Egzamin z Ekonometrii IiE18.06.2015

Pytania teoretyczne

1. Wyjaśnić, na czym polega założenie o niezależności od pozostałych alternatyw i podać przykład, dla którego założenie to nie jest spełnione.
2. Opisać dwustopniową procedurę estymacji modelu Heckmana.
3. Co modelowane jest w przypadku, kiedy zmienna zależna jest zmienną binarną (zerojedynkową)? Wyjaśnić, jaka jest relacja między zmienną obserwowalną a ukrytą w przypadku modeli dla zmiennych binarnych i jak tę relację można uzasadnić w oparciu o teorię ekonomii.
4. Wyjaśnić, dlaczego za pomocą estymatora efektów stałych nie można oszacować parametrów przy zmiennych nie zmieniających się w czasie.

ZADANIE 1 Mamy następujący model podaży i popytu na pracę

$$l_t = \alpha_0 + \alpha_1 w_t + \alpha_2 p_t + \varepsilon_{1t}$$
$$l_t = \beta_1 w_t + \beta_2 q_t + \varepsilon_{2t}$$

gdzie l_t oznacza logarytm zatrudnienia, w_t to logarytm płacy realnej, p_t logarytm wielkości siły roboczej, a q_t logarytm produkcji. Zmiennymi endogenicznymi są l_t i w_t a p_t i q_t są traktowane jako egzogeniczne.

1. Które z tych równań jest według ciebie równaniem popytu, a które podaży? Odpowiedź uzasadnij.
2. Sprawdź identyfikację poszczególnych równań.
3. Wyjaśnij jak wyglądać będzie forma zredukowana tego modelu
4. Wyjaśnij dlaczego równania popytu nie można poprawnie wyestymować za pomocą MNK.
5. Jeśli zastosowalibyśmy do estymacji tego równania MZI, to jakie zmienne mogłyby być użyte jako instrumenty?

Rozwiązanie:

1. Podaż pracy zależy od wielkości siły roboczej oraz wysokości płacy. Popyt na pracę zależy od wysokości płac i wielkości produkcji. Wynika z tego, że równanie pierwsze jest rówanie podaży a równanie drugie równaniem popytu.
2. Podział na zmienne endo i egzogeniczne

zmienne egzogeniczne (wliczamy stałą)	$1, p_t, q_t$
zmienne endogeniczne	l_t, w_t

   G_s liczba zmiennych endogenicznych w s-tym równaniu, K_s liczba zmiennych ezaogenicznych w s-tym równaniu:

$$\begin{array}{lll} K = 3 & G_1 = 2 & K_1 = 2 \\ & G_2 = 2 & K_2 = 1 \end{array}$$

   Identyfikacja:

$$\begin{array}{ll} 3 = 2 + 2 - 1 & (+) \\ 3 > 2 + 1 - 1 & (+) \end{array}$$

   Równanie popytu i podaży na pracę są zidentyfikowane.

3. Równania formy zredukowanej dla pracy i płacy:

$$l_t = \pi_{10} + \pi_{11} p_t + \pi_{12} q_t + \epsilon_{1t}$$
$$w_t = \pi_{20} + \pi_{21} p_t + \pi_{22} q_t + \epsilon_{1t}$$

IMIĘ NAZWISKO..

4. Równania popytu na pracę nie da się wyestymować standardowym MNK, ponieważ zmienna objaśniająca w_t (płaca) jest skorelowana z błędem losowym ε_{1t} z powodu występowania sprzeżenia zwrotnego między popytem na pracę i wysokością płacy (wysokość płacy wpływana na popyt na pracę a równocześnie wielkość popytu na pracę wpływa na płacę). Taka korelacja oznacza występowanie problemu równoczesności, który powoduje, że estymator MNK nie jest zgodny.

5. Instrumentami w przypadku estmacji rówania popytu na pracę powinny być wszystkie zmienne egzaogeniczne występujące w modelu a więc $1, p_t, q_t$.

