imię, nazwisko, nr indeksu:____

Ekonometria — egzamin
01/02/2019

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, 01/02/2019, ________________________________
podpis

Powodzenia :-)

Wartości krytyczne testu Durbina - Watsona.

n	d_L	d_U
50	1,50	1,59
100	1,65	1,69
200	1,76	1,78

Wartości krytyczne rozkładu χ^2.

liczba stopni swobody	$\alpha = 0{,}95$	$\alpha = 0{,}99$
1	3,84	6,64
2	5,99	9,21
3	7,81	11,35
4	9,49	13,28
5	11,07	15,09
6	12,59	16,81
7	14,07	18,48
8	15,51	20,09
9	16,91	21,67
10	18,31	23,21

Wartości krytyczne rozkładu F.

liczba stopni swobody	$\alpha = 0{,}95$
F(1,20)	4,35
F(2,20)	3,49
F(3,20)	3,10
F(20,1)	248,02
F(20,2)	19,45
F(20,3)	8,66
F(1,500)	3,86
F(2,500)	3,01
F(3,500)	2,62
F(1, ∞)	3,84
F(5, ∞)	2,21

Pytanie 1.

Udowodnij, że w modelu liniowym ze stałą średnia wartość zmiennej zależnej jest równa średniej z wartości dopasowanych.

Najczęstszym błędem był brak odwołania się do własności modelu. Dodatkowo $E(\varepsilon) = 0$ nie implikuje, że $\sum e_i = 0$.

Pytanie 2.

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

Oczekiwałem wskazania na ustalone parametry modelu i składnik (odchylenia) losowe oraz na ich losowe oszacowania.

Pytanie 3.

Wyjaśnij co to jest autokorelacja składnika losowego i opisz jakie są jej konsekwencje dla własności estymatorów MNK

Autokorelacja nie powoduje obciążenia estymatora wektora parametrów, obciążenie estymatora macierzy wariancji-kowariancji wektora parametrów

Pytanie 4.

Opisz korzyści i niebezpieczeństwa związane z nakładaniem ograniczeń na parametry modelu.

W tym pytaniu nie chodziło o to, że korzyścią jest możliwość weryfikacji teorii ekonomii, tylko o opis w jaki sposób nakładanie ograniczeń wpływa na własności statystyczne estymatorów.

1	2	3	$\sum$

Zadanie 1

Niech N oznacza zbiór liczb nieparzystych, oraz P oznacza zbiór liczb parzystych. Dany jest następujący model:

$$\begin{aligned} y_t &= \beta_1 + \beta_2 d_i + \varepsilon_i \quad \text{dla} \quad i = 1, \ldots, 2K \\ d_i &= \begin{cases} 1 & \text{dla} \quad i \in P \\ -1 & \text{dla} \quad i \in N \end{cases} \\ Var\left(\varepsilon_i\right) &= \sigma^2 \boldsymbol{I} \end{aligned}$$

1. Oblicz wartości estymatorów MNK dla parametrów β_1 i β_2 przyjmując, że $K = 50$, $\sum_{i=1}^{2K} y_t = 100$, $\sum_{i \in P} y_i = 60$.
2. Udowodnij, że estymatory MNK są nieobciążone
3. Podaj postać macierzy wariancji-kowariancji dla estymatorów parametrów β_1 i β_2, jeśli spełnione są założenia KMRL, przyjmując, że $\sigma = 1$. Czy składnik losowy jest homoscedastyczny?

IMIĘ I NAZWISKO ____________________

1	2	3	4	$\sum$

Zadanie 2 W modelu liniowym o postaci

$$y_i = \beta x_i + \varepsilon$$

zakładamy, że x_i są nielosowe dla $i = 1, \dots, n,\ E(\varepsilon) = 0,\ var(\varepsilon) = \sigma^2 \mathbb{I}$.

1. Wyprowadź postać estymatora MNK oraz sprawdź czy jest nieobciążony (uzasadnij dokonując obliczeń). Oblicz jego wartość jeżeli $\sum_{i=1}^n x_i y_i = 30,\ \sum_{i=1}^n x_i = 15\ \sum_{i=1}^n x_i^2 = 25\ \sum_{i=1}^n y_i = 13$
2. Pozostajemy przy modelu z pkt (a), ale zakładamy, że $E(\varepsilon) = \theta,\ var(\varepsilon) = \sigma^2 \mathbb{I}$, oraz x_i dla $i = 1, \dots, n$ i θ są nielosowe. Zakładamy, że ε ma rozkład normalny. Czy przy powyższych założeniach estymator MNK dla modelu z pkt 1 będzie nieobciążony? Zaproponuj sposób estymacji, dzięki któremu można uzyskać nieobciążone oszacowanie parametru β (uzasadnij dokonując obliczeń).

