

imię, nazwisko, nr indeksu: _____

Ekonometria — egzamin
semestr pierwszy
14/06/09

1. Przed przystąpieniem do pisania egzaminu należy podpisać **wszystkie kartki** arkusza egzaminacyjnego (na dole w przewidzianym miejscu). Złożenie podpisu pod regulaminem oznacza jego akceptację. Do egzaminu mogą przystąpić osoby, które akceptują regulamin.
2. 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.
3. Egzamin trwa 90 minut.
4. Rozwiązywanie zadań należy rozpocząć po ogłoszeniu początku egzaminu a zakończyć wraz z ogłoszeniem końca egzaminu. Złamanie tej zasady skutkuje usunięciem z sali i unieważnieniem arkusza egzaminacyjnego.
5. W trakcie egzaminu wolno używać jedynie długopisu o innym kolorze atramentu niż czerwony oraz kalkulatora.
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. 33,33 % punktów z części zadaniowej.

Regulamin zrozumiałam(em)

Warszawa, 14/06/2009, _____

podpis

Powodzenia :-)

Użyteczne wartości krytyczne

$\chi^2(1) = 3,84$
$\chi^2(2) = 5,99$
$\chi^2(3) = 7,81$
$\chi^2(4) = 9,49$
$\chi^2(5) = 11,07$
$\chi^2(6) = 12,59$
$\chi^2(7) = 14,07$
$\chi^2(8) = 15,51$
$\chi^2(9) = 16,92$
$\chi^2(10) = 18,31$
\vdots
$\chi^2(99) = 123,22$
$\chi^2(100) = 124,34$

Wartości krytyczne testu Durбина - Watsona.

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

Pytanie 1.

Podaj wzajemne relacje między wartościami obserwowanymi zmiennej zależnej, oszacowaniami parametrów, wartościami dopasowanymi i resztami.

Pytanie 2.

Wyprowadź postać macierzy wariancji kowariancji estymatorów b i podaj interpretację jej elementów.

Pytanie 3.

Za pomocą jakich testów weryfikujemy stabilność parametrów? Jakiemu założeniu KMRL odpowiada H_0 w tych testach? Jakie są hipotezy alternatywne w tych testach?

Pytanie 4.

Kiedy mówimy, że w modelu występuje problem równoczesności? Jakie są jej dwie najczęstsze przyczyny i jakie ma ona konsekwencje dla własności estymatorów MNK?

1	2	3	4	Σ

ZADANIE 1 Dany jest następujący model konsumpcji szacowany na podstawie danych przekrojowych

$$kons = \beta_0 + \beta_1 ldochg + \beta_2 los + \beta_3 ld + \beta_4 miasto + \varepsilon$$

gdzie *kons* jest logarytmem wartości konsumpcji gospodarstwa domowego, *ldochg* logarytmem dochodu gospodarstwa domowego, *los* liczbą osób w gospodarstwie domowym, *ld* liczbą osób dorosłych w gospodarstwie domowym, a *miasto* zmienną zero jedynekową przyjmującą wartość 1 dla gospodarstw zlokalizowanych w miastach.

1. Jakie ograniczenia muszą spełniać parametry modelu aby konsumpcja na osobę była uzależniona wyłącznie od dochodu i liczby dzieci w gospodarstwie domowym.
2. Jakie ograniczenia muszą być spełnione aby poziom konsumpcji zależał wyłącznie od liczby osób dorosłych i miejsca zamieszkania gospodarstwa.

Na podstawie danych z 520 gospodarstw oszacowano pełen model, oraz modele z punktu 1 oraz 2 uzyskując statystyki *F* odpowiednio 930 dla pełnego modelu, 740 dla modelu z punktu 1, oraz 530 dla modelu z punktu 2. Wartość krytyczna z rozkładu wynosi $F^* = 4,43$

3. Oceń, który z modeli jest w najlepszym stopniu dopasowany do danych empirycznych.
4. Sprawdź czy ograniczenia z pkt (1) oraz (2) są spełnione.

