Egzamin z ekonometrii 31.01.2007
I semestr

Regulamin egzaminu

Pytania teoretyczne

1. Wyjaśnić, dlaczego R^2 nie można używać do porównywania modeli.
2. Udowodnić, że w KMRL estymator $\boldsymbol{b}$ jest nieobciążony.
3. Parametry przy zmiennych x_1 i x_2 są dodatnie. Zmienne są ujemnie skorelowane. Jaki będzie wpływ pominięcia zmiennej x_1 na oszacowanie parametru przy zmiennej x_2?
4. Jak niesferyczność będów losowych wpływa na własności MNK?

ZADANIE 1 Dla modelu

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon},$$

stworzyliśmy macierz $\boldsymbol{X}^* = \boldsymbol{XA}$, gdzie $\boldsymbol{A}$ jest pewną macierzą nieosobliwą. Udowodnić, że jeśli oszacujemy regresją $\boldsymbol{y}$ na $\boldsymbol{X}^*$ to:

1. $\boldsymbol{b}^* = \boldsymbol{A}^{-1}\boldsymbol{b}$, gdzie $\boldsymbol{b}^*$ pochodzi z regresji $\boldsymbol{y}$ na $\boldsymbol{X}^*$ a $\boldsymbol{b}$ z regresji $\boldsymbol{y}$ na $\boldsymbol{X}$.
2. Policzyć $\operatorname{Var}(\boldsymbol{b}^*)$.
3. Pokazać, że R^2 otrzymane w obu regresjach jest identyczne.
4. Skomentować wynik z punktu (3).

Rozwiązanie:

1. Estymator $\boldsymbol{b}^*$ jest równy

$$\begin{aligned}\boldsymbol{b}^* &= \left(\boldsymbol{A}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{A}\right)^{-1}\boldsymbol{A}'\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{A}^{-1}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{A}'^{-1}\boldsymbol{A}'\boldsymbol{X}'\boldsymbol{y}\\ &= \boldsymbol{A}^{-1}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{A}^{-1}\boldsymbol{b}\end{aligned}$$

2. $\operatorname{Var}(\boldsymbol{b}^*) = \operatorname{Var}\left(\boldsymbol{A}^{-1}\boldsymbol{b}\right) = \boldsymbol{A}^{-1}\operatorname{Var}(\boldsymbol{b})\,\boldsymbol{A}^{-1'} = \sigma^2\boldsymbol{A}^{-1}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{A}^{-1'}$
3. Reszty z drugiej regresji są równe

$$\boldsymbol{e}^* = \boldsymbol{y} - \boldsymbol{X}^*\boldsymbol{b}^* = \boldsymbol{y} - \boldsymbol{XAA}^{-1}\boldsymbol{b} = \boldsymbol{y} - \boldsymbol{Xb} = \boldsymbol{e}$$

   Ponieważ zmienna zależna jest ta sama i reszty są te same więc R^2 muszą być także identyczne.
4. Jednoznaczne liniowe przekształcenia zmiennych nie mają wpływu na dopasowanie.

ZADANIE 2 Oszacowano regresję stopy wzrostu nakładów w wyrażeniu *realnym* (zmienna y) na stopie procentowej (zmienna x_1) i inflacji (zmienna x_2). Regresję przeprowadzono na danych kwartalnych dla Polski z okresu 1995-2002. Wprowadzono także do modelu zmienne sezonowe przyjmujące wartość 1 dla odpowiedniego kwartału, a 0 dla pozostałych. Wyniki regresji znajdują się w tabeli poniżej.

