Egzamin z ekonometrii 12.06.2009
2 semestr, Informatyka i Ekonometria

Pytania teoretyczne

1. Podaj dwie najważniejsze zalety estymatorów MNW. Jakie założenie, które należy spełnić, by oszacować model tą metodą, wydaje ci się najmniej realistyczne w zastosowaniach ekonometrycznych?
2. Na czym polega różnica między wielomianowym i warunkowym modelem logitowym?
3. Kiedy selekcja próby staje się problemem i jaki model można stosować w przypadku samoselekcji próby?
4. Co rozumiemy przez estymatory Quasi ML i kiedy je stosujemy?

ZADANIE 1 Postanowiono oszacować model Poissona na próbie panelowej. Założono, że obserwacje dla poszczególnych jednostek dobrane są losowo. Założono też, że model Poissona ma następująco postać:

$$\Pr\left(y = y_{it} \middle| \boldsymbol{x}_{it}, u_i\right) = \frac{\exp\left[-\mu\left(\boldsymbol{x}_{it}, u_i\right)\right]\left[\mu\left(\boldsymbol{x}_{it}, u_i\right)\right]^{y_{it}}}{y_{it}!}$$

a

$$\mu\left(\boldsymbol{x}_{it}, u_i\right) = \mu\left(\boldsymbol{x}_{it}\right) u_i$$

gdzie $\mu\left(\boldsymbol{x}_{it}\right) = \exp\left(\boldsymbol{x}_{it}\boldsymbol{\beta}\right)$, y_{it} jest zmienną zależną a $\boldsymbol{x}_{it}$ jest wektorem zmiennych niezależnych. Szacowany jest parametr $\boldsymbol{\beta}$ oraz parametry u_i dla $i = 1, \ldots n$.

1. Sformułuj logarytm funkcji wiarygodności dla tego modelu.
2. Wyprowadź estymator MNW parametrów u_i, $i = 1, \ldots n$. Znaleź warunki, z których można policzyć estymator $\widetilde{\boldsymbol{\beta}}$.
3. W przypadku estymatora parametru u_i pokaż, że dla $n \longrightarrow \infty$ i T skończonego estymator ten jest asympotycznie nieobciążony ($\lim\limits_{n\to\infty} \mathrm{E}\left(\widetilde{u}_i\right) = u_i$) ale nie jest zgodny.

   Podpowiedź: załóż, że estymator parametru $\boldsymbol{\beta}$ jest zgodny. W związku z tym przyjmij, że dla $n \to \infty$ parametr $\boldsymbol{\beta}$ jest znany. Ponadto przyjmij, że jeśli wariancją estymatora nie dąży do zera dla $n \to \infty$, to nie jest on zgodny. Skorzystaj także z tego, że dla modelu Poissona $\mathrm{Var}\left(y = y_{it} \middle| \boldsymbol{x}_{it}\right) = \exp\left(\boldsymbol{x}_{it}\boldsymbol{\beta}\right) u_i$.

Rozwiązanie:

1. W przypadku tego estymatora

$$\Pr\left(\boldsymbol{y} \middle| \boldsymbol{X}\right) = \prod_{i=1}^{n}\prod_{t=1}^{T} \frac{\exp\left[-\mu\left(\boldsymbol{x}_{it}, u_i\right)\right]\left[\mu\left(\boldsymbol{x}_{it}, u_i\right)\right]^{y_{it}}}{y_{it}!}$$