Badacz doszedł do wniosku, że lepiej będzie oszacować model o formie

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon$$

zakładamy, że x_i są nielosowe dla $i = 1, \dots, n,\ E(\varepsilon) = 0,\ var(\varepsilon) = \sigma^2 \mathbb{I}$. Z teorii jednak wynika, że $\beta_0 = 4$.

3. Oblicz wartość estymatora MNK dla parametru β_1 w modelu z ograniczeniem $\beta_0 = 4$ dla danych z podpunktu (1). Czy suma reszt będzie równa zero (skomentuj)? Sprawdź od czego zależy obciążenie estymatora. W tym celu oblicz obciążenie, gdy ograniczenie jest prawdziwe, oraz gdy jest ono fałszywe.
4. Oblicz wariancję estymatora β_1 przy założeniu o prawdziwości ograniczenia i porównaj z wariancją estymatora w modelu bez ograniczeń równą

$$\frac{n\sigma^2}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}$$

wiedząc, że $n \sum_{i=1}^n x_i^2 \geq (\sum_{i=1}^n x_i)^2$. Skomentuj uzyskany wynik, czy jest on sprzeczny z twierdzeniem Gaussa-Markowa?

1	2	3	4	5	$\sum$

Zadanie 3 Na podstawie danych pochodzących z Badania Budżetów Gospodarstw Domowych badacze poszukiwali determinantów wysokości wydatków konsumpcyjnych na osobę w gospodarstwie domowym `wk`. Po przeprowadzeniu analizy podobnych badań dla innych krajów europejskich w zbiorze potencjalnych regresorów uwzględnili: dochód gospodarstwa domowego na osobę `doch_os`, `wiek` głowy gospodarstwa domowego i jego interakcje z dochodem, lokalizację gospodarstwa domowego: zmienne 0-1 dla gospodarstw miejskich i gospodarstw z Warszawy (poziom bazowy: wieś), źródło utrzymania gospodarstwa: zmienne 0-1 dla gospodarstw w których głowa pracuje `praca`, oraz głowa pobiera emeryturę `emerytura`, oraz poziom wykształcenia w gospodarstwie: zmienna 0-1 wskazująca na posiadanie przez głowę gospodarstwa wykształcenia wyższego.

Uzyskali następujące wyniki:

```
Number of obs   =  37,189  R-squared       =    0.2756
F(9, 37179)     = 1571.80  Adj R-squared   =    0.2754
Prob > F        =  0.0000  Root MSE        =    725.65
------------------------------------------------------
           wk_os |    Coef.  Std. Err.      t    P>|t|
-----------------+------------------------------------
         doch_os |    0,089     0,006    14,32   0,000
            plec |   30,414     8,002     3,80   0,000
            wiek |   -1,996     0,380    -5,26   0,000
c.doch_os#c.wiek |    0,002     0,000    17,46   0,000
          miasto |  123,858     8,029    15,43   0,000
        warszawa |  485,326    17,849    27,19   0,000
           praca |  -13,723    11,287    -1,22   0,224
       emerytura |  -28,222    14,204    -1,99   0,047
          wyzsze |  423,622    10,324    41,03   0,000
           _cons |  683,757    26,614    25,69   0,000
------------------------------------------------------
RESET test F(10, 37170) =  1115,56  P = 0,0000
Breusch-Pagan chi2(9)   = 1.16e+06  P = 0,0000
Jarque-Bera chi(2)      > 1.00e+16  P = 0,0000
Mean VIF                  2.80
```

gdzie P oznacza wartość P.

Odpowiedz na poniższe pytania, uzasadniając swoje odpowiedzi stosownymi obliczeniami lub wartościami statystyk. Przyjmij poziom istotności 1%.

1. Ile wynosi efekt krańcowy wydatków konsumpcyjnych na osobę względem dochodów na osobę?

2. Ile wynosi oczekiwana różnica w wydatkach konsumpcyjnych na osobę między gospodarstwem z Warszawy a gospodarstwem miejskim.

