Egzamin z ekonometrii dla IiE 31.01.2007
I semestr

Pytania teoretyczne

1. Dlaczego w modelu nie powinno się umieszczać jednocześnie stałej i wszystkich zmiennych zerojedynkowych, związanych z poziomami zmiennej dyskretnej?
2. Udowodnić, że rozkład sumy kwadratów reszt jest rozkładem χ^2_{N-K}, niezależnym od rozkładu $\boldsymbol{b}$.
3. W jaki sposób przeprowadza się test prognoz?
4. Pokazać, że dla znanej próby $\boldsymbol{X}$ estymator MNK jest nieobciążony, nawet jeśli ta próba jest losowa. Podać założenia konieczne do tego dowodu.

ZADANIE 1 Za pomocą MNK uzyskano na podstawie dwóch niezależnych próbek oszacowania $\boldsymbol{b}_1$ i $\boldsymbol{b}_2$ parametru $\boldsymbol{\beta}$ w modelu. Zakładamy, że w obu próbkach spełnione są założenia $KMRL$.

$$y_i = \boldsymbol{x}_i\boldsymbol{\beta} + \varepsilon_i$$

Uwaga: przez niezależność próbek rozumie się, że błędy losowe w obu próbach są od siebie niezależne.

1. Pokazać, że estymatory $\boldsymbol{b}_1$ i $\boldsymbol{b}_2$ są od siebie niezależne.
2. Załóżmy, że parametry $\boldsymbol{\beta}$ i σ^2 są takie same w obu próbkach, jaka będzie wartość oczekiwana i wariancja $\boldsymbol{b}_1 - \boldsymbol{b}_2$?
3. Przyjmując założenie z punktu (2) i dodatkowo przyjmując, że $\boldsymbol{\varepsilon}$ ma rozkład normalny, jaki rozkład $\boldsymbol{b}_1 - \boldsymbol{b}_2$?
4. Zaproponować statystykę, którą możnaby zastosować do przetestowania hipotezy o równości β w obu podpróbkach, przy zakłożeniu, że σ^2 jest znane.

   Podpowiedź: dla wektora losowego $\boldsymbol{u}_{[g\times 1]} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma})$, $\boldsymbol{u}'\boldsymbol{\Sigma}^{-1}\boldsymbol{u} \sim \chi^2_g$

Rozwiązanie:

1. Estymator $\boldsymbol{b}_1 = \boldsymbol{\beta}_1 + \left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'\varepsilon_1$ a $\boldsymbol{b}_2 = \boldsymbol{\beta}_2 + \left(\boldsymbol{X}_2'\boldsymbol{X}_2\right)^{-1}\boldsymbol{X}_2'\varepsilon_2$, z $KMRL$ macierze $\boldsymbol{X}_1$ i $\boldsymbol{X}_2$ są deterministyczne a więc $\boldsymbol{b}_1$ i $\boldsymbol{b}_2$ są deterministycznymi funkcjami niezależnych zmiennych losowych $\boldsymbol{\varepsilon}_1$ i $\boldsymbol{\varepsilon}_2$ i w związku z tym też muszą być niezależne.
2. Na mocy nieobciążoności estymatora MNK przy spełnionych założeniach $KMRL$ wartość oczekiwana jest równa
$$\mathrm{E}(\boldsymbol{b}_1 - \boldsymbol{b}_2) = \mathrm{E}(\boldsymbol{b}_1) - \mathrm{E}(\boldsymbol{b}_2) = \boldsymbol{\beta} - \boldsymbol{\beta} = 0$$
   Wariancja jest równa
$$\begin{aligned}\mathrm{Var}(\boldsymbol{b}_1 - \boldsymbol{b}_2) &= \mathrm{Var}(\boldsymbol{b}_1) + \mathrm{Var}(\boldsymbol{b}_2) - 2\,\mathrm{Cov}(\boldsymbol{b}_1, \boldsymbol{b}_2)\\ &= \mathrm{Var}(\boldsymbol{b}_1) + \mathrm{Var}(\boldsymbol{b}_2) = \sigma^2\left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1} + \sigma^2\left(\boldsymbol{X}_2'\boldsymbol{X}_2\right)^{-1}\end{aligned}$$
   Przy czym skorzystaliśmy z tego, że kowariancja między zmiennymi niezależnymi jest równa zeru.
3. Jeśli $\boldsymbol{\varepsilon}$ ma rozkład normalny to $\boldsymbol{b}_1$ i $\boldsymbol{b}_2$ jako liniowe i deterministyczne funkcje $\boldsymbol{\varepsilon}$ mają także rozkład normalny. Różnica niezależnych zmiennych o rozkładzie normalnym ma także rozkład normalny:
$$\boldsymbol{b}_1 - \boldsymbol{b}_2 \sim N\left(\boldsymbol{0}, \sigma^2\left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1} + \sigma^2\left(\boldsymbol{X}_2'\boldsymbol{X}_2\right)^{-1}\right)$$
4. Podstawiając $\boldsymbol{u} = \boldsymbol{b}_1 - \boldsymbol{b}_2$ uzyskujemy statystykę
$$(\boldsymbol{b}_1 - \boldsymbol{b}_2)'\left[\sigma^2\left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1} + \sigma^2\left(\boldsymbol{X}_2'\boldsymbol{X}_2\right)^{-1}\right]^{-1}(\boldsymbol{b}_1 - \boldsymbol{b}_2) \sim \chi^2_g$$
   gdzie g jest liczbą zmiennych w modelu.

ZADANIE 2 Szacowana jest funkcja produkcji typu Cobba-Douglasa:

$$Q_t = A_t L_t^\alpha K_t^\beta$$

gdzie Q_t oznacza produkcję, L_t - wielkość siły roboczej a K_t - kapitał zaangażowany w produkcję, A_t - poziom technologii. Dane dotyczące kapitału nie są dostępne ale znamy dane dotyczące stopy procentowej r. Krańcowa produktywność kapitału jest równa $\frac{\partial Q_t}{\partial K_t} = A_t \beta L_t^\alpha K_t^{\beta-1}$.

1. Przy założeniu, że wynagrodzenie kapitału (stopa procentowa) jest równa jego produktywności krańcowej przekształcić model tak, żeby stał się on modelem liniowym względem logarytmów zmiennych Q, A, L i r.

2. Korzystając z prawa malejącego krańcowowego produktu oraz z tego, że produkt krańcowy powinien być dodatni, pokazać, że współczynnik przy logarytmie L powinien być dodatni a przy logarytmie r ujemny.

3. Jakie warunki powinny spełniać parametry przekształconego modelu, aby funkcja produkcji charakteryzowała się stałymi przychodami skali?

4. Dla jakiej postaci funkcji A_t stopa wzrostu technologicznego jest w tym modelu stała w czasie?

5. Przy użyciu danych dotyczących wielkości produktu narodowego w wyrażeniu nominalnym (PKB), siły roboczej, stopy referencyjnej banku centralnego oraz deflatora CPI dla 36 obserwacji z kwartałów 1995.1-2004.3 otrzymano następujące wyniki estymacji w modelu na logarytmach.

Zmienna	**Współczynnik**	**(Bł. stand.)**
l	1.288144	0.428196
r	0.005671	0.019787
t	0.011029	0.001778
stała	-5.817455	4.156405

Na podstawie tych wyników przetestować hipotezy, że parametry przy logarytmach zmiennych mają oczekiwane znaki oraz, że stopa wzrostu technologicznego jest dodatnia oraz, że przychody skali są stałe. Wartość krytyczne dla jednostronnych przedziałów krytycznych wynoszą $t_{0.05}(31) = 1.696$, $t_{0.05}(32) = 1.694$, $t_{0.05}(33) = 1.692$, natomiast dla dwustronnego obszaru krytycznego $t_{0.05}(31) = 2.035$, $t_{0.05}(32) = 2.037$, $t_{0.05}(33) = 2.035$. Odpowiedzi uzasadnić.