1	2	3	4	5	Σ

ZADANIE 2 Na podstawie próbki z badania BAEL zbudowano model zarobków. Logarytm zarobków (zmienna *lzarobki*) powiązано z płcią respondenta - *plec* (1 - kobieta, 0 - mężczyzna), jego wiekiem - *wiek*, oraz kwadratem wieku - *wiek2*; poziomem wykształceniem (poziom bazowy wykształcenie wyższe), oraz miejscem zamieszkania zmienną *k1m* o wartości 0 dla osób mieszkających w mieście pow 100 tys mieszkańców, wyższa cyfra oznacza mniejsze miasto, a 9 oznacza respondentów mieszkających na wsi

Source	SS	df	MS	Number of obs =	179
Model	9.43884787	14	.674203419	F(14, 164) =	5.98
Residual	18.4803748	164	.112685212	Prob > F =	0.0000
-----				R-squared =	0.3381
-----				Adj R-squared =	0.2816
Total	27.9192227	178	.156849566	Root MSE =	.33569

lzarobki	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
<i>plec</i>	-.3384568	.0566641	-5.97	0.000	-.4503421	-.2265715
<i>wiek</i>	.0289047	.0176726	1.64	0.104	-.0059904	.0637998
<i>wiek2</i>	-.0002773	.0002236	-1.24	0.217	-.0007188	.0001643
<i>_Iwyksztal^2</i>	-.3264285	.1523042	-2.14	0.034	-.6271584	-.0256986
<i>_Iwyksztal^3</i>	-.2840807	.0918304	-3.09	0.002	-.465403	-.1027584
<i>_Iwyksztal^4</i>	-.1481433	.1146872	-1.29	0.198	-.3745971	.0783105
<i>_Iwyksztal^5</i>	-.4155959	.0894161	-4.65	0.000	-.5921511	-.2390406
<i>_Iwyksztal^6</i>	-.504672	.1034214	-4.88	0.000	-.7088812	-.3004628
<i>_Iklm_12_1</i>	-.1375864	.082137	-1.68	0.096	-.2997687	.0245959
<i>_Iklm_12_2</i>	-.1929347	.0937871	-2.06	0.041	-.3781206	-.0077488
<i>_Iklm_12_3</i>	-.1130659	.1133965	-1.00	0.320	-.3369712	.1108393
<i>_Iklm_12_4</i>	-.2032006	.1233155	-1.65	0.101	-.4466913	.0402901
<i>_Iklm_12_5</i>	-.2707854	.1772258	-1.53	0.128	-.6207238	.0791531
<i>_Iklm_12_9</i>	-.2191738	.0671312	-3.26	0.001	-.3517268	-.0866209
<i>_cons</i>	5.808663	.3509937	16.55	0.000	5.115614	6.501712

Ramsey RESET test F(5, 160) = 0.38 Prob > F = 0.8630
 Breusch-Pagan test chi2(15) = 1173.91 Prob > chi2 = 0.0000

Na podstawie uzyskanych wyników udziel odpowiedzi na poniższe pytania, odpowiednio je uzasadniając wynikami testów. Przyjmij poziom istotności 5%.

1. Oceń łączną istotność parametrów modelu, oraz wypisz zmienne statystycznie istotne
2. Zinterpretuj parametr dla płci i klasy miejscowości oznaczającej wieś
3. Oceń dopasowanie modelu do danych empirycznych
4. Ocen poprawność funkcyjna modelu.
5. Jakie właściwości ma składnik losowy i jakie ma to konsekwencje dla poprawności oszacowań?