$$\begin{aligned}
\ell\left(\boldsymbol{\beta}, \boldsymbol{u}\right) &= \sum_{i=1}^{n}\sum_{t=1}^{T}\left[-\mu\left(\boldsymbol{x}_{it}, u_i\right) + y_{it}\ln\mu\left(\boldsymbol{x}_{it}, u_i\right) - \ln y_{it}!\right] \\
&= \sum_{i=1}^{n}\sum_{t=1}^{T}\left[-\exp\left(\boldsymbol{x}_{it}\boldsymbol{\beta}\right)u_i + y_{it}\boldsymbol{x}_{it}\boldsymbol{\beta}u_i - \ln y_{it}!\right] \\
&= \sum_{i=1}^{n}\ln u_i\sum_{t=1}^{T} y_{it} + \sum_{i=1}^{n}\sum_{t=1}^{T}\left[y_{it}\boldsymbol{x}_{it}\,\boldsymbol{\beta} - u_i\exp\left(\boldsymbol{x}_{it}\boldsymbol{\beta}\right)\right] - \sum_{i=1}^{n}\sum_{t=1}^{T}\ln y_{it}!
\end{aligned}$$

2. Warunki pierwszego rzędu na maksymalizację $\ell\left(\boldsymbol{\beta}, \boldsymbol{u}\right)$ względem $\boldsymbol{u}$ są następujące:

$$\frac{\partial\ell\left(\boldsymbol{\beta}\right)}{\partial u_i} = \frac{1}{u_i}\sum_{t=1}^{T} y_{it} - \sum_{t=1}^{T}\exp\left(\boldsymbol{x}_{it}\boldsymbol{\beta}\right) = 0$$

$$\sum_{t=1}^{T} y_{it} = u_i\sum_{t=1}^{T}\exp\left(\boldsymbol{x}_{it}\boldsymbol{\beta}\right)$$

Estymator MNW parametru u_i jest więc dany wzorem:

$$\widetilde{u}_i = \frac{\sum_{t=1}^{T} y_{it}}{\sum_{t=1}^{T} \exp\left(\boldsymbol{x}_{it}\widetilde{\boldsymbol{\beta}}\right)}$$

Warunki pierwszego rzędu na maksymalizację $\ell(\boldsymbol{\beta}, \boldsymbol{u})$ względem $\boldsymbol{\beta}$ są następujące:

$$\frac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n}\sum_{t=1}^{T}\left[y_{it} - u_i \exp(\boldsymbol{x}_{it}\boldsymbol{\beta})\right]\boldsymbol{x}'_{it} = \boldsymbol{0}$$

Estymator MNW znajdujemy rozwiązując warunki pierwszego rzędu dla wszystkich parametrów. Ponieważ mamy już rozwiązanie dla $\widetilde{u}_i$ można sformułować równanie dla $\widetilde{\boldsymbol{\beta}}$ w sposób następujący:

$$\sum_{i=1}^{n}\sum_{t=1}^{T}\left[y_{it} - \frac{\sum_{t=1}^{T} y_{it}}{\sum_{t=1}^{T} \exp\left(\boldsymbol{x}_{it}\widetilde{\boldsymbol{\beta}}\right)}\exp\left(\boldsymbol{x}_{it}\widetilde{\boldsymbol{\beta}}\right)\right]\boldsymbol{x}'_{it} = \boldsymbol{0}$$

rozwiązując (numerycznie) ten warunek znajdujemy oszacowanie $\widetilde{\boldsymbol{\beta}}$.

3. Wartość oczekiwana i wariancja $\widetilde{u}_i$ dla $n \to \infty$ jest równa

$$\lim_{n\to\infty} \mathrm{E}\left(\widetilde{u}_i | \boldsymbol{X}\right) = \frac{\sum_{t=1}^{T}\exp(\boldsymbol{x}_{it}\boldsymbol{\beta})\, u_i}{\sum_{t=1}^{T}\exp(\boldsymbol{x}_{it}\boldsymbol{\beta})} = u_i$$
$$\lim_{n\to\infty} \mathrm{Var}\left(\widetilde{u}_{it} | \boldsymbol{X}\right) = \frac{u_i}{\sum_{t=1}^{T}\exp(\boldsymbol{x}_{it}\boldsymbol{\beta})}$$

ponieważ wyniki te nie zależą od $\boldsymbol{X}$ więc

$$\lim_{n\to\infty} \mathrm{E}\left(\widetilde{u}_{it}\right) = u_i$$
$$\lim_{n\to\infty} \mathrm{Var}\left(\widetilde{u}_{it}\right) = \frac{u_i}{\sum_{t=1}^{T}\exp(\boldsymbol{x}_{it}\boldsymbol{\beta})} \neq 0$$

Asymptotyczna wariancja estymatora nie jest równa zero dla $n \to \infty$, estymator nie jest zgodny.