ZADANIE 2 Badane są czynniki skłaniające responedentów do emigracji. W badaniu wykorzystywane są dane z badania "Diagnoza społeczna" z lat 2000, 2003, 2005, 2007. Poniżej znajdują się oszacowania parametrów i efektów cząskowych policzonych dla średnich w próbie dla modelu probitowego oszacowanego dla tego problemu, w którym zmienną zależną jest odpowiedź na pytanie, czy dany członek gospodarstwa w momencie badania pracuje zagranicą (1 pracuje, 0 nie pracuje). Zmiennymi objaśniającymi w modelu są: logarytm dochodu na głowę w gospodarstwie, liczba lat nauki, zmienna zerojedynkowa związana z płcią respondenta (1 kobieta, 0 mężczyzna) oraz zmienne zerojedynkowe związane z rokiem przeprowadzenia badania.

	parametr	**efekt cząstkowy**	**z**	**p**
ln(dochodu)	-0.292	-0.004	-8.387	0.000
lata nauki	0.064	0.001	8.332	0.000
kobieta	-0.334	-0.005	-7.417	0.000
2003	-0.118	-0.002	-1.240	0.215
2005	0.388	0.008	4.885	0.000
2007	0.598	0.012	7.958	0.000
Constant	-1.550		-7.544	0.000
N	38916	38916		
$R^2_{\text{McKelvey'a i Zavoiny}}$	0.134			
$R^2_{\text{liczebnościowe}}$	0.992			
$\overline{R}^2_{\text{liczebnościowe}}$	0.000			

Testy przeprowadzamy na poziomie istotności $\alpha = 0.05$.

1. Wypisz założenia modelu probitowego.

2. Które założenie $KMRL$ będzie najprawdopodobniej fałszywe dla modelu liniowego opisującego zależność p-stwa emigracji od charakterystyk respondenta?

3. Na jakie problemy natkniemy się, jeśli model ten oszacujemy za pomocą MNK?

4. Jakie wnioski można wyciągnąć na podstawie oszacowań parametrów w tym modelu?

5. Podaj intepretację $R^2_{\text{McKelvey'a i Zavoiny}}$ i $R^2_{\text{liczebnościowe}}$. Wyjaśnij dlaczego występuje tak duża rozbieżność między $R^2_{\text{liczebnościowe}}$ i miarą skorygowaną $\overline{R}^2_{\text{liczebnościowe}}$.

   Podpowiedź: W badanej próbie znajduje się jedynie 317 obserwacji, dla których respondenci w momencie badania pracowali zagranicą.

6. Podaj intepretację efektu cząstkowego dla płci (kodowanie 0 - mężczyzna, 1 kobieta).

7. Zinterpretuj efekt cząskowy dla logarytmu dochodu na głowę w gospodarstwie.

8. Jaki model można by zastosować, by zbadać hipotezę, że wykształcenie wpływa w różny sposób na skłonność do emigracji kobiet i mężczyzn?

9. Przetestowano hipotezę, że rok badania nie ma wpływu na skłonność do emigracji respondentów uzyskano i statystykę $LR = 154.95$. Zweryfikuj powyższą hipotezę.

   Podpowiedź: $\chi^2_{0.95}(1) = 3.84$, $\chi^2_{0.95}(2) = 5.99$, $\chi^2_{0.95}(3) = 7.81$.

Rozwiązanie:

IMIĘ NAZWISKO...

1. Zmienna ukryta

$$\begin{aligned} y^* &= \boldsymbol{x\beta} + \varepsilon \\ \varepsilon &\sim N(0,1) \end{aligned}$$

i poszczególne obserwacje są niezależne. Obserwujemy

$$\begin{aligned} y = 0 &\quad \text{dla} \quad y^* \leq 0 \\ y = 1 &\quad \text{dla} \quad y^* > 0 \end{aligned}$$

Alternatywna odpowiedź: prawdopodobieństwa zajścia pojedynczego zdarzenia:

$$\Pr(y_i) = \begin{cases} 1 - \Phi(\boldsymbol{x\beta}) & \text{dla} \quad y = 0 \\ \Phi(\boldsymbol{x\beta}) & \text{dla} \quad y = 1 \end{cases}$$

i poszczególne obserwacje są niezależne.

2. W przypadku modelu liniowego (Liniowego Modelu Prawdopodobieństwa - LPM) fałszywe jest założenie o homoskedastyczności, ponieważ $\mathrm{Var}(y) = \mathrm{Var}(\varepsilon) = p(\boldsymbol{x})[1 - p(\boldsymbol{x})] = \boldsymbol{x\beta}(1 - \boldsymbol{x\beta})$ i jest zależne od $\boldsymbol{x}$.