naklady	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
stopa	-2.403391	.4816086	-4.99	0.000	-3.393352	-1.413431
inf	4.336979	.7105403	6.10	0.000	2.876443	5.797516
_Iq_2	-1.760541	2.614987	-0.67	0.507	-7.135725	3.614642

```
      _Iq_3 |   -1.06475   2.613665    -0.41   0.687    -6.437216    4.307716
      _Iq_4 |   .3517396   2.616217     0.13   0.894    -5.025972    5.729451
      _cons |   7.183309   3.081168     2.33   0.028     .8498766    13.51674
------------------------------------------------------------------------------

Durbin-Watson (6,32)      =   .78506
Breusch-Godfrey, chi2(2) = 10.305 [0.006]
Breusch-Pagan, chi2(1)   =  0.420 [0.517]
Ramsey RESET, F(3, 23)   =  1.060 [0.385]
Jarque-Bera,  chi2(2)    =  1.239 [0.538]
Chow, t=1999.1, F(4,22)  =  1.500 [0.236]
```

Przy założonym poziomie istotności $\alpha = 0.01$ przeprowadzić analizę wyników. Każdą z odpowiedzi należy uzasadnić za pomocą odpowiedniego testu.
Podpowiedź: wartości d_L i d_U dla testu DW przy 32 obserwacjach, 5 zmiennych i stałej oraz $\alpha = 0.01$ wynoszą $d_L = 0.917, d_U = 1.597$.

1. Określić, czy model jest dobrze dopasowany oraz 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 modelu są stabilne.
8. Zinterpretować współczynnik przy zmiennej x_1.
9. W jaki sposób należałoby zweryfikować hipotezę o neutralności pieniądza w gospodarce (tzn., że zmienne nominalne nie wpływają na zmienne realne)?
10. W jaki sposób należałoby weryfikować hipotezę mówiącą o tym, że zmienne sezonowe są nieistotne w modelu?
11. Jeśli model nie spełnia założeń KMRL określić:
    (a) które założenie nie jest spełnione?
    (b) jakie ma to konsekwencje dla wnioskowania statystycznego?
    (c) jakie są metody radzenia sobie z tym problemem?

Rozwiązanie:

1. Unieważnione
2. Istotnymi zmiennymi objaśniającymi jest stopa i inflacja (*p-value* dla statystyk t są mniejsze od 0.01).
3. W modelu występuje dodatnia autokorelacja I rzędu ponieważ statystyka $DW \approx 0.78 \in (0, 917)$ a więc przyjmuje wartość, dla której odrzucamy hipotezę zerową. Odpowiednie wartości krytyczne dla tego testu wynoszą ($d_L = 0.917, d_U = 1.597$). H_0 o braku autokorelacji odrzucamy także na podstawie *p-value* dla testu Breuscha-Godfreya $[.006 < 0.01]$.
4. Nie ma podstaw do odrzucenia hipotezy o homoskedastyczności składnika losowego w związku z wartością *p-value* dla testu Breuscha-Godfreya $[0.517 > 0.01]$.
5. Nie ma podstaw do odrzucenia H_0 o poprawności specyfikacji modelu, o czym świadczy *p-value* dla testu RESET $[.385 > 0.01]$.

6. Nie ma podstaw do odrzucenia hipotezy o normalności rozkładu błędu losowego, o czym świadczy *p-value* dla testu Jarque'a-Bera $[.538 > 0.01]$.

7. Nie ma podstaw do odrzucenia hipotezy o stabilności parametrów modelu, o czym świadczy *p-value* dla testu Chowa $[.236 > 0.01]$.

8. Współczynnik przy zmiennej x_1 mówi o ile punktów procentowych zmaleje stopa wzrostu nakładów jeśli stopa procentowa wzrośnie o 1 punkt procentowy. A więc zgodnie z oszacowaniem wzrost stopy procentowej o 1 punkt procentowy powoduje spadek stopy wzrostu nakładów o 2.4 punkta procentowego.

9. Jeśli rzeczywiście zmienne nominalne nie wpływają na zmienne realne, to na wzrost nakładów w wyrażeniu relanym powinna wpływać jedynie realna stopa procentowa. Jeśli tak jest to

$$\begin{aligned} nakady_t &= \beta_0 + \beta_1 stopa_t + \beta_2 inf_t + \sum_{s=1}^{3} \gamma_s q_{st} + \varepsilon_i \\ &= \beta_0 + \beta_1 (stopa_t - inf_t) + \sum_{s=1}^{3} \gamma_s q_{st} + \varepsilon_i \end{aligned}$$

   ponieważ relana stopas procentowa jest (w przybliżeniu) równa różnicy między stopą nominalną i inflacją. Ograniczenie to będzie prawdziwe, jeśli $\beta_1 = -\beta_2$.