ZADANIE 2 Szacujemy model

$$y_i = \beta_0 + x_{1i}^{\beta_1} + \beta_2 x_{2i} + \varepsilon_i,$$
$$\mathrm{E}\left(\varepsilon_i | x_{1i}, x_{2i}\right) = 0.$$

1. Sformułuj warunki narzucone na momenty teoretyczne oraz równania, na podstawie których można (numerycznie) policzyć estymator UMM dla tego modelu.
2. Czy dla estymatora UMM suma reszt $e_i = y_i - \widetilde{\beta}_0 - x_{1i}^{\widetilde{\beta}_2} - \widetilde{\beta}_2 x_{1i}$ w tym modelu będzie równa zeru?
3. Dla jakiej hipotezy model ten stanie się modelem liniowym?
4. Chcemy oszacować nieco inną warsję modelu:

$$y_i = \beta_0 + \beta_1 x_{1i}^{\beta_2} + \varepsilon_i,$$
$$\mathrm{E}\left(\varepsilon | x_i\right) = 0.$$

Sformułuj równania na podstawie których można by policzyć estymator UMM dla tego modelu.

Podpowiedź: zwróć uwagę na problem identyfikacji oraz skorzystaj z faktu, że jeśli $\mathrm{E}\left(\varepsilon_i | x_{1i}\right) = 0$, to $\mathrm{E}\left(w\left(x_{1i}\right)\varepsilon_i\right) = 0$ dla dowolnej funkcji w.

Rozwiązanie:

1. Wykorzystując standardowe sposób postępowania w przypadku modeli UMM uzyskujemy następujące ograniczenia narzucone na bezwarunkowe momenty teoretyczne

$$\begin{aligned} \mathrm{E}\left[\mathrm{E}\left(\varepsilon \mid \boldsymbol{x}_i\right)\right] &= \mathrm{E}\left(\varepsilon\right) = 0 \\ \mathrm{E}\left(\boldsymbol{x}_i' \varepsilon\right) &= 0 \end{aligned}$$

   oraz

$$\varepsilon_i = y_i - \beta_0 - x_{1i}^{\beta_1} - \beta_2 x_{2i}$$

   Na podstawie tych ograniczeń możemy formułujemy następujące ograniczenia narzucone na momenty empiryczne

$$\begin{aligned} \frac{1}{n}\sum\left(y_i - \widetilde{\beta}_0 - x_{1i}^{\widetilde{\beta}_1} - \widetilde{\beta}_2 x_{2i}\right) &= 0 \\ \frac{1}{n}\sum\left(y_i - \widetilde{\beta}_0 - x_{1i}^{\widetilde{\beta}_1} - \widetilde{\beta}_2 x_{2i}\right) x_{1i} &= 0 \\ \frac{1}{n}\sum\left(y_i - \widetilde{\beta}_0 - x_{1i}^{\widetilde{\beta}_1} - \widetilde{\beta}_2 x_{2i}\right) x_{2i} &= 0 \end{aligned}$$

   rozwiązując te równania dla $\widetilde{\beta}_0$, $\widetilde{\beta}_1$, $\widetilde{\beta}_2$ znajdujemy estymator UMM

2. Z pierwszego równania, które powinno być spełnione dla estymatora UMM wynika, że istotnie suma reszt powinna być równa zeru.