3. Szacując model LPM natkniemy się na dwa problemy: po pierwsze na wspomnianą wyżej heteroskeadstyczność a po drugie możemy uzyskać nieintepretowalen dopasowane wartości prawdopodobieństw (spoza przedziału $[0,1]$)

4. Na podstawie oszacowań uzyskanych z modelu możemy zbadać istotność oraz kierunek wpływu poszczególnych zmiennych. Istotne są zmienne: lata nauki, kobieta, 2005, 2007, oraz stała. Dodatni wpływ na p-stwo posiadania emigracji ma zmienna lata nauki, 2005, 2007. Zmienna kobieta ma ujemny wpływ na p-stwo emigracji. Zmienna 2003 ma nieistotny wpływ na p-stwo emigracji co oznacza, że nie było istotnej różnicy między rokiem 2000 i 2003 jeśli chodzi o p-stwo emigracji osób o identycznych charakterystykach.

5. Intepretacja $R^2_{\text{McKelvey'a i Zavoiny}}$: $13,4\%$ zmienności nieobserwowalnej zmiennej ukrytej została wyjaśniona przez zmienne niezależne. Intepretacja $R^2_{\text{liczebnościowe}}$: w przypadku $99,2\%$ obserwacji model prawidłowo przewidział sukces bądź porażkę. $\overline{R}^2_{\text{liczebnościowe}}$ jest skorygowane o liczebność najbardziej liczebnej kategorii. W analizowanym przypadku najliczniejsza kategoria (porażka), to $\frac{38916-317}{38916} = 99.2\%$ obserwacji. Model, który zawsze wskazywałby na tą kategorię przewidywałby prawidłowo 99.2% obserwacji. Po zastosowaniu korekty okazuje się, że zmienne mają bardzo mały wpływ na liczbę prawidłowo przewidzianch zdarzeń.

6. Wielkość efektu cząstkowego dla zmiennej kobieta oznacza, że kobiety o charakterystykach na poziomie średnich w próbie mają o 0.5 punkta procentowego mniejsze prawdopodobieństwo emigracji w stosunku do mężczyzn o charakterystykach na poziomie średnich w próbie.

7. Wielkość efektu cząstkowego dla zmiennej $\ln(doch\acute{o}d)$ oznacza, że wzrost dochodu na głowę w gospodarstwie o 1% związany jest ze spadkiem p-stwa emigracji o 0.4 punktu procentowego.

8. Należy do modelu wstawić interakcję między płcią a latami nauki (kobieta$\times$lata nauki) oszacować tak rozbudowany model i przetestować istotność współczynnika przy dodanej zmiennej

9. Testując hipotezę łączną o nieistotności roku badania badamy hipotezę łączną o niestotności parametrów przy trzech zmiennych: 2003, 2005, 2007. Statystyka testu LR ma więc rozkład $\chi^2(3)$. Ponieważ $LR = 154.95 > 7.81$ więc H_0 o niestotności roku badania odrzucamy.

ZADANIE 3 Zmienna losowa y_i pochodzi z rozkładu Poissona z parametrem $\lambda_i = \exp(\boldsymbol{x}_i\boldsymbol{\beta})$ oraz zakładamy, że y_i i y_j są niezależne dla $i \neq j$.

1. Pokazać, że estymator MNW w tym modelu znajdujemy rozwiązując względem $\boldsymbol{\beta}$ następujący (nieliniowy) układ równań:

$$\sum_{i=0}^{N} e_i \boldsymbol{x}_i = \boldsymbol{0} \qquad (*)$$

gdzie $e_i = y_i - \lambda_i$.

Podpowiedź: $\Pr(y_i = k) = \frac{e^{-\lambda_i}\lambda_i^k}{k!}$

IMIĘ NAZWISKO..

2. Pokazać, że estymator UMM w tym modelu możemy znaleźć rozwiązując układ równań identyczny do układu (*).

   Podpowiedź: W rozpatrywanym modelu $\mathrm{E}\left(y_i|\,\boldsymbol{x}_i\right) = \lambda_i$.

