Egzamin z ekonometrii 5.03.2010

Pytania teoretyczne

1. Pokaż, że w modelu ze stałą suma reszt jest równa zeru.
2. Dlaczego zmienną dyskretną rozkodowujemy na zmienne zerojedynkowe?
3. Udowodnij, że jeśli spełnione są założenia KMRL, to s^2 jest nieobciążonym estymatorem σ^2.
4. W jaki sposób można, w przypadku znanej macierzy $\mathbf{\Omega}$, sprowadzić model z niesferycznymi błędami losowymi do modelu spełniającego założenia KMRL.

ZADANIE 1 Estymujemy model trendu liniowego postaci

$$y_t = \beta_1 + \beta_2 t + \varepsilon_t$$

dla $t = 1, 2, \ldots, T$. Załóżmy, że $T = \sum_{t=1}^T y_t = \sum_{t=1}^T t y_t = 11$.

1. Wyprowadź wzory dla estymatorów b_1, b_2. Policz wielkości tych estymatorów dla podanych wartości $T, \sum_{t=1}^T y_t, \sum_{t=1}^T t y_t$.

   UWAGA: skorzystaj z tego, że $\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$, $\sum_{t=1}^T t = \frac{T(T+1)}{2}$ a $\sum_{t=1}^T t^2 = \frac{T(2T+1)(T+1)}{6}$.
2. Policz macierz wariancji kowariancji tych estymatorów. Załóż, że spełnione są założenia KMRL i $\sigma^2 = 1$.

Rozwiązanie:

1. Macierz obserwacji dla tego modelu ma postać

$$\boldsymbol{X} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & T \end{bmatrix}$$

W związku z tym

$$\boldsymbol{X}'\boldsymbol{X} = \begin{bmatrix} T & \sum_{t=1}^T t \\ \sum_{t=1}^T t & \sum_{t=1}^T t^2 \end{bmatrix} = \begin{bmatrix} T & \frac{T(T+1)}{2} \\ \frac{T(T+1)}{2} & \frac{T(2T+1)(T+1)}{6} \end{bmatrix}$$

$$\begin{aligned} \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} &= \frac{1}{\frac{T^2(2T+1)(T+1)}{6} - \frac{T^2(T+1)^2}{4}} \begin{bmatrix} \frac{T(2T+1)(T+1)}{6} & -\frac{T(T+1)}{2} \\ -\frac{T(T+1)}{2} & T \end{bmatrix} \\ &= \frac{1}{T(T-1)} \begin{bmatrix} 2(2T+1) & -6 \\ -6 & \frac{12}{T+1} \end{bmatrix} \end{aligned}$$

$$\boldsymbol{X}'\,\boldsymbol{y} = \begin{bmatrix} \sum_{t=1}^T y_t \\ \sum_{t=1}^T t y_t \end{bmatrix}$$

$$\boldsymbol{b} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{x}'\,\boldsymbol{y} = \frac{2}{T(T-1)} \begin{bmatrix} (2T+1)\sum_{t=1}^T y_t - 3\sum_{t=1}^T t y_t \\ -3\sum_{t=1}^T y_t + \frac{6}{T+1}\sum_{t=1}^T t y_t \end{bmatrix}$$

$$b_1 = \frac{2\,[23 \times 11 - 3 \times 11]}{11 \times 10} = \frac{2\,[20]}{10} = 4$$

$$b_2 = \frac{2\left[-3 \times 11 + \frac{1}{2} \times 11\right]}{11 \times 10} = \frac{5}{10} = \frac{1}{2}$$

$$\mathrm{Var}\,(b) = \frac{2\sigma^2}{T(T-1)} \begin{bmatrix} 2T+1 & -3 \\ -3 & \frac{6}{T+1} \end{bmatrix}$$

$$\mathrm{Var}\,(b) = \begin{bmatrix} \frac{23}{110} & -\frac{3}{110} \\ -\frac{3}{110} & \frac{1}{220} \end{bmatrix}$$

ZADANIE 2 Podstawowym równaniem ilościowej teorii pieniądza jest równanie *równanie wymiany* sformułowane w 1911 przez Irvinga Fishera:

$$MV = PT, \tag{*}$$

gdzie M jest podażą pieniądza, V szybkością obiegu pieniądza, P poziomem cen a T liczbą transakcji w gospodarce. Według ilościowej teorii pieniądza V jest stałe, a P dostosowuje się do zmian M i T.
W badaniu za miarę podaży pieniądza przyjęto $M3$, za miarę poziomu cen CPI a za miarę liczby transakcji, PKB w wyrażeniu realnym. Model postanowiono oszacować przy wykorzystaniu danych dotyczących stóp wzrostu $pkb_t = \frac{PKB_t}{PKB_{t-1}}$, $m3_t = \frac{M3_t}{M3_{t-1}}$, $p_t = \frac{CPI_t}{CPI_{t-1}}$.
Podpowiedź: hipotezy testujemy na poziomie istotności $\alpha = 0.05$. Wartość krytyczna $t_{0.05}(35) = 2.03$

1. Przyjmując, że ilościowa teoria pieniądza jest prawdziwa, która zmienna powinna być uznana za zmienną endogeniczną? Odpowiedź uzasadnij.