ZADANIE 3 Estymowany jest model bez stałej i z jedną zmienną niezależną

$$y_i = \beta x_i + \varepsilon_i$$

Przyjęto, że zmienna x_i jest nielosowa oraz, że błąd losowy nie podlega autokorelacji. Dane, na których estymowany jest model można podzielić na dwie podpróbki, dla których błąd losowy ma różną wariancję:

$$\text{Var}(\varepsilon_i) = \begin{cases} \sigma_1^2 & \text{dla } i = 1, \dots, n_1 \\ \sigma_2^2 & \text{dla } i = n_1 + 1, \dots, n_1 + n_2 \end{cases}$$

1. Pokaż, że estymator UMNK dla tego modelu ma postać $b_{UMNK} = b_1 w_1 + b_2 w_2$, gdzie b_1 i b_2 są estymatorami MNK policzonymi odpowiednio dla $i = 1, \dots, n_1$ oraz $i = n_1 + 1, \dots, n_1 + n_2$ a $w_1 = \frac{\gamma_1}{\gamma_1 + \gamma_2}$, $w_2 = \frac{\gamma_2}{\gamma_1 + \gamma_2}$ i $\gamma_1 = \frac{1}{\text{Var}(b_1)}$, $\gamma_2 = \frac{1}{\text{Var}(b_2)}$.
2. Pokaż, że wyprowadzony estymator UMNK jest nieobciążony.
3. Pokaż, że wariancja estymatora MNK zastosowanego do całej próby ma postać:

$$\text{Var}(b_{MNK}) = \frac{\sigma_1^4 \gamma_1 + \sigma_2^4 \gamma_2}{(\sigma_1^2 \gamma_1 + \sigma_2^2 \gamma_2)^2}$$

a estymatora UMNK

$$\text{Var}(b_{UMNK}) = \frac{1}{\gamma_1 + \gamma_2}$$

4. Pokaż, że

$$\text{Var}(b_{MNK}) \geq \text{Var}(b_{UMNK})$$

Jakie twierdzenie podaje w ogólniejszej formie wniosek, który wyprowadziliśmy dla tego szczególnego przypadku?

Rozwiązania zadań

Zadanie 1

1. Ponieważ model jest wyrażony w logarytmach zmiennych, aby przejść na wartości na osobę wystarczy przenieść zmienną *los* na lewą stronę równania, pamiętając że $\beta_2 = 1$. Drugim ograniczeniem jest $\beta_4 = 0$.
2. W tym przypadku $\beta_1=0$, oraz $\beta_2 = 0$, ponieważ odpowiednie zmienne mają nie być związane ze zmienną zależną.
3. Należy skorzystać z zależności

$$R^2 = \frac{\frac{k-1}{N-k}F}{1 + \frac{k-1}{N-k}F}$$

- W przypadku pełnego modelu $N = 520$, $k = 5$, $F = 930$, zatem $R^2 \approx 0,88$
- Dla modelu z pkt (1) $N = 520$, $k = 3$, $F = 740$, zatem $R^2 \approx 0,74$
- Dla modelu z pkt (2) $N = 520$, $k = 3$, $F = 530$, zatem $R^2 \approx 0,67$

Zatem najlepiej dopasowany do danych empirycznych jest pełny model.

4. W celu sprawdzenia ograniczeń można posłużyć się wynikami z poprzedniego podpunktu

$$F = \frac{(RSS_R - RSS_U)/J}{RSS_U/(N - k)} = \frac{(R_U^2 - R_R^2)/J}{R_U^2/(N - k)}$$

- Dla modelu z pkt (1)

$$F = \frac{(0,88 - 0,74)/2}{0,88/47} \approx 4,30 < F^* = 4,43$$

Zatem ograniczenia są prawdziwe

- Dla modelu z pkt (2)

