odpowiedź uzasadnij (wystarczy argumentacja opisowa - nie wyprowadzać żadnych własności w sposób formalny!)

2. kiedy pominięcie zmiennej x_2 nie wpłynie na obciążenia estymatora (proszę wymienić dwie sytuacje)?

   Następnie badacz estymował model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \varepsilon$

3. Wytłumacz dlaczego w modelu uwzględnił element $\beta_3 x_1 x_2$. Odpowiadając odwołaj się do tego jaki związek może zachodzić pomiędzy zmienną objaśnianą a zmiennymi objaśniającymi. Występowanie jakiego efektu możemy analizować za pomocą tego elementu?

4. Czy wszystkie parametry modelu możemy interpretować oddzielnie? Odpowiedź uzasadnij! Wskaż poprawną odpowiedź (tylko jedna prawdziwa):

   - tak - wszystkie
   - nie
   - oddzielnie możemy interpretować tylko parametry $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$, ale nie parametr $\hat{\beta}_3$.

   Zinterpretuj parametry modelu, jeżeli wiadomo, że $\hat{\beta}_0 = 3{,}79$ $\hat{\beta}_1 = 0{,}05$ $\hat{\beta}_2 = -0{,}46$ $\hat{\beta}_3 = -0{,}01$

   Badacz ustalił, że zmienna x_2 jest nieistotna w tłumaczeniu zmienności y. Literatura jednak wskazuje jeszcze jedno źródło zmienności - płeć absolwenta. Badacz wyestymował więc model:

   $y = \gamma_0 + \gamma_1 x_1 + \gamma_2 plec + \gamma_3 plec \times x_1 + \varepsilon$

5. Czemu umieścił w modelu element $\gamma_3 plec \times x_1$?

6. Zinterpretuj parametry modelu wiedząc, że $\hat{\gamma}_0 = 3{,}79$ $\hat{\gamma}_1 = 0{,}05$ $\hat{\gamma}_2 = -0{,}46$ $\hat{\gamma}_3 = 0{,}01$

1	2	3	4	5	$\sum$

ZADANIE 3 **Zadanie 1.**

Październikowe Badanie Wynagrodzeń jest badaniem płac w przedsiębiorstwach zatrudniających powyżej 10 pracowników. Losowana próba jest dwuwartstwowa, w pierwszym etapie losowane są firmy, następnie pracownicy wybranych firmy. Na podstawie danych zgromadzonych w trakcie badania w roku 2004 oszacowano parametry równania płac dla Polski, uzyskując następujące wyniki:

```
      Source |       SS       df       MS              Number of obs =  609764
-------------+------------------------------           F(  7,609756) =10831.84
       Model |  19719.2446     7  2817.03494           Prob > F      =  0.0000
    Residual |  158579.135609756  .260069823           R-squared     =  0.1106
-------------+------------------------------           Adj R-squared =  0.1106
       Total |   178298.38609763  .292406033           Root MSE      =  .50997

------------------------------------------------------------------------------
    lzarobki |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        wiek |   .0099372   .0001445    68.76   0.000      .009654    .0102204
     _Iplec_2 |  -.1543351   .1358100    -1.13   0.000    -.1569969   -.1516733
        dosw |   .0040469     .00025    16.19   0.000      .003557    .0045369
       dosw2 |  -.0000909   5.62e-06   -16.18   0.000    -.0001019   -.0000799
    prywatna |   .1654016   .0013657   121.11   0.000     .1627249    .1680784
_Iwielkosc_2 |   .1253921   .0023077    54.34   0.000      .120869    .1299151
_Iwielkosc_3 |   .2222128   .0021699   102.41   0.000     .2179598    .2264657
       _cons |   7.016897   .0044526  1575.90   0.000      7.00817    7.025624
------------------------------------------------------------------------------
Breusch-Pagan     chi2(7)             = 55777.15     Prob > chi2  = 0.0000
Ramsey RESET test F(3, 609753       ) =    775.40    Prob > F =     0.0000
Durbin-Watson d-statistic(  8,609764) =  .8848615
Breusch-Godfrey LM test (4)           = 266838.52    Prob > F =     0.0000

Cameron & Trivedi's decomposition of IM-test
---------------------------------------------------
              Source |       chi2     df      p
---------------------+-----------------------------
  Heteroskedasticity |   39060.14     29    0.0000
            Skewness |   14375.81      7    0.0000
            Kurtosis |    1473.80      1    0.0000
---------------------+-----------------------------
               Total |   54909.75     37    0.0000
---------------------------------------------------
```

Zmienna *plec* przyjmuję wartość 2 dla kobiety, *dosw* to doświadczenie zawodowe pracownika w latach, *dosw*2 to kwadrat doświadczenia, *wielkosc* 1 oznacza firmę do 20 pracowników, *wielkosc* 2 oznacza firmę zatrudniającą od 20 do 100 pracowników, a *wielkosc* 3 firmę zatrudniającą powyżej 100 pracowników.

