Egzamin z ekonometrii dla IiE 14.09.2007
I semestr

ZADANIE 1 W modelu ze stałą:

$$y_i = \beta_0 + \boldsymbol{x}_i\boldsymbol{\beta} + \varepsilon_i$$

1. pokazać, że spełnione jest, że
$$\overline{y} = b_0 + \overline{\boldsymbol{x}}\boldsymbol{b}$$
gdzie b_0 i $\boldsymbol{b}$ są oszacowaniami MNK parametrów β_0 i $\boldsymbol{\beta}$.

2. pokazać, że wektor reszt e_i z regresji y_i na $\boldsymbol{x}_i$ spełnia:
$$\boldsymbol{e} = \boldsymbol{y^*} - \boldsymbol{X}^*\boldsymbol{b}$$
gdzie $y_i^* = y_i - \overline{y}$, $\boldsymbol{x_i^*} = \boldsymbol{x}_i - \overline{\boldsymbol{x}}$

3. pokazać, że układ równań normalnych zapisany jako $\boldsymbol{X^*}'\boldsymbol{e}^* = 0$ jest spełniony dla $\boldsymbol{e}^* = \boldsymbol{y^*} - \boldsymbol{X^*b}$

   Podpowiedź: zwróć uwagę na to, że $\boldsymbol{X^*} = \boldsymbol{X} - \frac{1}{n}\boldsymbol{ll'X}$

4. Jaka jest relacja między estymatorami uzyskanymi z regresji na poziomach i regresji na odchyleniach od średnich?

   Podpowiedź: zastosuj wyniki z dwóch poprzednich punktów.

Rozwiązanie:

1. Jeśli b_0 i $\boldsymbol{b}$ są estymatorami MNK, to zachodzi $\sum e_i = 0$. Zatem:
$$\begin{aligned} y_i &= b_0 + \boldsymbol{x}_i\boldsymbol{b} + e_i, \\ \sum y_i &= \sum b_0 + \sum \boldsymbol{x}_i\boldsymbol{b} + \sum_i e, \\ \frac{\sum y_i}{n} &= \frac{nb_0}{n} + \frac{\sum \boldsymbol{x}_i\boldsymbol{b}}{n} + \frac{0}{n}, \\ \overline{y} &= b_0 + \overline{\boldsymbol{x}}\boldsymbol{b}. \end{aligned}$$

2. Jeśli $\boldsymbol{b}$ jest estymatorem MNK, to:
$$\begin{aligned} y_i &= b_0 + \boldsymbol{x}_i\boldsymbol{b} + e_i, \\ y_i - \overline{y} &= b_0 + \boldsymbol{x}_i\boldsymbol{b} + e_i - \overline{y}, \\ y_i^* &= \boldsymbol{x}_i\boldsymbol{b} + e_i - \overline{\boldsymbol{x}}'\boldsymbol{b}, \\ y_i^* &= (\boldsymbol{x}_i - \overline{\boldsymbol{x}})\,\boldsymbol{b} + e_i, \\ y_i^* &= \boldsymbol{x}_i^*\boldsymbol{b} + e_i \Longrightarrow \\ e_i &= y_i^* - \boldsymbol{x}_i^*\boldsymbol{b} \end{aligned}$$

3. Z poprzedniego punktu wynika, że $\boldsymbol{X}^{*\prime}\boldsymbol{e}^* = \boldsymbol{X}^{*\prime}\boldsymbol{e}$. Dalej
$$\begin{aligned} \boldsymbol{X}^{*\prime}\boldsymbol{e}^* = \boldsymbol{X}^{*\prime}\boldsymbol{e} &= \left(\boldsymbol{X}' - \frac{1}{n}\boldsymbol{Xll}'\right)\boldsymbol{e} \\ &= \underbrace{\boldsymbol{X}'\boldsymbol{e}}_{0} - \frac{1}{n}\boldsymbol{Xl}\,\underbrace{\boldsymbol{l}'\boldsymbol{e}}_{0} = 0 \end{aligned}$$

4. Z dwóch poprzednich punktów wnioskujemy, że układ równań normalnych dla regresji $\boldsymbol{x_i^*}$ na y_i^* jest spełniony dla $\boldsymbol{b}$
$$\boldsymbol{X}^{*\prime}\left(\boldsymbol{y^*} - \boldsymbol{X^*b}\right) = 0$$
rozwiązując te równanie $\boldsymbol{b}$ uzyskujemy $\boldsymbol{b} = \left(\boldsymbol{X}^{*\prime}\boldsymbol{X}^*\right)\boldsymbol{X}^{*\prime}\boldsymbol{y^*} = \boldsymbol{b}^*$ a więc estymatorem MNK w regresji y_i^* na $\boldsymbol{x_i^*}$ jest $\boldsymbol{b}$ - oszacowaniu współczynniki przy $\boldsymbol{x}_i$ w regresji y_i na stałej i $\boldsymbol{x}_i$.

ZADANIE 2 Szacujemy model w którym zmienną objaśnianą jest wynagrodzenie (zmienna $\ln wyn_i$), a zmienną objaśniającą jest wykształcenie pracownika (zmienna edu_i) zakodowane w następujący sposób: 1 - podstawowe, 2 - gimnazjum, 3 - średnie, 4 - wyższe.