6. Dodając do modelu zmienną t^2 starano się sprawdzić, czy postęp technologiczny przyspiesza, czy też zwalnia w czasie. Uzyskano następujące wyniki regresji (małymi literami oznaczone są logarytmy zmiennych):

Zmienna	**Współczynnik**	**Bł. stand.**	**t**	**P>|t|**
l	1.248485	0.443130	2.82	0.008
r	-0.007822	0.036839	-0.21	0.833
t	0.014521	0.008199	1.77	0.086
t2	-0.000085	0.000195	-0.44	0.665
stała	-5.502959	4.271199	-1.29	0.207

oraz wielkości statystyk VIF

Zmienna	**VIF**	**1/VIF**
t	91.40	0.010941
t2	83.81	0.011932
r	4.42	0.226141
l	3.90	0.2562577
Średni VIF	45.88	

Wyjaśnić, jakie wnioski płyną z tej regresji w odniesieniu do pytania o zmiany szybkości postępu technologicznego. Odpowiedź uzasadnić?

Rozwiązanie:

1. Mnożąc wzór $r = \frac{\partial Q}{\partial K} = A\beta L^\alpha K^{\beta-1}$ przez K uzyskujemy: $rK = A\beta L^\alpha K^\beta$, ponieważ $Q = AL^\alpha K^\beta$ więc $rK = \beta Q$, w rezultacie $K = \beta \frac{Q}{r}$. Podstawiając do wzoru na produkcję uzyskujemy

$$Q = AL^\alpha \left(\beta \frac{Q}{r}\right)^\beta = AL^\alpha \left(\beta \frac{Q}{r}\right)^\beta = AL^\alpha \beta^\beta Q^\beta r^{-\beta}$$

rozwiązując dla Q uzyskujemy:

$$Q = \left(\beta^{\beta} A\right)^{\frac{1}{1-\beta}} L^{\frac{\alpha}{1-\beta}} r^{-\frac{\beta}{1-\beta}}$$

Logarytmując stronami uzyskujemy

$$\ln Q = \frac{1}{1-\beta} \ln\left(\beta^{\beta} A\right) + \frac{\alpha}{1-\beta} \ln L - \frac{\beta}{1-\beta} \ln r$$

2. Krańcowy przychód z kapitału i pracy jest dodatni tylko wtedy, gdy $\frac{\partial Q}{\partial K} = A\beta L^{\alpha} K^{\beta-1} > 0$ z czego wynika, że $\beta > 0$. Podobnie z $\frac{\partial Q}{\partial L} = A\alpha L^{\alpha-1} K^{\beta} > 0$ wynika, że $\alpha > 0$. Z kolei przychody krańcowe maleją, gdy drugie pochodne są ujemne $\frac{\partial^2 Q}{\partial^2 K} = A\beta\left(\beta - 1\right) L^{\alpha} K^{\beta-2} < 0$, co implikuje, że $\beta < 1$. Dla $\beta \in (0, 1)$ i $\alpha > 0$ parametr przy $\ln L$ równy $\frac{\alpha}{1-\beta} > 0$ a parametr przy $\ln r$ równy $-\frac{\beta}{1-\beta}$ jest ujemny.