3. Pokazać, że w przypadku występowania stałej w modelu, średnia dla wartości dopasowanych zmiennej zależnej $\widehat{y}_i = \exp\left(\boldsymbol{x}_i\widetilde{\boldsymbol{\beta}}\right)$ uzyskana z oszacowanego modelu Poissona jest równa średniej z wielkości zaobserwowanych dla zmiennej zależnej y_i.

4. W przypadku modelu Poissona podnosi się często, że istotnym jego ograniczeniem jest założenie, że $\mathrm{E}\left(y_i|\,\boldsymbol{x}_i\right) = \mathrm{Var}\left(y_i|\,\boldsymbol{x}_i\right)$. Na podstawie wyników z poprzednich punktów wykazać, że estymator MNW parametru $\boldsymbol{\beta}$ uzyskany z modelu Poissona jest zgodny nawet wtedy, gdy założenie o równości wartości oczekiwanej i wariancji nie jest spełnione.

Rozwiązanie:

1. Zauważmy, że

$$\ln \Pr\left(y = k\right) = \ln\left(\frac{e^{-\lambda}\lambda^k}{k!}\right) = -\lambda + k\ln\lambda - \ln\left(k!\right)$$

$$\frac{\partial \ln \Pr\left(y = k\right)}{\partial\lambda} = \frac{k}{\lambda} - 1$$

   Logarytm funkcji wiarygodności jest więc równy

$$\frac{\partial\ell\left(\boldsymbol{\beta}\right)}{\partial\boldsymbol{\beta}} = \frac{\partial\sum_{i=1}^{N}\ln\Pr\left(y_i\right)}{\partial\boldsymbol{\beta}} = \sum_{i=1}^{N}\frac{\partial\ln\Pr\left(y_i\right)}{\partial\lambda_i}\frac{\partial\lambda_i}{\partial\boldsymbol{\beta}}$$

$$= \sum_{i=1}^{N}\left(\frac{y_i}{\lambda_i} - 1\right)\lambda_i\boldsymbol{x} = \sum_{i=1}^{N}\left(y_i - \lambda_i\right)\boldsymbol{x} = \sum_{i=1}^{N}e_i\boldsymbol{x}$$

   a warunek pierwszego rzędu na maksymalizację logarytmu funkcji wiarogodności ma rzeczywiście postać układu równań:

$$\sum_{i=1}^{N}e_i\boldsymbol{x} = \mathbf{0}$$

2. Moment warunkowy

$$\mathrm{E}\left(y_i|\,\boldsymbol{x}_i\right) = \lambda_i$$

   wynika z tego, że

$$\mathrm{E}\left(y_i - \lambda_i|\,\boldsymbol{x}_i\right) = \mathrm{E}\left(\varepsilon_i|\,\boldsymbol{x}_i\right) = 0$$

   gdzie $\varepsilon_i = y_i - \lambda_i$.

   Moment bezwarunkowy

$$\mathrm{E}\left(\varepsilon_i\boldsymbol{x}_i\right) = 0$$

   Odpowiednik próbkowy

$$\frac{1}{N}\sum_{i=1}^{N}e_i\boldsymbol{x}_i = 0$$

3. W przypadku, gdy w modelu jest stała, jednym z elementów wektora $\boldsymbol{x}_i$ będzie 1 a jednym z równań (*) będzie równanie $\sum_{i=0}^{N}e_i = 0$. To z kolei i definicja e_i implikują, że $\sum_{i=1}^{N}y_i = \sum_{i=1}^{N}\widehat{\lambda}_i = \sum_{i=1}^{N}\exp\left(\boldsymbol{x}_i\widetilde{\boldsymbol{\beta}}\right) = \sum_{i=1}^{N}\widehat{y}_i$. Dzieląc obie strony przez N uzyskujemy $\overline{y} = \overline{\widehat{y}}$.

4. Postać estymatora MNW w modelu Poissona jest taka sama jak estymatora UMM, jednak estymator UMM nie wykorzystuje założenia, że $\mathrm{Var}\left(y_i|\,\boldsymbol{x}_i\right) = \lambda_i$. Estymator UMM jest więc zgodny nawet, gdy założenie to nie jest spełnione a zatem to samo dotyczy estymatora MNW.

IMIĘ NAZWISKO...