3. Model ten stanie się modelem liniowym dla $\beta_1 = 1$

4. Postępując w podobny sposób co poprzednio uzyskujemy dwa równania UMM

$$\begin{aligned} \frac{1}{n}\sum\left(y_i - \widetilde{\beta}_0 - \widetilde{\beta}_1 x_{1i}^{\widetilde{\beta}_2}\right) &= 0 \\ \frac{1}{n}\sum\left(y_i - \widetilde{\beta}_0 - \widetilde{\beta}_1 x_{1i}^{\widetilde{\beta}_2}\right) x_{1i} &= 0 \end{aligned}$$

   Jednak nie udało nam się uzyskać w ten sposób identyfikacji ponieważ liczba równań jest mniejsza od liczby parametrów. Skoro jednak jeśli $\mathrm{E}\left(\varepsilon_i \mid x_{1i}\right) = 0$, to $\mathrm{E}\left(w\left(x_{1i}\right)\varepsilon_i\right) = 0$, to możmy wziąść dowolną funkcję x_{1i} np. x_{1i}^2 i uzyskać trzecie równanie

$$\frac{1}{n}\sum\left(y_i - \widetilde{\beta}_0 - \widetilde{\beta}_1 x_{1i}^{\widetilde{\beta}_2}\right) x_{1i}^2 = 0$$

ZADANIE 3 Następujący model popytu na ziemię uprawną (zmienna $land_i$ wyrażona w akrach) został opublikowany w roku 1981 przez Roda F. Ziemera and Freda C. White'a:

$$land_i = \beta_0 + \beta_1 price_i + \beta_2 size_i + \beta_3 purchased_i + \beta_4 age_i + \beta_5 age_i^2 + \sum_{k=1}^{4}\gamma_k income_{ki} + u_i$$

gdzie zmienna $price_i$ jest ceną za akr ziemi w tysiącach dolarów, $size$ jest całkowitą wielkością farmy w akrach, $purchased_i$ liczbą akrów ziemi zakupioną wcześniej przez farmera, age jest wiekiem farmera w latach, $income_{ki}$ są to zmienne zerojedynkowe związane z wysokością dochodu farma z innych źródeł niż farma (0 for $income = 0$, 1 for $0 < income \leq 10000\$$; 2 for $10000\$ < income \leq 20000\$$; 3 for $20000\$ < income \leq 50000\$$; 4 for $income > 50000\$$). Model został wyestymowany na podstawie danych z próby losowej zebranej w USA. Użyto jedynie obserwacji dotyczących farm w stanie Georgia, których właściciele poczynili zakupy ziemi w latach $1970 - 1978$. Wyniki estymacji znajdują się poniżej.
Odpowiedzi proszę uzasadniać wielkością statystyk. Założony poziom istotności $\alpha = 0.05$.

1. Wyjaśnij dlaczego użyta próba nie jest próbą losową dla farm w stanie Georgia.

2. Jaki rodzaj selekcji wystąpił w przypadku analizowej próbki?

3. Pokaż, że estymator MNK nie jest zgodny w tym modelu. Przyjmij następujące założenia.

   Zmienna ukryta y_i^*:

$$y_i^* = \boldsymbol{x}_i \boldsymbol{\beta} + \varepsilon_i$$
$$\varepsilon \sim N\left(0, \sigma^2\right)$$

   Zmienna obserwowalna y_i (to jest $land_i$):

$$y_i^* > 0, \quad y_i = y_i^*, \boldsymbol{x}_i \text{ są obserwowalne}$$
$$y_i^* \leq 0, \quad y_i = 0 \text{ ale } y_i, \boldsymbol{x}_i \text{ są nieobserwowalne}$$

   Podpowiedź:

$$\mathrm{E}\left(y_i | \boldsymbol{x}_i, \varepsilon_i > -\boldsymbol{x}_i\boldsymbol{\beta}\right) = \boldsymbol{x}_i\, \boldsymbol{\beta} + \sigma\lambda\left(\boldsymbol{x}_i\boldsymbol{\beta}\right) \text{ i } \lambda\left(\boldsymbol{x}_i\boldsymbol{\beta}\right) = \frac{\phi\left(\boldsymbol{x}_i\boldsymbol{\beta}/\sigma\right)}{\Phi\left(\boldsymbol{x}_i\boldsymbol{\beta}/\sigma\right)}$$

