imię, nazwisko, nr indeksu:____

Ekonometria — egzamin
wersja Informatyka i Ekonometria
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. Wyprowadź estymator MNK dla modelu z wieloma zmiennymi objaśniającymi.

   Szacujemy model $y = Xb + e \Rightarrow e = y - Xb$

   $RSS = e'e = (y - Xb)'(y - Xb) = (y' - b'X')(y - Xb) = y'y - y'Xb - b'X'y + b'X'Xb$

   Ponieważ $y'Xb$ oraz $b'X'y$ są skalarami otrzymujemy:

   $$RSS = y'y - 2b'X'y + b'X'Xb$$

   Następnie należy zminimalizować sumę kwadrat˘w reszt.

   $$\frac{\partial RSS}{\partial b} = -2X'y + 2X'Xb$$

   $$\frac{\partial^2 RSS}{\partial b \partial b'} = 2X'X$$

   ponieważ macierz X ma pe-en rząd kolumnowy, to macierz $X'X$ jest dodatnio określona więc jest odwracalna i $\frac{\partial RSS}{\partial b}$ jest szukanym minimum. Zapisujemy warunek pierwszego rzędu:

   $$-2X'y + 2X'Xb = 0$$

   $$X'y = X'Xb$$

   mnożymy obie strony równania przez macierz $(X'X)^{-1}$ z lewej strony.

   $$(X'X)^{-1}X'y = (X'X)^{-1}X'Xb$$

   $$b = (X'X)^{-1}X'y$$

2. Co to jest współliniowość i jakie są konsekwencje jej występowania dla własności statystycznych estymatorów MNK w KMRL.

   Współliniowość jest to zależność funkcyjna pomiędzy zmiennymi objaśniającymi zawartymi w modelu ekonometrycznym. Wspóliniowość dzielimy na dokładną, która występuje gdy jedną ze zmiennych objaśniających można przedstawić jako funkcję pozostałych zmiennych objaśniających oraz niedokładną, która występuję gdy jedna ze zmiennych objaśniających jest silnie skorelowana z pozostałymi zmiennymi objaśniającymi.

   Gdy występuje zjawisko współliniowości spełnione założenia KMRL, ale też mogą pojawić się następujące problemy:

   (a) niewielkie zmiany w zbiorze danych powodują duże zmiany w otrzymywanych oszacowaniach parametrów.

   (b) współczynniki równania regresji mają duże błędy standardowe, oraz mogą być pojedynczo nieistotne statystycznie, nawet gdy łącznie są istotne, a współczynnik determinacji modelu R^2 jest wysoki

   (c) współczynniki równania regresji mają „złe”, czyli niezgodne z teorią ekonomiczną znaki, albo są zbyt małe lub zbyt duże.

Pytania teoretyczne c.d.

3	4	$\sum$

3. Podaj postać przekształcenia Boxa-Coxa i wyjaśnij w jakim celu jest ono stosowane w ekonometrii??

   Przekształcenie Boxa-Coxa dane jest wzorem

$$g(x, \lambda) = \frac{x^\lambda - 1}{\lambda}$$

   To przekształcenie pozwala stwierdzić, która forma funkcyjna (liniowa, logarytmiczna, odwrotna) modelu jest najlepiej dopasowana do dostępnego zbioru danych empirycznych.

4. Na czym polega problem równoczesności?

   Problem równoczesności występuje, gdy zmienne objaśniające zawarte w modelu są skorelowane z błędem losowym

$$cov(x_i, \varepsilon_i) \neq 0$$

   W takim przypadku uzyskane oszacowania nie będą zgodne, czyli nawet w przypadku szacowania parametrów na podstawie dużej próby wartości oszacowań mogą być błędne. Problem ten występuje zazwyczaj w wyniku obecności sprzężeń zwrotnych.

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	$\sum$

Zadanie 2 **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.

1	2	3	4	5	$\sum$

ZADANIE 3 Na podstawie danych kwartalnych dla Polski w latach 1995q1 do 2007q2 oszacowano model wyjaśniający poziom spożycia indywidualnego wielkością PKB w danym kwartale. Najpierw oszacowano model na wielkościach nominalnych spożycia i PKB (w mln złotych) i otrzymano następujące wyniki regresji:

	b	se	t	p
pkb	0.606	0.022	28.042	0.000
stała	4173.576	4175.802	0.999	0.323
R^2	0.942			

```
Breusch-Godfrey chi2(1) = 2.555 [0.1100]
   Durbin-Watson (3, 50) = 2.436
  Breusch-Pagan chi2(1) = 6.97[0.0083]
  Ramsey RESET F(3, 45) = 4.50[0.0076]
```

Następnie zdecydowano się oszacować ten sam model, ale dla logarytmów poziomów spożycia i logarytmów realnego PKB (lrspoz_ind, lrpkb), przy czym wartości nominalne zdeflowano za pomocą deflatora CPI. Uzyskano następujące wyniki:

	b	se	t	p
lrpkb	0.763	0.045	16.850	0.000
stała	2.208	0.512	4.317	0.000
R^2	0.8554			

```
Breusch-Godfrey chi2(2) = 7.472 [0.0063]
   Durbin-Watson(3, 50) = 1.298
 Breusch-Pagan chi2(1) = 1.16 [0.2806]
        Ramsey RESET F(3, 45) = 2.79 [0.0515]
```

Na koniec oszacowano ten sam model, ale macierz wariancji i kowariancji oszacowano za pomocą estymatora Newey'a-Westa. Otrzymano następujące oszacowania:

	b	se	t	p
lrpkb	0.763	0.058	13.107	0.000
stała	2.208	0.663	3.329	0.002

UWAGA: Testowanie hipotez należy przeprowadzić przy poziomie istotności $\alpha = 0.05$. Wartości krytyczne statystyki DW dla 50 obserwacji i 2 zmiennych objaśniających i $\alpha = 0.05$ wynoszą: $d_L = 1.50$ i $d_U = 1.58$. W nawiasach kwadratowych pojawiają się wartości p dla danego testu.