$$F = \frac{(0,88 - 0,74)/2}{0,88/47} \approx 6,45 > F^* = 4,43$$

Zatem ograniczenia nie są prawdziwe

Zadanie 2

- (a) Parametry modelu są łącznie istotne, ponieważ *p-value* statystyki F wynosi $0 < \alpha = 0,05$. Pojedynczo istotne są oszacowania parametrów przy zmiennych: płeć *p-value* $0 < \alpha$, Iwysztal2 *p-value* $0,034 < \alpha$; Iwysztal3 *p-value* $0,002 < \alpha$; Iwysztal5 *p-value* $0 < \alpha$; Iwysztal6 *p-value* $0 < \alpha$; Ikml12-2 *p-value* $0,041 < \alpha$; Ikml12-9 *p-value* $0,001 < \alpha$; oraz stała *p-value* $0 < \alpha$;
- (b) Kobiety zarabiają przeciętnie 33 % mniej niż mężczyźni przy innych czynnikach stałych; Osoby mieszkające na wsi zarabiają przeciętnie 22% mniej niż mieszkańcy dużych miast.
- (c) Model jest dobrze dopasowany do danych empirycznych, świadczy o tym wysoka i istotna statystyka F , oraz $R^2 = 0,33$ oznaczające że ok $\frac{1}{3}$ zróżnicowania zarobków jest tłumaczona za pomocą zmiennych uwzględnionych w modelu.
- (d) Na podstawie *p-value* statystyki testu RESET równego $0,86 > \alpha$ brak jest podstaw do odrzucenia hipotezy o prawidłowości formy funkcyjnej, zatem należy ją uznać za poprawną.

- (e) Na podstawie wyniku testu B-P widać, że składnik losowy jest heteroscedastyczny, ponieważ $p=$ value statystyki testowej $0 < \alpha$ więc należy odrzucić hipotezę o stałej wariancji składnika losowego. Wobec tego oszacowania parametrów są prawidłowe, jednak oszacowania wariancji parametrów nie są prawidłowe.

Zadanie 3

- (a) Estymatory MNK dla próbek mają postać $b_1 = \frac{\sum_{i=1}^{n_1} x_i y_i}{\sum_{i=1}^{n_1} x_i^2}$, $b_2 = \frac{\sum_{i=n_1+1}^{n_1+n_2} x_i y_i}{\sum_{i=n_1+1}^{n_1+n_2} x_i^2}$ a ich wariancje mają postać $\text{Var}(b_1) = \frac{\sigma_1^2}{\sum_{i=1}^{n_1} x_i^2}$, $\text{Var}(b_2) = \frac{\sigma_2^2}{\sum_{i=n_1+1}^{n_1+n_2} x_i^2}$. Wynika z tego, że $\gamma_1 = \frac{1}{\sigma_1^2} \sum_{i=1}^{n_1} x_i^2$, $\gamma_2 = \frac{1}{\sigma_2^2} \sum_{i=n_1+1}^{n_1+n_2} x_i^2$.

Macierz $\text{Var}(\epsilon) = \Omega$ oraz macierz Ω^{-1} mają postać:

$$\Omega = \begin{bmatrix} \sigma_1^2 \mathbf{I} & 0 \\ 0 & \sigma_2^2 \mathbf{I} \end{bmatrix}, \Omega^{-1} = \begin{bmatrix} \frac{1}{\sigma_1^2} \mathbf{I} & 0 \\ 0 & \frac{1}{\sigma_2^2} \mathbf{I} \end{bmatrix}$$

a estymator UMNK dla tak przekształconych zmiennych sprowadza się do estymatora MNK

$$\begin{aligned} b_{UMNK} &= (\mathbf{X}\Omega^{-1}\mathbf{X})^{-1} \mathbf{X}\Omega^{-1}\mathbf{y} = \frac{\frac{1}{\sigma_1^2} \sum_{i=1}^{n_1} x_i y_i + \frac{1}{\sigma_2^2} \sum_{i=n_1+1}^{n_1+n_2} x_i y_i}{\frac{1}{\sigma_1^2} \sum_{i=1}^{n_1} x_i^2 + \frac{1}{\sigma_2^2} \sum_{i=n_1+1}^{n_1+n_2} x_i^2} \\ &= \frac{\sum_{i=1}^{n_1} x_i y_i}{\sum_{i=1}^{n_1} x_i^2} \frac{\frac{1}{\sigma_1^2} \sum_{i=1}^{n_1} x_i^2}{\frac{1}{\sigma_1^2} \sum_{i=1}^{n_1} x_i^2 + \frac{1}{\sigma_2^2} \sum_{i=n_1+1}^{n_1+n_2} x_i^2} + \frac{\sum_{i=n_1+1}^{n_1+n_2} x_i y_i}{\sum_{i=n_1+1}^{n_1+n_2} x_i^2} \frac{\frac{1}{\sigma_2^2} \sum_{i=n_1+1}^{n_1+n_2} x_i^2}{\frac{1}{\sigma_1^2} \sum_{i=1}^{n_1} x_i^2 + \frac{1}{\sigma_2^2} \sum_{i=n_1+1}^{n_1+n_2} x_i^2} \\ &= b_1 \frac{\gamma_1}{\gamma_1 + \gamma_2} + b_2 \frac{\gamma_2}{\gamma_1 + \gamma_2} = b_1 w_1 + b_2 w_2 \end{aligned}$$