2. Przekształć równanie (*) tak, by zależność dotyczyła stóp wzrostu a nie poziomów zmiennych

3. Pokaż, że uzyskane równanie można przekształcić do postaci równania liniowego względem przekształconych zmiennych

4. Jakie warunki powinny spełniać parametry modelu oszacowanego na odpowiednio przekształconych miarach cen, ilości transakcji i podaży pieniądza, by ilościowa teoria pieniądza okazała się zgodna z danymi?

5. Na danych polskich przeprowadzono regresję $\log(p_t)$ na stałej, $\log(m3_t)$ oraz $\log(pkb_t)$ i uzyskano następujący wynik:

Zmienna	**Współczynnik**	**Błąd stand.**	**t**	$P > \|t\|$
logM3	.4650093	.0472045	9.85	0.000
logPKB	.3149285	.226876	1.39	0.174
Stała	-.0047033	.0087901	-0.54	0.596

Test Breuscha-Godfrey'a	15.484	[0.0001]
Test Breusch-Pagan'a	2.44	[0.1181]
Test RESET	2.25	[0.1011]

   Testując hipotezy proste zbada, czy ograniczenia narzucone na parametry przez ilościową teorię pieniądza (patrz poprzedni podpunkt) są prawdziwe.

6. Zweryfikuj poprawność formy funkcyjnej modelu

7. Zbadaj, czy w modelu występuje heteroskedastyczność

8. Zbadaj, czy w modelu występuje autokorelacja

9. Co implikują, odnośnie wyniku estymacji i wnioskowania statystycznego, uzyskane wyniki testów diagnostycznych? Jak można rozwiązać wykryty przez nie problem?

Rozwiązanie:

1. Zgodnie z ilościową teorią pieniądza to ceny dostosowują się do liczby transakcji i podaży pieniądza. Cena więc jest więc zmienną endogeniczną w modelu.

2. Wychodząc od równania pierwotnego dla okresu t:

$$M_t V_t = P_t T_t$$

   dla okresu $t-1$

$$M_{t-1} V_t = P_{t-1} T_{t-1}$$

   Dzieląc stronami uzyskujemy

$$\frac{M_t}{M_{t-1}} \frac{V_t}{V_{t-1}} = \frac{P_t}{P_{t-1}} \frac{T_t}{T_{t-1}}$$

3. Logarytmując stronami uzyskujemy, że

$$\log\left(\frac{M_t}{M_{t-1}}\right) + \log\left(\frac{V_t}{V_{t-1}}\right) = \log\left(\frac{P_t}{P_{t-1}}\right) + \log\left(\frac{T_t}{T_{t-1}}\right)$$

   zmienną zależną są ceny:

$$\log\left(\frac{P_t}{P_{t-1}}\right) = \log\left(\frac{M_t}{M_{t-1}}\right) + \log\left(\frac{V_t}{V_{t-1}}\right) - \log\left(\frac{T_t}{T_{t-1}}\right)$$

   Zgodnie z ilościową teorią pieniądza V_t jest stałe a więc $\log\left(\frac{V_t}{V_{t-1}}\right) = \log(1) = 0$ i model ma postać

$$\log\left(\frac{P_t}{P_{t-1}}\right) = \log\left(\frac{M_t}{M_{t-1}}\right) - \log\left(\frac{T_t}{T_{t-1}}\right)$$

4. Stosując zaproponowane miary cen, podaży pieniądza i ilości transakcji uzyskujemy model

$$\log(p_t) = \beta_1 + \beta_2 \log(m3_t) + \beta_3 \log(pkb_t) + \varepsilon_t$$

   Model ten będzie zgodny z ilościową teorią pieniądza jeśli prawdziwe będą ograniczenia: $\beta_1 = 0, \beta_2 = 1, \beta_3 = -1$