1. Wyjaśnić dlaczego nie powinno się estymować modelu postaci $\ln wyn_i = \alpha + \beta edu_i + \varepsilon_i$.
2. Podać prawidłową formę modelu i zinterpretować jego współczynniki.
3. Przy założeniu, że osoby z wykształceniem podstawowym poświęciły na naukę 6 lat, z gimazjalnym 9 lat, ze średnim 12 lat, a z wyższym 17 lat, zaproponować sposób przetestowania hipotezy, że istotna jest jedynie liczba lat spędzonych na nauce (nowo zdefiniowana na podstawie zmiennej edu_i zmienna $lata_i$), a nie poziom osiągniętego wykształcenia.
4. Jaki jest wynik tego testu jeśli wyestymowana przy ograniczeniu wynikającym z H_0 suma kwadratów reszt wyniosła $RSS = 28$, a w modelu bez ograniczeń $RSS = 20$, liczba obserwacji $N = 24$. Wartość krytyczna $F_{0.05}(1, 20) = 4.35$, $F_{0.05}(2, 20) = 3.49$.

Rozwiązanie:

1. Dyskretną zmienną niezależną powinno się rozkodować na zmienne zerojedynkowe. Estymując model

$$\ln wyn_i = \alpha + \beta edu_i + \varepsilon_i$$

zakładamy, że oczekiwane wynagrodzenie

$$\mathrm{E}(\ln wyn_1) - \mathrm{E}(\ln wyn_2) = \mathrm{E}(\ln wyn_3) - \mathrm{E}(\ln wyn_4) = \mathrm{E}(\ln wyn_4) - \mathrm{E}(\ln wyn_5) = \beta$$

przy czym jako wyn_s oznaczyliśmy tu wynagrodzenie respondenta, który osiągną poziom wykształcenia s. Procentowy oczekiwany przyrost płacy dla dodatkowego osiągniętego poziomu edukacji jest więc taki sam niezależnie od tego jaki poziom edukacji respondent już uzyskał.

2. Prawidłowa forma modelu jest następująca

$$\ln wyn_i = \alpha + \gamma_1 d_{1i} + \gamma_2 d_{2i} + \gamma_3 d_{3i} + \varepsilon_i$$

gdzie zmienne zero jedynkowe posiadają następujące definicje (najwyższy posiadany poziom wykształcenia): d_{1i} - gimnazjalne, d_{2i} - średnie, d_{3i} - wyższe. Interpretacja współczynników: α logarytm średniej płacy osób z wykształceniem podstawowym, γ_1 procentowa różnica między płacą osoby z wykształceniem podstawowym i gimnazjalnym, γ_2 procentowa różnica między płacą osoby z wykształceniem podstawowym i średnim, γ_3 procentowa różnica między płacą osoby z wykształceniem podstawowym i wyższym. Poziom d_{0i} - wykształcenie podstawowe traktujemy tutaj jako bazowe.

3. Musimy znaleźć takie wielkości γ_k, dla których model

$$\ln wyn_i = \alpha + \gamma_1 d_{1i} + \gamma_2 d_{2i} + \gamma_3 d_{3i} + \varepsilon_i$$

sprowadzi się do modelu

$$\begin{aligned}
\ln wyn_i &= \alpha + \beta lata_i + \varepsilon_i \\
&= \alpha + \beta 6 d_{0i} + \beta 9 d_{1i} + \beta 12 d_{2i} + \beta 17 d_{3i} + \varepsilon_i \\
&= \alpha + \beta\left(6d_{0i} + 9d_{1i} + 12d_{2i} + 17d_{3i}\right) + \varepsilon_i \\
&= \alpha + 6\beta + \beta\left(3d_{1i} + 6d_{2i} + 11d_{3i}\right) + \varepsilon_i \\
&= \alpha^* + 3\beta d_{1i} + 6\beta d_{2i} + 11\beta d_{3i} + \varepsilon_i \\
&= \alpha^* + \beta^* d_{1i} + 2\beta^* d_{2i} + \frac{11}{3}\beta^* d_{3i} + \varepsilon_i
\end{aligned}$$

widać z tego, że H_0 będzie prawdziwe jeśli spełnione będzie

$$H_0 : \begin{cases} \gamma_2 = 2\gamma_1 \\ \gamma_3 = \frac{11}{3}\gamma_1 \end{cases}$$

4. Ilość ograniczeń w modelu wynosi $g = 2$, statystyka $F = \frac{(S_R - S)/g}{S/(N-K)} = \frac{(28-20)/2}{20/(24-4)} = \frac{4}{1} = 4 > F_{0.05}(2, 20) = 3.49$ - hipotezę zerową odrzucamy.