1. Proszę wyjaśnić dlaczego, z punktu widzenia teorii estymacji, oszacowania uzyskane na pierwszym etapie zostały uznane za niesatysfakcjonujące.

2. Zbadać, czy wyniki uzyskane w modelu dla zmiennych realnych można uznać za bardziej satysfakcjonujące. Odpowiedź uzasadnij wynikami testów statystycznych.

3. Wyjaśnij, dlaczego zdecydowano się oszacować macierz wariancji i kowariancji inną metodą. Jaki problem mógłby się pojawić, gdyby nie zastosowano tej metody?

4. Podać interpretację oszacowań współczynników (poza stałymi) uzyskanych z modelu oszacowanego na wartościach nominalnych i logarytmach wartości realnych.

5. Policz elastyczność spożycia indywidualnego na podstawie wyników z pierwszej regresji wiedząc, że średni poziom PKB w analizowanym okresie wyniósł 184592,6 mln zł.

IMIĘ I NAZWISKO ____________________

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:

$$RSS^\star = \underset{\min}{b_1, b_2, b_3} \sum_{i=1}^{n} (y_i - b_1 - b_2x_{2i} - b_3x_{3i})^2$$

$$RSS^\star = \underset{\min}{b_2, b_3} \sum_{i=1}^{n} (y_i - 3{,}14 - b_2x_{2i} - b_3x_{3i})^2 = \sum_{i=1}^{n} (y_i - b_1 - b_2x_{2i} - b_3x_{3i})^2$$

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 $RSS^\star \le RSS^{\star\star}$

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:

$$RSS\star \le RSS\star\star \;\Leftrightarrow\; \frac{RSS\star}{TSS} \le \frac{RSS\star\star}{TSS} \;\Leftrightarrow\; 1 - \frac{RSS\star}{TSS} \ge 1 - \frac{RSS\star\star}{TSS} \;\Leftrightarrow\; R^{2\star} \ge R^{2\star\star}$$

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:

$$RSS^\star = \underset{\min}{\gamma_1, \gamma_2, \gamma_3} \sum_{i=1}^{N+1} (y_i - \gamma_1 - \gamma_2x_{2i} - \gamma_3x_{3i})^2 =$$

$$\underset{\min}{\gamma_1, \gamma_2, \gamma_3} \sum_{i=1}^{N} ((y_i - \gamma_1 - \gamma_2x_{2i} - \gamma_3x_{3i})^2 + (y_{N+1} - \gamma_1 - \gamma_2x_{2N+1} - \gamma_3x_{3N+1})^2) =$$

$$\sum_{i=1}^{N} ((y_i - \gamma_1 - \gamma_2x_{2i} - \gamma_3x_{3i})^2 + (\bar{y} - \gamma_1 - \gamma_2\bar{x_2} + \gamma_3\bar{x_3})^2)$$

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. 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 *dosw*2 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.

Zadanie 3

1. W pierwszej regresji występują następujące problemy:

   (a) błędna forma funkcyjna na co wskazuje test RESET (`0.0076`<`0.05`)

   (b) heteroskedastyczność na co wskazuje wynik testu Breuscha-Pagana (`0.0083`<`0.05`)

   Błędna forma funkcyjna implikuje brak zgodności estymatora

2. W drugiej regresji:

   (a) nie ma podstaw do odrzucenia hipotezy o poprawności formy funkcyjnej na co wskazuje wynik testu RESET (`0.0515`>`0.05`)

   (b) nie ma podstaw do odrzucenia hipotezy o homoskedastyczności na co wskazuje wynik testu Breuscha-Pagana (`0.2806`>`0.05`)

   (c) Hipoteza o braku autokorelacji jest odrzucana przez test DW (`1.298`<1.50) a także przez test Breuscha-Godfrey'a (`0.0063`<`0.05`)

3. W trakcie analizowania wyników regresji badacz doszedł do wniosku, że w modelu występuje autokorelacja. W związku z tym zastosowano odporną na autokorelację metodę szacowania macierzy wariancji i kowariancji Newey'a-Westa. W razie, gdyby nie zastosowano tej poprawki oszacowania macierzy wariancji i kowariacji byłyby błędne a tym samym błędne mogłyby wyniki wnioskowania.

4. Pierwszej regresja sugeruje, że wzrost pkb o mln złotych powoduje wzrost spożycia indywidualnego o 0.606 mln złotych. Z drugiej regresji wynika, że wzrost realnego pkb o 1% powoduje wzrost realnego spożycia indywidualnego o 0.763%.

5. Z definicji elastyczność jest równa:

$$e_{y,x_k} = \frac{\partial\,\mathrm{E}\,(y)}{\partial x_k} \bigg/ \frac{y}{x_k} = \frac{x_k}{y}\beta$$

licząc elastyczność dla wielkości średnich otrzymujemy w modelu ze stałą i jedną zmienną objaśniającą:

$$\widehat{e}_{y,x} = \frac{\overline{x}\widehat{\beta}_1}{\overline{y}} = \frac{\overline{x}\widehat{\beta}_1}{\overline{\widehat{y}}} = \frac{\overline{x}\widehat{\beta}}{\widehat{\beta}_0 + \overline{x}_k\widehat{\beta}}$$

a więc $\widehat{e}_{y,x} = \frac{0.606\times184592.6}{4173.576+0.606\times184592.6} = 0.964\,03$