Przyjmując poziom istotności $\alpha = 0{,}05$ odpowiedz na poniższe pytania:

1. Oceń dopasowanie danych empirycznych do modelu. Określ które zmienne można uznać za statystycznie istotne?
2. Czy forma funkcyjna oszacowanego powyżej modelu jest liniowa?
3. Czy powyższy model spełnia założenia KMRL (wymień, które spełnia, a których nie spełnia)
4. Zaproponuj sposób weryfikacji hipotezy o braku wpływu doświadczenia zawodowego na wysokość uzyskiwanych zarobków.
5. Jeżeli model nie spełnia założeń KMRL opisz jakie to ma konsekwencje dla poprawności wnioskowania statystycznego, oraz jakie są metody radzenia sobie z tym problemem.

Każdą odpowiedź uzasadnij wynikiem odpowiednich testów diagnostycznych zapisując po wartość statystyki testowej lub jej *p-value,* oraz jej interpretację.

Rozwiązania zadań

Zadanie 1

1. Wprowadzamy następujące oznaczenia: $RSS^\star$ - suma kwadratów reszt dla regresji ($\star$); $RSS^{\star\star}$ - suma kwadratów reszt dla regresji ($\star\star$) Sumy kwadratów reszt są rozwiązaniem następującego problemu optymalizującego:

$$RSS^\star = \underset{\min}{b_1, b_2, b_3} \sum_{i=1}^{n} (y_i - b_1 - b_2 x_{2i} - b_3 x_{3i})^2$$

$$RSS^\star = \underset{\min}{b_2, b_3} \sum_{i=1}^{n} (y_i - 3{,}14 - b_2 x_{2i} - b_3 x_{3i})^2 = \sum_{i=1}^{n} (y_i - b_1 - b_2 x_{2i} - b_3 x_{3i})^2$$

W przypadku regresji ($\star$) minimalizujemy tę samą funkcję co w przypadku regresji ($\star$), ale wprowadzamy ograniczenie $b_1 = 3{,}14$. Czyli $RSS^\star$ to wartość minimalizowanej funkcji dla problemu optymalizacyjnego bez ograniczeń, a $RSS^{\star\star}$ to wartość minimalizowanej funkcji dla problemu optymalizacyjnego z ograniczeniami, więc $RSS^\star \leq RSS^{\star\star}$

2. $TSS^\star = \sum (y_i - \bar{y})^2$ - całkowita suma kwadratów dla regresji ($\star$). W regresji ($\star\star$) zmienną zależną jest $\tilde{y} = y_i - 3{,}14$. Średnia wartość zmiennej zależnej dla regresji ($\star\star$)

$$\bar{\tilde{y}} = \frac{1}{N} \sum_{i=1}^{N} (y_i - 3{,}14) = \bar{y} - 3{,}14$$

oraz TSS

$$TSS^{\star\star} = \sum_{i=1}^{N} (y_i - 3{,}14 - (\bar{y} - 3{,}14))^2 = \sum_{i=1}^{N} (y_i - \bar{y})^2$$

Otrzymujemy więc, że w obu regresjach TSS jest takie same. Przyjmijmy następujące oznaczenia $TSS = TSS^\star = TSS^{\star\star}$. Korzystając z wyników podpunktu 1 mamy:

$$RSS\star \leq RSS\star\star \;\Leftrightarrow\; \frac{RSS\star}{TSS} \leq \frac{RSS\star\star}{TSS} \;\Leftrightarrow\; 1 - \frac{RSS\star}{TSS} \geq 1 - \frac{RSS\star\star}{TSS} \;\Leftrightarrow\; R^{2\star} \geq R^{2\star\star}$$

3. Niech b_1, b_2 b_3 oznaczają oszacowania parametrów uzyskane w regresji na próbie liczącej N obserwacji, natomiast i to oszacowania tej samej regresji, ale na próbie liczącej $N+1$ obserwacji. Musimy pokazać, że $\gamma_1 = b_1$,$\gamma_2 = b_2$, $\gamma_3 = b_3$.