- (b) Wartość oczekiwana estymatora UMNK jest więc następująca:

$$E(b_{UMNK}) = E(b_1 w_1 + b_2 w_2) = \underbrace{E(b_1)}_{\beta} w_1 + \underbrace{E(b_2)}_{\beta} w_2 = \beta (w_1 + w_2) = \beta$$

a więc jest rzeczywiście nieobciążony.

- (c) Zauważmy, że $\text{Var}(y) = \text{Var}(x_i \beta + \epsilon_i) = \text{Var}(\epsilon_i)$. Wariancję można uzyskać bezpośrednio z formy estymatora MNK:

$$\text{Var}\left(\frac{\sum_{i=1}^{n_1} x_i y_i}{\sum_{i=1}^{n_1} x_i^2}\right) = \frac{\sum_{i=1}^{n_1} x_i^2 \text{Var}(y_i)}{(\sum_{i=1}^{n_1} x_i^2)^2} = \frac{\sigma_1^2 \sum_{i=1}^{n_1} x_i^2 + \sigma_2^2 \sum_{i=n_1+1}^{n_1+n_2} x_i^2}{(\sum_{i=1}^{n_1+n_2} x_i^2)^2} = \frac{\sigma_1^4 \gamma_1 + \sigma_2^4 \gamma_2}{(\sigma_1^2 \gamma_1 + \sigma_2^2 \gamma_2)^2}$$

Wariancja estymatora UMNK jest równa

$$\begin{aligned} \text{Var}(b_{UMNK}) &= \text{Var}(b_1 w_1 + b_2 w_2) = \text{Var}(b_1) w_1^2 + \text{Var}(b_2) w_2^2 \\ &= \frac{1}{\gamma_1} \left(\frac{\gamma_1}{\gamma_1 + \gamma_2}\right)^2 + \frac{1}{\gamma_2} \left(\frac{\gamma_2}{\gamma_1 + \gamma_2}\right)^2 = \frac{\gamma_1 + \gamma_2}{(\gamma_1 + \gamma_2)^2} = \frac{1}{\gamma_1 + \gamma_2} \end{aligned}$$

- (d) Wariancja MNK przekracza wariancję UMNK jeśli

$$\begin{aligned} \frac{\sigma_1^4 \gamma_1 + \sigma_2^4 \gamma_2}{(\sigma_1^2 \gamma_1 + \sigma_2^2 \gamma_2)^2} &\geq \frac{1}{\gamma_1 + \gamma_2} \iff \\ (\sigma_1^4 \gamma_1 + \sigma_2^4 \gamma_2) (\gamma_1 + \gamma_2) &\geq (\sigma_1^2 \gamma_1 + \sigma_2^2 \gamma_2)^2 \iff \\ \sigma_1^4 \gamma_1^2 + \sigma_2^4 \gamma_2^2 + (\sigma_1^4 + \sigma_2^4) \gamma_1 \gamma_2 &\geq \sigma_1^4 \gamma_1^2 + \sigma_2^4 \gamma_2^2 + 2\sigma_1^2 \sigma_2^2 \gamma_1 \gamma_2 \iff \\ \sigma_1^4 + \sigma_2^4 &\geq 2\sigma_1^2 \sigma_2^2 \iff \\ \sigma_1^4 + \sigma_2^4 - 2\sigma_1^2 \sigma_2^2 &= (\sigma_1^2 - \sigma_2^2)^2 \geq 0 \end{aligned}$$

Ostatnia nierówność jest oczywista.