3. Czy formę funkcyjną modelu można uznać za poprawną?

4. Czy spełnione jest założenie o (hiper)sferyczności składnika losowego?

5. Jakie konsekwencje niesie niespełnienie założeń KMRL w tym modelu i w jaki sposób można rozwiązać ten problem.

Rozwiązania zadań

Zadanie 1

1. Macierz $X'X$ ma postać

$$X'X = \begin{bmatrix} 1\dots & \dots 1 \\ \underbrace{1\dots 1}_{K} & \underbrace{-1\dots -1}_{K} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \\ 1 & -1 \\ \vdots & \vdots \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 2K & 0 \\ 0 & 2K \end{bmatrix}$$

Zatem macierz $(X'X)^{-1}$ ma postać

$$(X'X)^{-1} = \frac{1}{4K^2}\begin{bmatrix} 2K & 0 \\ 0 & 2K \end{bmatrix} = \begin{bmatrix} \frac{1}{2K} & 0 \\ 0 & \frac{1}{2K} \end{bmatrix}$$

Macierz $X'y$ przyjmuje postać

$$X'y = \begin{bmatrix} 1\dots & \dots 1 \\ 1\dots 1 & -1\dots -1 \end{bmatrix} \begin{bmatrix} \sum y_P \\ \sum y_N \end{bmatrix} = \begin{bmatrix} \sum y_t \\ \sum y_P - \sum y_N \end{bmatrix}$$

Zatem

$$b = \begin{bmatrix} \frac{1}{2K} & 0 \\ 0 & \frac{1}{2K} \end{bmatrix} \begin{bmatrix} \sum y_t \\ \sum y_P - \sum y_N \end{bmatrix} = \begin{bmatrix} \frac{\sum y_t}{2K} \\ \frac{\sum y_P - \sum y_N}{2K} \end{bmatrix} = \begin{bmatrix} 1.0 \\ 0.2 \end{bmatrix}$$

2. Standardowy dowód nieobciążoności

$$E(b) = E\left[(X'X)^{-1}X'y\right] = E\left[(X'X)^{-1}X'(X\beta+\varepsilon)\right] = E\left[(X'X)^{-1}X'X\beta\right] = \beta$$

albo

$$E(b) = E\begin{bmatrix} \frac{y_t}{2K} \\ \frac{y_P - y_N}{2K} \end{bmatrix} = \begin{bmatrix} \frac{y_t}{2K} \\ \frac{y_P - y_N}{2K} \end{bmatrix}$$

ponieważ w macierzy nie ma elementów losowych.

3.

$$\sigma^2(X'X)^{-1} = \sigma^2\begin{bmatrix} \frac{1}{2K} & 0 \\ 0 & \frac{1}{2K} \end{bmatrix} = \begin{bmatrix} \frac{1}{100} & 0 \\ 0 & \frac{1}{100} \end{bmatrix}$$

Składnik losowy na mocy założeń modelu ma stałą wariancję, czyli jest homoscedastyczny.

Zadanie 2

1. Macierz $\mathbb{X}$ składa się z jednej kolumny i n wierszy więc estymator MNK:

$$b = \left(\begin{bmatrix} x_1 & \dots & x_n \end{bmatrix} \begin{bmatrix} x_1 \\ \dots \\ x_n \end{bmatrix} \right)^{-1} \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ \dots \\ y_n \end{bmatrix} = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{30}{25} = 1,2$$

$$E(b) = \frac{E(\sum x_i y_i)}{E(\sum x_i^2)} = \frac{\sum x_i E(y_i)}{\sum x_i^2} = \frac{\sum x_i E(\beta x_i + \varepsilon)}{\sum x_i^2} = \beta\frac{\sum x_i^2}{\sum x_i^2}$$

więc estymator jest nieobciążony.

2. W tym przypadku wartość oczekiwana zmiennej zależnej wynosi $E(y_i) = \beta x_i + \theta$. Wobec tego wartość oczekiwana estymatora jest równa

$$E(b) = \frac{E(\sum x_i y_i)}{E(\sum x_i^2)} = \frac{\sum x_i E(y_i)}{\sum x_i^2} = \frac{\beta \sum x_i^2 + \theta \sum x_i}{\sum x_i^2} = \beta + \frac{\theta \sum x_i}{\sum x_i^2}$$

A zatem estymator będzie obciążony. Aby uzyskać nieobciążony estymator np. wystarczy zauważyć, że $\varepsilon_i = \xi_i + \theta$, gdzie $\xi_i \sim \mathcal{N}(0, \sigma^2)$. Wobec tego model można zapisać jako

$$y_i = \beta x_i + \theta + \xi_i$$

Ten model spełnia założenia modelu KMRL dla modelu ze stałą (θ), więc jego oszacowania będą nieobciążone.