Problem optymalizacyjny dla regresji z $N+1$ obserwacjami:

$$RSS^\star = \underset{\min}{\gamma_1, \gamma_2, \gamma_3} \sum_{i=1}^{N+1} (y_i - \gamma_1 - \gamma_2 x_{2i} - \gamma_3 x_{3i})^2 =$$

$$\underset{\min}{\gamma_1, \gamma_2, \gamma_3} \sum_{i=1}^{N} ((y_i - \gamma_1 - \gamma_2 x_{2i} - \gamma_3 x_{3i})^2 + (y_{N+1} - \gamma_1 - \gamma_2 x_{2N+1} - \gamma_3 x_{3N+1})^2) =$$

$$\sum_{i=1}^{N} ((y_i - \gamma_1 - \gamma_2 x_{2i} - \gamma_3 x_{3i})^2 + (\bar{y} - \gamma_1 - \gamma_2 \bar{x_2} + \gamma_3 \bar{x_3})^2)$$

Pierwszy składnik przyjmuje swoje minimum dla $\gamma_1 = b_1$,$\gamma_2 = b_2$, $\gamma_3 = b_3$, gdyż jest to suma kwadratów reszt dla regresji na próbie liczącej N obserwacji. Drugi składnik przyjmuje najmniejszą wartość równą zero dla $\gamma_1 = b_1$,$\gamma_2 = b_2$, $\gamma_3 = b_3$, ponieważ $\bar{y} = b_1 + b_2 \bar{x_2} + b_3 \bar{x_3}$(punkt będący średnią z próby należy do hiperpłaszczyzny regresji w regresji ze stałą).

4. Pokażemy, że R^2 w obu regresjach będzie takie samo. Zaczynamy od analizy sumy kwadratów reszt w obu regresjach:

$$RSS_1 = \sum_{i=1}^{N}(y_i - b_1 - b_2x_{2i} - b_3x_{3i})^2$$

$$RSS_2 = \sum_{i=1}^{N+1}(y_i - b_1 - b_2x_{2i} - b_3x_{3i})^2 = \sum_{i=1}^{N}((y_i - b_1 - b_2x_{2i} - b_3x_{3i})^2 + (y_{N+1} - b_1 - bx_{2N+1} + b_3x_{3N+1})^2 = RSS_1$$

Wystarczy pokazać jeszcze, że TSS dla obu regresji jest takie same:

$$TSS_1 = \sum_{i=1}^{N}(y_i - \bar{y})^2$$

$$TSS_2 = \sum_{i=1}^{N+1}(y_i - \bar{y_2})^2$$

gdzie $\bar{y_2} = \frac{\sum_{i=1}^{N} y_i + \bar{y}}{N+1} = \bar{y}$ Przekształcając TSS_2 otrzymujemy

$$TSS_2 = \sum_{i=1}^{N+1}(y_i - \bar{y})^2 = \sum_{i=1}^{N}(y_i - \bar{y})^2 + (\bar{y} - \bar{y})^2$$

Ponieważ RSS i TSS w obu regresjach jest takie samo, więc R^2 również.

5. Korzystając z liniowości wartości oczekiwanej oraz nielosowości x:

$$E(\alpha) = E(\frac{\bar{Y_2} - \bar{Y_1}}{\bar{X_2} - \bar{X_1}}) = \frac{1}{\bar{X_2} - \bar{X_1}}E(\bar{Y_2}) - \frac{1}{\bar{X_2} - \bar{X_1}}E(\bar{Y_1})$$

Niech n oznacza liczebności obu prób. Wówczas

$$E(Y_1) = E(\frac{1}{N}\sum_{i=1}^{N}(x_i\beta + \varepsilon_i)) = \bar{x_1}\beta$$

$$E(Y_2) = E(\frac{1}{N}\sum_{i=N+1}^{2N}(x_i\beta + \varepsilon_i)) = \bar{x_2}\beta$$

Ostatecznie otrzymujemy

$$E(\alpha) = \frac{1}{\bar{X_2} - \bar{X_1}}\beta - \frac{1}{\bar{X_2} - \bar{X_1}}\beta = \beta$$

5b) Najprościej skorzystać z twierdzenia Gaussa-Markowa (odpowiednie założenia KMRL są spełnione). Pokażemy, że estymator α jest liniowy (czyli, że można go przedstawić w postaci Cy, gdzie C jest pewną nielosową macierzą).