ZADANIE 3 W trakcie badania ekonometrycznego szacowano funkcję produkcji Cobba-Douglasa:

$$Q = AK^{\alpha}L^{\beta}.$$

Po transformacji i wyestymowaniu otrzymano oszacowany model postaci:

$$\ln Q_i = \underset{(0.1)}{-0.2} + \underset{(0.06)}{0.3} \ln K_i + \underset{(0.2)}{0.7} \ln L_i.$$

1. Podać interpretację współczynnika $\ln K_i$.
2. Jak należałoby weryfikować hipotezę, że wyestymowana funkcja produkcji charakteryzuje się stałymi przychodami skali?
3. W jaki sposób (zakładając, że wynagrodzenia są równe produktom krańcowym) należałoby weryfikować hipotezę mówiącą o tym, że całkowite wynagrodzenie kapitału wynosi w gospodarce $1/4$ całkowitego wynagrodzenia pracy?
4. W modelu postanowiono uwzględnić wpływ ewentualnej większej efektywności firm prywatnych w procesie produkcji. Jeśli wprowadzić zmienną zerojedynkową:

$$D_{pi} = \begin{cases} 1 & \text{własność prywatna} \\ 0 & \text{własność państwowa} \end{cases}, \quad D_{si} = \begin{cases} 0 & \text{własność prywatna} \\ 1 & \text{własność państwowa} \end{cases},$$

   to które z poniższych równań mogą być użyte do tak rozszerzonego modelu? Odpowiedź uzasadnić.

$$\begin{aligned} &\text{A:} && \ln Q_i = b_0 + b_1 \ln K_i + b_2 \ln L_i + b_3 D_{pi} + \varepsilon_i \\ &\text{B:} && \ln Q_i = b_0 + b_1 \ln K_i + b_2 \ln L_i + b_3 D_{si} + \varepsilon_i \\ &\text{C:} && \ln Q_i = b_0 + b_1 \ln K_i + b_2 \ln L + b_3 D_{pi} + b_4 D_{si} + \varepsilon_i. \end{aligned}$$

5. Do szacowanej funkcji produkcji (+) dodano trend, który stanowić ma przybliżenie postępu technicznego. W tak rozszerzonym modelu $RSS = 30$, podczas gdy w pierwotnym modelu $RSS = 50$. Liczba użytych obserwacji wynosi 54. Policzyć statystykę F i zweryfikować, czy w modelu tym trend jest istotny?
6. Czy ujemna wartość stała jest sensowna w kontekście untepretacji ekonomicznej szacowanego modelu? Odpowiedź uzasadnić.
7. Model (+) oszacowano na podstawie szeregu czasowego, a następnie przeprowadzono test prognoz, przy czym okresem prognozy były 2 okresy następujące po okresie estymacji. Uzyskano statystykę $F_{2,54} = 2$. Co ten wynik sugeruje na temat własności modelu?

   Podpowiedź: $F_{0.05}(2, 54) \approx 3.18$

Rozwiązanie:

1. Współczynnik β jest elastycznością wielkości produkcji względem pracy, zatem wzrost nakładu pracy o 1% spowoduje wzrost produkcji o 0.7%.
2. Hipoteza o stałych przychodach skali powinna być testowana za pomocą testu F badającego następujące ograniczenie na parametry modelu: $\alpha + \beta = 1$.
3. Krańcowy produkt pracy równy wynagrodzeniu pracy jest dla tej funkcji produkcji równy $w = \frac{\partial Q}{\partial L} = \beta \frac{Q}{L}$. Całkowite wynagrodzenie pracy jest równe $wL = \alpha Q$. Podobnie wynagrodzenie całkowite wynagrodzenie kapitału jest równe $rK = \alpha Q$. Stosunek całkowitych wynagrodzeń jest więc równy $\frac{rK}{wL} = \frac{\alpha}{\beta}$. Przetestowanie hipotezy o stosunku wynagrodzeń sprowadza się więc do przetestowania hipotezy $\frac{\alpha}{\beta} = \frac{1}{4}$ lub $\beta = 4\alpha$ za pomocą statystyki F.
4. Tylko model C nie nadaje się do weryfikacji hipotezy mówiącej o wpływie rodzaju własności na proces produkcji, ponieważ tylko w tym modelu występuje pełna współliniowość. Suma kolumn ze zmiennymi zerojedynkowymi D_{pi} i D_{si} jest równa kolumnie jedynek.

5. W celu zweryfikowania hipotezy zerowej mówiącej braku wpływu postępu technologicznego przybliżonego trendem na produkcję, należy policzyć wartość statystyki testującej $F = \frac{(RSS-RSS_R)/j}{RSS_R/(n-k)} = 20 > F_{0.95,1,50} \approx 4.04$. Należy zatem odrzucić hipotezę zerową i uznać, że postęp technologiczny ma wpływ na produkcję

6. Oszacowana ujemna stała nie jest problemem, "nie skutkuje ujemną produkcją" ponieważ zmienną objaśnianą jest logarytm produkcji.

7. Test prognoz służy do sprawdzenia stabilności parametrów modelu w próbie z kilkoma dodatkowymi obserwacjami. Hipoteza zerowa mówi o tym, że parametry modelu są stabilne. Uzyskana wartość statystyki testującej jest mniejsza od wartości krytycznej (5% poziom istotności) wynoszącej $F_{0.05}(2, 54) \approx 3.18$, zatem parametry modelu są stabilne