10. Hipotezę o łącznej nieistotności zmiennych sezonowych należy testować za pomocą testu F, nakładając łączne ograniczenie na parametry przy zmiennych $seasonal_2$, $seasonal_3$, $seasonal_4$, - przyrównując je jednocześnie do zera.

11. (a) Model nie spełnia założenia braku autokorelacji składnika losowego.
    (b) Oszacowanie macierzy wariancji-kowariancji MNK jest obciążone, co powoduje, że niewłaściwe są błędy standardowe parametrów i tym samym niewłaściwe są statystyki t.
    (c) Jeśli nie chcemy usunąć autokorelacji, należy posłużyć się tzw. odporną macierzą wariancji-kowariancji Newey'a-Westa. Można też podjąć próbę usunięcia autokorelacji za pomocą Stosowalnej UMNK.

ZADANIE 3 Można pokazać, że w modelu Solowa w stanie stacjonarnym dochód na zatrudnionego jest dany następującym równaniem:

$$\left(\frac{Y_t}{L_t}\right)^* = A_t \left(\frac{s}{g+\delta+n}\right)^{\frac{\alpha}{1-\alpha}}$$

gdzie Y_i dochód narodowy, L_i siła robocza, s stopa oszczędności, g stopa wzrostu technologicznego, δ stopa deprecjacji, n - stopa wzrostu siły roboczej, $A_t = Ae^{a+gt}$. Po zlogarytmowaniu lewej i prawej strony otrzymujemy:

$$\ln\left(\frac{Y_t}{L_t}\right) = \ln A + a + gt + \left(\frac{\alpha}{1-\alpha}\right)\ln(s) - \left(\frac{\alpha}{1-\alpha}\right)\ln(n+g+\delta)$$

Załóżmy, że $g + \delta = 0.05$. Mankiw, Romer i Weil (1992) zaproponowali, by parametry tego równania oszacować na postawie próby przekrojowej, przy założeniu, że s, n i A_i różnią się między krajami, natomiast g, δ i a są takie same dla wszystkich krajów. Oszacowano równanie:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i \qquad (*)$$

gdzie $y_i = \ln\left(\frac{Y_i}{L_i}\right)$, $x_{1i} = \ln(s_i)$, $x_{2i} = \ln(n_i + 0.05)$, $\varepsilon_i = \ln(A_i) + \eta_i$. Na podstawie danych dotyczących 182 krajów z Penn Word Tables 6.2 z roku 2003 uzyskano następujące oszacowania:

zmienna	**współczynnik**	t	Prob($\|t\| > t^*$)
$\ln(s_i)$	.716	6.14	0.000
$\ln(n_i + 0.05)$	-2.431	-5.62	0.000
stała	.0976	0.09	0.931

Regresja	RSS
y_i na stałej, x_{1i} i x_{2i}	163.3
y_i na stałej i x_{1i}	192.1
y_i na $x_{1i} - x_{2i}$	221.4
y_i na stałej i $x_{1i} - x_{2i}$	174.7

1. Czy znaki uzyskane w regresji są zgodne z teorią? Odpowiedź uzasadnić.
2. Jakie ograniczenie wynika z teorii w odniesieniu do parametrów modelu? Zweryfikować to ograniczenie na podstawie wyników regresji.