$$E(\alpha) = \frac{1}{N(\bar{X_2} - \bar{X_1})}\sum_{i=N+1}^{2N} y_i - \frac{1}{N(\bar{X_2} - \bar{X_1})}\sum_{i=1}^{N} y_i$$

Zaproponowany w podpunkcie a) estymator jest liniowy i nieobciążony, więc na mocy twierdzenia Gaussa-Markowa zachodzi

$$var\alpha \geq var(b)$$

Zadanie 2

1. obciążenie ujemne, ponieważ wpływ nieuwzględnionej zmiennej zostaje częściowo przejęty przez zmienna uwzględnioną w modelu, ponieważ korelacja pomiędzy nimi jest ujemna (różna od zera). Oznacza to że wzrost x_1 współwystępuje ze spadkiem x_2, a x_2 jest dodatnio skorelowane z y. W związku z tym wraz ze wzrostem x_1 rośnie y ale mniej niż, gdyby w modelu znajdowała się zmienna x_2.

2. brak korelacji pomiędzy x_1 i x_2, brak korelacji x_2 i y (inaczej: x2 jest nieistotną zmienną w modelu wyjaśniającym y) 1pkt

3. spodziewamy się nieliniowego wpływu regresorów jednoczesny wzrost inteligencji i pracowitości, to więcej niż proporcjonalny wzrost średniej ocen, **występowanie efektu synergii**

4. nie,ponieważ parametry β_1 i β_3, oraz parametry β_2 i β_3 znajdują się przy tych samych zmiennych objaśniających i opisują związek tej samej zmiennej objaśniającej z y.

5. Stałej nie interpretujemy, dla $x_1: \ \beta_1 + \beta_3 x_2$ przy jednostkowym wzroście x_1 y wzrasta o $\beta_1 +$ dodatkowo $\beta_3 * wiek$, dla x_2 analogicznie.

6. interakcja, która ma wychwycić różnicę w tym jak zmiana IQ wpływa na średnią kobiet i mężczyzn lub interakcja zmiennej dyskretnej ze zmienną ciągłą (quasi-ciągłą) umożliwiająca podział populacji absolwentów na podpopulacje, w których potencjalnie związek IQ ze średnią jest inny.

7. Stałej nie interpretujemy. γ_1 wzrost IQ mężczyzny o 1 wywołuje wzrost średniej o 0,05, $gamma_2$ nie interpretujemy, γ_3 wartość o jaką różni się przyrost średniej i kobiet wywołany jednostkową zmianą IQ.

Zadanie 3

1. Zmienne objaśniające wyjaśniają wariancję zarobków w 11 %, o czym świadczy wielkość statystyki R^2. Statystycznie istotne są wszystkie zmienne objaśniające, poza zmienną płeć, gdyż statystyki t są większe od 2, a ich *p-value* wynosi 0.

2. Założenie o poprawności formy funkcyjnej można weryfikować testem RESET. Do modelu dodawany jest zbiór dodatkowych zmiennych Z i sprawdzana jest statystyczna istotność współczynników przy dodatkowych zmiennych (oznaczmy je γ) $H_0: \ \gamma = 0$ - forma funkcyjna jest poprawna, $H_A: \ \gamma \neq 0$ - forma funkcyjna jest niepoprawna. Wartość statystyki testowej wynosi 775.40, a jej *p-value* $Prob > F = 0.0000$ wobec tego brak jest podstaw do odrzucenia hipotezy o poprawności formy funkcyjnej.

3. Model spełnia założenie o liniowości (test RESET), nie są spełnione założenie o normalności (test JB), o homoscedastyczności (White,Breusch-Pagan), o braku autokorelacji (DW, B-Godfrey) składnika losowego. Założenie o nielosowości macierzy obserwacji jest nietestowalne.

4. Aby przetestować hipotezę o braku wpływu doświadczenia na wysokość zarobków należy sprawdzić czy suma współczynników przy zmiennych $dosw$ oraz $dosw2$ wynosi zero. Mając oszacowania modelu bez ograniczeń można oszacować model z narzuconymi ograniczeniami, a następnie na podstawie sumy kwadratów reszt obu modeli zbudować statystykę $F = \frac{(RSS_R - RSS_U)}{RSS_U} / \frac{N-k}{J}$ i porównać z wartością krytyczną z rozkładu F.

5. Składnik losowy modelu jest heteroscedastyczny i zawiera autokorelację, więc uzyskane oszacowania MNK nie są zgodne. Należy zastosować odporną metodę estymacji Praisa-Winstena.