3. Przychody skali są stałe, gdy $Q\left(\lambda L, \lambda K\right) = \lambda Q\left(L, K\right) = A\left(\lambda L\right)^{\alpha} \left(\lambda K\right)^{\beta} = AL^{\alpha} K^{\beta} \lambda^{\alpha+\beta}$. Dla dowolnej λ zachodzi to jedynie dla $\alpha + \beta = 1$. Ponieważ parametr przy $\ln L$ równy jest $\frac{\alpha}{1-\beta}$ więc dla stałych przychodów skali powinien być równy $\frac{\alpha}{1-\beta} = \frac{\alpha}{\alpha} = 1$

4. Stopa wzrostu wydajności jest stała, gdy $\frac{Q_{t+1}}{Q_t} = const$ dla stałego K i L. Zachodzi to dla $A_t = e^{\gamma_0+\gamma_1 t}$, ponieważ wtedy $\frac{Q_{t+1}}{Q_t} = \frac{e^{\gamma_0+\gamma_1(t+1)} L^{\alpha} K^{\beta}}{e^{\gamma_0+\gamma_1 t} L^{\alpha} K^{\beta}} = e^{\gamma}$. Wynika z tego, że można przyjąć, że $\ln\left(A_t\right) = \gamma_0 + \gamma_1 t$

5. Wielkości statystyk t jest następująca: dla l 3.01, dla r 0.29, dla t 6.20, dla stałej -1.40. Ponieważ 3.01$>t_{0.05}\left(32\right) = 1.694$ więc możemy dla parametru przy $\ln L$ odrzucić $H_0 : \frac{\alpha}{1-\beta} = 0$ na rzecz hipotezy alternatywnej, że $\frac{\alpha}{1-\beta} > 0$. W przypadku parametru przy $\ln r$, ponieważ 0.29$>t_{0.05}\left(32\right) = -1.694$ nie możemy odrzucić $H_0 : -\frac{\beta}{1-\beta} = 0$ na rzecz $H_1 : -\frac{\beta}{1-\beta} < 0$. Estymacja nie potwierdziła ujemnego znaku przy stopie procentowej. Wszystkie hipotezy o znaku testujemy stosując jednostronne obszary krytyczne!

   Jeśli przychody skali są stałe to parametr przy $\ln\left(L\right)$ powinien być równy jeden. Testujemy tą hipotezę tworząc statystykę t postaci: $\frac{1.288144-1}{0.428196} = 0.672\,93 < t_{0.05}\left(32\right) = 2.037$. Hipotezy o stałych przychodach skali nie można odrzucić.

6. Po wprowadzeniu t^2 do modelu współczynnik przy t^2 okazał się nieistotny (0.086>0.05 i 0.0665>0.05) w związku z tym nie da się stwierdzić czy postęp techniczny zwalnia, czy przyspiesza, ponieważ przy testowaniu hipotezy $H_0 : \beta_{t^2} = 0$ przy $H_1 : \beta_{t^2} > 0$ i hipotezy $H_0 : \beta_{t^2} = 0$ przy $H_1 : \beta_{t^2} < 0$ w obu przypadkach otrzymamy wynik, że $\beta_{t^2} = 0$. Uzyskany wynik nie świadczy jednak o też o tym, że postęp techniczny jest stały w czasie. W drugiej regresji współczynnik przy t stał się nieistotny mimo, że był istotny w pierwszej regresji a statystyka VIF okazała się bardzo wysoka zarówno dla t jak i t^2. Wnioskujemy z tego, że drugiej w regresji występuje współliniowość między t i t^2. Z tego powodu współczynniki w drugiej regresji są nieprecyzyjnie oszacowane i dlatego też nie da się na jej podstawie wyciągnąć wyraźniejszych wniosków na temat cech postępu technicznego.

ZADANIE 3 Estymowany jest model, w którym zmiennymi objaśniającymi są jedynie zmienne zerojedynkowe, powstałe w wyniku rozkodowania zmiennej dyskretnej x_i. Formalnie, model ten można zapisać w sposób następujący:

$$y_i = \sum_{s=1}^{K} \gamma_s D_{s,i} + \varepsilon_i$$

gdzie $D_{s,i} = \begin{cases} 1 & \text{jeśli} \quad x_i = s \\ 0 & \text{jeśli} \quad x_i \neq s \end{cases}$. Zmienna x_i może być interpretowana jako zmienna, która dzieli próbę na K rozdzielnych podpróbek o określonych cechach.