   Podpowiedź: $F_{0.01}(1,179) = 6.77$, $F_{0.01}(2,179) = 4.72$.
3. Załóżmy, że s_i i n_i są losowe. Jaki warunek powinno spełniać $\ln(A_i)$, aby estymator MNK był zgodny?
4. Oszacować wielkość α. Udowodnić, że oszacowanie to jest zgodne, jeśli tylko spełnione są warunki zgodności estymatora MNK w modelu (*)

Rozwiązanie:

1. Zgodnie z teorią parametr α musi znajdować się w przedziale $[0,1]$. Wynika z tego, że współczynnik $\beta_1 = \frac{\alpha}{1-\alpha}$ przy $\ln(s_i)$ powinien być dodatni a parametr $\beta_2 = -\left(\frac{\alpha}{1-\alpha}\right)$ przy $\ln(n_i + 0.05)$powinien być ujemny. Uzyskane w regresji znaki parametrów są zgodne z teorią
2. Z postaci modelu wynika, że $\beta_1 = -\beta_2$. Zweryfikować te ograniczenie można licząc statystykę F na postawie sum kwadratów reszt w modelu z ograniczeniami i bez ograniczeń. $S = 163.3$, $S_R = 174.7$. Statystyka
   $$F(1,179) = \frac{(174.7 - 163.3)/1}{163.3/(182-3)} = 12.496 > 6.77$$
   Hipotezę zerową $H_0 : \beta_1 = -\beta_2$ odrzucamy
3. Aby estymator MNK był zgodny w przypadku losowych s_i i n_i musi zachodzić, że
   $$Cov(s_i, \varepsilon_i) = Cov(s_i, \ln(A_i) + \eta_i) = Cov(s_i, \eta_i) + Cov(s_i, \ln(A_i)) = 0$$
   i
   $$Cov(n_i, \varepsilon_i) = Cov(n_i, \ln(A_i) + \eta_i) = Cov(n_i, \eta_i) + Cov(n_i, \ln(A_i)) = 0$$
   Wynika z tego, że koniecznym warunkiem zgodności estymatora MNK jest w tym modelu $Cov(s_i, \ln(A_i)) = 0$ i $Cov(n_i, \ln(A_i)) = 0$ - brak korelacji między oszczędnościami i stopą wzrostu siły roboczej a stanem technologii w danym kraju (teoretycznie mogłoby się zdarzyć, że $Cov(n_i, \eta_i) = -Cov(n_i, \ln(A_i))$ i kowariancja $Cov(s_i, \varepsilon_i) = 0$ mimo, że $Cov(s_i, \ln(A_i)) \neq 0$ jednak w praktyce wystąpienie tego przypadeku jest mało nieprawdopodobne).
4. $\beta_1 = \frac{\alpha}{1-\alpha}$ a więc $\alpha = \frac{\beta_1}{1+\beta_1}$. Estymatorem β może więc być $\widehat{\alpha} = \frac{\widehat{\beta}_1}{1+\widehat{\beta}_1}$, drugi estymator można znaleźć z zależności $\beta_2 = -\frac{\alpha}{1-\alpha}$ a więc $\widehat{\alpha} = \frac{\widehat{\beta}_2}{\widehat{\beta}_2 - 1}$. Z pierwszego wzoru otrzymujemy $\widehat{\alpha} = \frac{0.716}{1+0.716} = 0.41725$ z drugiego $\widehat{\alpha} = \frac{-2.431}{-2.431-1} = 0.70854$. Przy założeniu, że estymator MNK w tym modelu jest zgodny oba estymatory będą zgodne, ponieważ jeśli $\operatorname{plim}\left(\widehat{\beta}_1\right) = \beta_1$, $\operatorname{plim}\left(\widehat{\beta}_2\right) = \beta_2$, to
   $$\operatorname{plim}\left(\frac{\widehat{\beta}_1}{1+\widehat{\beta}_1}\right) = \frac{\operatorname{plim}\left(\widehat{\beta}_1\right)}{1+\operatorname{plim}\left(\widehat{\beta}_1\right)} = \frac{\beta_1}{1+\beta_1} = \alpha$$
   $$\operatorname{plim}\left(\frac{\widehat{\beta}_2}{\widehat{\beta}_2 - 1}\right) = \frac{\operatorname{plim}\left(\widehat{\beta}_2\right)}{\operatorname{plim}\left(\widehat{\beta}_2\right) - 1} = \frac{\beta_2}{\beta_2 - 1} = \alpha$$