3. Minimalizowana jest suma kwadratów reszt przy ograniczeniu $b_0 = 4$, czyli $S_R = \sum(y_i - 4 - b_1 x_i)^2$. Pochodna względem b_1 jest równa:

$$\frac{\partial S}{\partial b_1} = -2\sum(y_i - 4 - b_1 x_i)x_i = 0$$

więc

$$\sum y_i x_i - 4\sum x_i - b_1 \sum x_i^2 = 0$$

wobec tego

$$b_1 = \frac{\sum y_i x_i - 4\sum x_i}{\sum x_i^2} = \frac{30 - 4 * 15}{25} = -1.2$$

Suma reszt w tym modelu nie będzie równa zero

$$\sum e_i = \sum y_i - 4 * n + 1.2\sum x_i = 13 - 4 * n + 1.2 * 15 = 31 - 4 * n$$

Pomimo tego, iż w modelu jest stała, to zostało na nią narzucone ograniczenie, przez co przy szukaniu minimum sumy kwadratów reszt nie znajdujemy optimum, w związku z tym suma reszt nie musi być równa zero (inaczej: narzucenie ograniczenia na stałą sprowadza model do modelu bez stałej).

Przy prawdziwym ograniczeniu wartość oczekiwana zmiennej objaśnianej jest równa

$$E(b_1) = \frac{\sum E(y_i)x_i - 4\sum x_i}{\sum x_i^2} = \frac{4\sum x_i + \beta \sum x_i^2 - 4\sum x_i}{\sum x_i^2} = \beta_1$$

Przy fałszywym ograniczeniu wartość oczekiwana zmiennej zależnej będzie równa

$$E(b_1) = \frac{\sum E(y_i)x_i - 4\sum x_i}{\sum x_i^2} = \frac{\beta_0 \sum x_i + \beta_1 * \sum x_i^2 - 4\sum x_i}{\sum x_i^2} = \beta_1 + \frac{(\beta_0 - 4)\sum x_i}{\sum x_i^2}$$

czyli estymator będzie obciążony. Obciążenie będzie proporcjonalne do różnicy pomiędzy rzeczywistą a zakładaną wartością parametru β_0.

4. Wariancja estymatora b_1 jest równa

$$var(b_1) = var\Big(\frac{\sum y_i x_i - 4\sum x_i}{\sum x_i^2}\Big) = var\Big(\frac{\sum(4 + \beta_1 x_i + \varepsilon_i)x_i}{\sum x_i^2}\Big) = var\Big(\frac{\sum \varepsilon_i x_i}{\sum x_i^2}\Big)$$

$$var(b_1) = \frac{\sum x_i^2 var(\varepsilon_i)}{(\sum x_i^2)^2} = \frac{\sigma^2}{(\sum x_i^2)}$$

Czyli wariancja w modelu z ograniczeniami jest mniejsza lub równa wariancji dla modelu bez ograniczeń. Nie jest to sprzeczne z twierdzeniem Gaussa-Markowa, ponieważ podczas konstrukcji modelu z ograniczeniami wykorzystywane są dodatkowe informacje pochodzące spoza próby.

Zadanie 3

1.
$$\frac{\partial Ewk\_os}{\partial doch\_os} = \beta_{doch\_os} + \beta_{c.doch\_os\#\#c.wiek}\bar{wiek} = 0{,}089 + 0{,}002\bar{wiek}$$

2.
$$E(wk\_os|warszawa) - E(wk\_os|miasto) = \beta_{warszawa} - \beta_{miasto} = 485{,}326 - 123{,}858 = 361{,}468$$

   Oczekiwana różnica w wydatkach konsumpcyjnych na osobę między gospodarstwami z Warszawy a miejskimi wynosi 361,468 zł.

   UWAGA: Uznawane były inne odpowiedzi, jeżeli wskazane były przyjęte założenia i przy tych założeniach obliczenia były poprawne

3. Co prawda wynik testu RESET (*p-value*=0) wskazuje na niepoprawną formę funkcyjną, lecz biorąc pod uwagę, że model został skonstruowany na podstawie teorii oraz szacowany jest na podstawie próby o dużej liczbie obserwacji formę funkcyjną należy uznać za poprawną.

4. Wynik testu Breuscha-Pagana (*p-value*=0) wskazuje na heteroscedastyczność składnika losowego, zatem składnik losowy nie jest hipersferyczny

5. Konsekwencją heteroscedastyczności są nieprawidłowe oszacowania błędów standardowych, zatem wyniki testów istotności t oraz F mogą być zniekształcone. Aby temu zaradzić należy wykorzystać odporny na heteroscedastyczność estymator macierzy wariancji-kowariancji, np. estymator White'a.