1. Pokazać, że w tego rodzaju modelu estymatorem MNK parametru γ_s będzie średnia y_i policzona dla podpróbki s, a więc takiej, dla której $x_i = s$.

2. Przez y_{sj} oznaczono obserwację j dla podpróbki s liczebnosć tej podpróbki wynosi n_s. Pokazać, że całkowitą sumę kwadratów w takim modelu można zdekomponować w sposób następujący:

$$TSS = \sum_{s=1}^{K} \sum_{j=1}^{n_s} \left(y_{sj} - \overline{y}_{\bullet\bullet}\right)^2 = \sum_{s=1}^{K} n_s \left(\overline{y}_{s\bullet} - \overline{y}_{\bullet\bullet}\right)^2 + \sum_{s=1}^{K} \sum_{j=1}^{n_s} \left(\overline{y}_{s\bullet} - \overline{y}_{s\bullet}\right)^2$$

gdzie $\overline{y}_{\bullet\bullet} = n^{-1}\sum_{s=1}^{K}\sum_{i=1}^{n_s} y_{sj}$ jest średnią w próbie, natomiast $\overline{y}_{s\bullet} = n_s^{-1}\sum_{j=1}^{n_s} y_{sj}$ średnią dla danej podpróbki.

3. Wyjaśnić ile wynosi RSS i TSS w tym modelu. Podać wzór na test, za pomocą którego możnaby przetestować hipotezę:

$$H_0 : \gamma_1 = \gamma_2 = \ldots = \gamma_K$$

Podać rozkład tego testu.

4. Załóżmy, że x_i jest losowe. Jaki warunek musi spełniać ε_i, by estymator MNK parametru γ_i był nieobciążny? Jaki warunek musi spełniać ε_i by estymator MNK parametru γ_i był zgodny.

Rozwiązanie:

1. Zapiszmy macierz obserwacji $\boldsymbol{X} = \begin{bmatrix} \boldsymbol{D}_1 & \boldsymbol{D}_2 & \boldsymbol{D}_3 & \boldsymbol{D}_4 \end{bmatrix}$, gdzie $\boldsymbol{D}_s$ są wektorami obserwacji dla poszczególnych zmiennych zerojedynkowych (złożonymi tylko z zer i jedynek). Z własności hiperpłaszczyny regresji

$$\boldsymbol{X}'\boldsymbol{e} = \boldsymbol{0}$$

a więc

$$\begin{bmatrix} \boldsymbol{D}_1'\boldsymbol{e} \\ \boldsymbol{D}_2'\boldsymbol{e} \\ \vdots \\ \boldsymbol{D}_K'\boldsymbol{e} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} D_{1,i}\left(y_{1i} - b_1\right) \\ \sum_{i=1}^{n} D_{2,i}\left(y_{2i} - b_2\right) \\ \vdots \\ \sum_{i=1}^{n} D_{3,i}\left(y_{3i} - b_3\right) \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{n_1}\left(y_{1j} - b_1\right) \\ \sum_{j=1}^{n_2}\left(y_{2j} - b_2\right) \\ \vdots \\ \sum_{j=1}^{n_4}\left(y_{4j} - b_1\right) \end{bmatrix} = \boldsymbol{0}$$

Rozwiązując ostatni układ równań otrzymujemy

$$b_s = \frac{\sum_{j=1}^{n_s} \overline{y}_{sj}}{n_s} = \overline{y}_{s\bullet}$$