5. Dla testowanych trzech hipotez prostych statystyki testowe będą miały postać

$$t_i = \frac{b_i - \beta_i}{se(b_i)}$$

   A więc dla

$$\begin{array}{ll} H_0 : \beta_1 = 0 & t_1 = \frac{-.0047033}{.008790} = -.\,535073 \\ H_0 : \beta_2 = 0 & t_2 = \frac{.4650093-1}{.0472045} = -11.\,333 \\ H_0 : \beta_3 = -1 & t_3 = \frac{.31492856+1}{.2268765} = 5.\,795 \end{array}$$

   $|t_1| = .\,535073 < 2.03$, hipotezy $H_0 : \beta_1 = 0$ nie można odrzucić

   $|t_2| = 11.\,333 > 2.03$ hipotezę $H_0 : \beta_2 = 0$ odrzucamy

   $|t_3| = 5.\,795 > 2.03$ hipotezę $H_0 : \beta_3 = -1$ odrzucamy

6. Wartość $p = 0.1011 > 0.05$ wskazuje, że hipotezy o poprawności formy funkcyjnej nie można odrzucić

7. Wartość $p = 0.1181 > 0.05$ wskazuje, że hipotezy braku heteroskedastyczności nie można odrzucić

8. Wartość $p = 0.0001 < 0.05$ wskazuje, że hipotezę o braku autokorelacji musimy odrzucić. Przyjmujemy hipotezę alternatywną o występowaniu autokorelacji

9. Testy wykazały, że w modelu występuje autokorelacja. W tym przypadku współczynniki są poprawnie oszacowane ale niepoprawnie oszacowana jest macierz wariancji kowariancji, w rezultacie niepoprawne są standardowe metody testowania. Możemy ten problem rozwiązać, używając odpornej macierzy Newey'a-Westa, bądź Uogólnionej Metodę Najmniejszych Kwadratów.

ZADANIE 3 Badacz postanowił oszacować funkcję popytu na energię elektryczną na podstawie danych pochodzących z gospodarstw domowych. Badacz przyjął, że funkcja popytu na energię elektryczną ma następującą postać:

$$q_i = \beta_1 + \beta_2 doch_i + \beta_3 p_i + \varepsilon_i$$

gdzie zmienna q_i oznacza konsumpcję energii elektrycznej, $doch_i$ oznacza dochód i-tego gospodarstwa a p_i indeks cen energii elektrycznej w przypadku i-tego gospodarstwa. Wszystkie zmienne wyrażone są w logarytmach. Indeks cen energii został zdefiniowany w taki sposób, by jego wartość oczekiwana była dla całej badanej populacji równa zero, a więc $\mathrm{E}(p_i) = 0$. Niestety badacz nie ma dostępu do danych na temat wysokości tego indeksu dla poszczególnych gospodarstw domowych. Załóżmy, że zmienne objaśniające w modelu są losowe oraz spełnione są założenia przy których estymator MNK jest zgodny.

1. Podaj warunki, przy których można uzyskać zgodny estymator parametru β_2 w przypadku, gdy badacz pominie zmienną p_i.

2. Przyjmijmy, że dochody w dużych miastach są średnio wyższe niż w pozostałych miejscowościach a ceny energii elektrycznej w dużych miastach są średnio niższe niż pozostałych miejscowościach. Jakiego kierunku obciążenia estymatora MNK parametru β_2 należy spodziewać się w takim przypadku? Odpowiedź uzasadnij.

3. Załóżmy, że między wpływem dochodu i cen na poziom konsumpcji energii elektrycznej zachodzi interakcja. Pokaż, że w przypadku pominięcia zmiennej p_i i $doch_i \times p_i$ estymator MNK parametru β_2 nie będzie zgodny, gdy $\operatorname{Cov}(doch_i, p_i) = 0$ ale będzie zgodny, jeśli zmienne $doch_i$ i p_i są od siebie niezależne.

   Podpowiedź: skorzystaj z tego, że $\operatorname{Cov}(x_1, x_2) = \operatorname{E}(x_1 x_2) - \operatorname{E}(x_1)\operatorname{E}(x_2)$ oraz z tego, że dla zmiennych niezależnych x_1 i x_2 wartość oczekiwana $\operatorname{E}(x_1 x_2) = \operatorname{E}(x_1)\operatorname{E}(x_2)$.

4. Pokaż, że w modelu z pominiętymi zmiennymi opisanym w poprzednim punkcie wystąpi heteroskedastyczność błędu losowego.

   Podpowiedź: spróbuj wyprowadzić wariancję warunkową błędu losowego w tym modelu, skorzystaj z tego, że $\operatorname{Var}(x_1 x_2 | x_2) = x_2^2 \operatorname{Var}(x_1)$ oraz $\operatorname{Cov}(x_1, x_2 x_3 | x_3) = x_3 \operatorname{Cov}(x_1, x_2)$.