2. Unieważnione (w egzaminie był błąd w definicji)

Rozpiszmy wzór na TSS

$$\begin{aligned}
\sum_{j=1}^{K}\sum_{i=1}^{n_j}\left(y_{ij} - \overline{y}_{\bullet\bullet}\right)^2 &= \sum_{s=1}^{K}\sum_{j=1}^{n_s}\left[\left(\overline{y}_{s\bullet} - \overline{y}_{\bullet\bullet}\right) + \left(y_{ij} - \overline{y}_{s\bullet}\right)\right]^2 \\
&= \sum_{s=1}^{K} n_s\left(\overline{y}_{s\bullet} - \overline{y}_{\bullet\bullet}\right)^2 + \sum_{s=1}^{K}\sum_{j=1}^{n_s}\left(y_{sj} - \overline{y}_{s\bullet}\right)^2 \\
&+ 2\sum_{s=1}^{K}\sum_{j=1}^{n_s}\left(\overline{y}_{s\bullet} - \overline{y}_{\bullet\bullet}\right)\left(y_{sj} - \overline{y}_{s\bullet}\right)
\end{aligned}$$

ale

$$\sum_{s=1}^{K}\sum_{j=1}^{n_s}\left(\overline{y}_{s\bullet} - \overline{y}_{\bullet\bullet}\right)\left(y_{sj} - \overline{y}_{s\bullet}\right) = \sum_{j=0}^{K}\left[\left(\overline{y}_{s\bullet} - \overline{y}_{\bullet\bullet}\right)\sum_{j=1}^{n_s}\left(y_{sj} - \overline{y}_{s\bullet}\right)\right] = 0$$

ponieważ $\sum_{j=1}^{n_s}\left(y_{sj} - \overline{y}_{s\bullet}\right) = n_s\overline{y}_{s\bullet} - n_j\overline{y}_{s\bullet} = 0$.

3. Z definicji

$$TSS = RSS + ESS$$

ale w tym modelu $\widehat{y}_i = \overline{y}_{s\bullet}$ jeśli obserwacja i ma charakterystykę s. W rezultacie

$$TSS = \sum_{j=1}^{K}\sum_{i=1}^{n_j}\left(y_{ij} - \overline{y}_{\bullet\bullet}\right)^2$$

$$ESS = \sum_{s=1}^{K} n_s\left(\overline{y}_{s\bullet} - \overline{y}_{\bullet\bullet}\right)^2$$

$$RSS = \sum_{s=1}^{K}\sum_{j=1}^{n_s}\left(y_{sj} - \overline{y}_{s\bullet}\right)^2$$

Z kolei wzór na statystykę F ma postać

$$\begin{aligned} F\left(K-1, N-K\right) &= \frac{\left[\sum_{j=1}^{K} \sum_{i=1}^{n_j} \left(y_{ij} - \overline{y}_{\bullet\bullet}\right)^2 - \sum_{s=1}^{K} \sum_{j=1}^{n_s} \left(y_{sj} - \overline{y}_{s\bullet}\right)^2\right] \Big/ \left(K-1\right)}{\sum_{s=1}^{K} \sum_{j=1}^{n_s} \left(y_{sj} - \overline{y}_{s\bullet}\right)^2 \Big/ \left(n-K\right)} \\ &= \frac{\sum_{s=1}^{K} n_s \left(\overline{y}_{s\bullet} - \overline{y}_{\bullet\bullet}\right)^2 \Big/ \left(K-1\right)}{\sum_{s=1}^{K} \sum_{j=1}^{n_s} \left(y_{sj} - \overline{y}_{s\bullet}\right)^2 \Big/ \left(n-K\right)} \end{aligned}$$

4. Estymator $\boldsymbol{b}$ jest nieobciążony jeśli $\mathrm{E}\left(\varepsilon_i | \, x_i\right) = 0$. W anlizowanym przypadku oznacza to, że $\mathrm{E}\left(\varepsilon_i | \, x_i = s\right) = 0$ dla każdego $s = 1, \ldots, K$. Warunek konieczny zgodności sprowadza się do warunku, że korelacja empiryczna $\operatorname{plim} \frac{1}{N} \boldsymbol{X}' \boldsymbol{\varepsilon} = \operatorname{plim} \frac{1}{N} \sum_{i=1}^{N} x_i \varepsilon_i = 0$.