Rozwiązanie:

1. Estymator MNK pozostaje zgodny w przypadku pominięcia istotnej zmiennej objaśniającej jeśli zmienna ta nie jest skorelowana ze zmiennymi, które zostały uwzględnione w modelu. W naszym przypadku warunek ten jest spełniony jeśli $\operatorname{Cov}(doch_i, p_i) = 0$.

2. Z warunków opisanych w tym podpunkcie wynika, że należy się spodziewać iż współczynnik korelacji $\rho_{doch,p} < 0$. Z teorii wyboru konsumenta wiemy, że cenowa elastyczność popytu jest mniejsza od zera dla dóbr nie będących dobrami Giffena ($\beta_3 < 0$). Ze wzoru na obciążenie wynika więc, że $\operatorname{E}\left(\widetilde{\boldsymbol{\beta}}_2\right) - \beta_2 = \beta_{doch} \frac{s_{doch}}{s_p} \rho_{doch,p} > 0$. Ponieważ energia elektryczna jest dobrem normalnym ($\beta_2 > 0$) więc oszacowanie elastyczności dochodowej będzie prawdopodobnie przeszacowane.

3. Niezbędnym warunkiem zgodności jest brak korelacji między błędem losowym i zmiennymi objaśniającymi. Model z interakcją między $doch_i$ i p_i ma postać:

   $$q_i = \beta_1 + \beta_2 doch_i + \beta_3 p_i + \beta_4 doch_i p_i + \varepsilon_i$$

   model z pominiętymi zmiennymi ma postać:

   $$\begin{aligned} q_i &= \beta_1 + \beta_2 doch_i + u_i \\ u_i &= \beta_3 p_i + \beta_4 doch_i p_i + \varepsilon_i \end{aligned}$$

   Oszacowania parametrów w tym modelu będą zgodne, gdy

   $$\begin{aligned} \operatorname{Cov}(u_i, doch_i) &= \operatorname{Cov}(\beta_3 p_i + \beta_4 doch_i p_i + \varepsilon_i, doch_i) = \beta_3 \underbrace{\operatorname{Cov}(p_i, doch_i)}_{0} + \beta_4 \operatorname{Cov}(doch_i p_i, doch_i) \\ &= \operatorname{E}\left(doch_i^2 p_i\right) - \operatorname{E}(doch_i p_i)\operatorname{E}(doch_i) \neq 0 \end{aligned}$$

   Jeśli jednak zmienne $doch_i$ i p_i są od siebie niezależne, to $\operatorname{E}\left(doch_i^2 p_i\right) = \operatorname{E}\left(doch_i^2\right)\operatorname{E}(p_i)$ oraz $\operatorname{E}(doch_i p_i) = \operatorname{E}(doch_i)\operatorname{E}(p_i)$. W tym szczególnym przypadku

   $$\begin{aligned} \operatorname{Cov}(u_i, doch_i) &= \operatorname{E}\left(doch_i^2 p_i\right) - \operatorname{E}(doch_i p_i)\operatorname{E}(doch_i) = \left\{\operatorname{E}\left(doch_i^2\right) - [\operatorname{E}(doch_i)]^2\right\}\operatorname{E}(p_i) \\ &= \operatorname{Var}(doch_i) \underbrace{\operatorname{E}(p_i)}_{0} = 0 \end{aligned}$$

   ponieważ z założeń $\operatorname{E}(p_i) = 0$. Wynika z tego, że w modelu z interakcją dla niezależnych $doch_i$ i p_i estymator MNK jest zgodny.

4. W przypadku modelu z interakcją i pominiętą zmienną p_i oraz interakcją wariancja błędu losowego jest równa

$$\begin{aligned}
\operatorname{Var}(u_i|\, doch_i) &= \operatorname{Var}(\beta_3 p_i + \beta_4 doch_i p_i + \varepsilon_i|\, doch_i) \\
&= \beta_3^2 \operatorname{Var}(p_i + \varepsilon_i|\, doch_i) + \beta_4^2 \operatorname{Var}(doch_i p_i|\, doch_i) + 2\beta_3\beta_4 \operatorname{Cov}(p_i + \varepsilon_i, doch_i p_i|\, doch_i) \\
&= \beta_3^2 [\operatorname{Var}(p_i) + \operatorname{Var}(\varepsilon_i)] + \beta_4^2 doch_i^2 \operatorname{Var}(p_i) + 2\beta_3\beta_4 doch_i \operatorname{Var}(p_i) \neq const
\end{aligned}$$

a więc wariancja warunkowa zależy od zmiennej objaśniającej $doch_i$ i tym samym w modelu występuje heteroskedastyczność.