4. Podaj, na postawie wyników estymacji podanych poniżej, które zmienne okazała się istotne w modelu.

5. Zintepretuj efekty cząstkowe dla $\mathrm{E}\left(y_i\right)$ oraz $\mathrm{E}\left(y_i | y_i > 0\right)$ dla zmiennych $price$ i $size$. których oszacowania podano poniżej.

Zmienna	Współczynnik	odch. std	z	wartość p
stała	7493.640	556.394	13.468	0.000
price	-0.940	0.119	-7.899	0.000
size	0.976	0.064	15.250	0.000
purchased	-0.569	0.053	-10.736	0.000
age	-185.295	21.255	-8.718	0.000
age^2	2.173	0.229	9.489	0.000
income$_1$	-2949.731	181.364	-16.264	0.000
income$_2$	-3029.300	185.999	-16.287	0.000
income$_3$	-2592.900	184.554	-14.050	0.000
income$_4$	-3285.570	223.135	-14.725	0.000

Zmienna	Efekt cząstkowy dla $\mathrm{E}(y)$	Efekt cząstkowy dla $\mathrm{E}(y \| y > 0)$
price	-0.938	-0.862
size	0.974	0.895
purchased	-0.568	-0.522
age	-184.924	-169.916
age^2	2.169	1.993
income$_1$	-2944.030	-2705.086
income$_2$	-3023.241	-2777.868
income$_3$	-2587.714	-2377.689
income$_4$	-3278.999	-3012.868

Rozwiązanie:

1. Próba nie jest próbą losową ponieważ uwględniono w niej jedynie farmy, których właściciele poczynili zakupy ziemi w latach $1970 - 1978$.

2. Selekcja próby w tym przypadku została dokonana na podstawie wielkości zmiennej zależnej.

3. Warunkowa wartość oczekiwana dla wartości zmiennej zależnej pod warunkiem dobrania jej do próbki jest równa

$$\mathrm{E}\left(y_i | \boldsymbol{x}_i, \varepsilon_i > -\boldsymbol{x}_i\boldsymbol{\beta}\right) = \boldsymbol{x}_i\, \boldsymbol{\beta} + \sigma\lambda\left(\boldsymbol{x}_i\boldsymbol{\beta}\right),$$

   ponieważ tylko dla $y_i^* = \boldsymbol{x}_i\, \boldsymbol{\beta} + \varepsilon_i > 0$ a więc dla $\varepsilon_i > -\boldsymbol{x}_i\boldsymbol{\beta}$ możliwe jest znalezienie się obserwacji w próbie. W standardowym modelu liniowym zakładamy, że $\mathrm{E}\left(y_i | \boldsymbol{x}_i\right) = \boldsymbol{x}_i\boldsymbol{\beta}$ a tym samym pomijamy element $\sigma\lambda\left(\boldsymbol{x}_i\boldsymbol{\beta}\right)$. Pominięcie tego elementu prowadzi do braku zgodności oszacowania parametru $\boldsymbol{\beta}$**,** ponieważ zazwyczaj $\boldsymbol{x}_i$ i pominięty element $\sigma\lambda\left(\boldsymbol{x}_i\boldsymbol{\beta}\right)$ są skorelowane a zatem powstaje problem zmiennej pominiętej.

4. Wszystkie zmienne w modelu okazały się istotne co widać po wartościach p, które w każdym przypadku są równe 0.000 a tym samym mniejsze od 0.05.

5. Wzrost ceny akra o jeden tysiąc dolarów zmniejsza oczekiwane zakupy ziemi przeciętnego farmera w Georgi o 0.938 akra i zmiejsza oczekiwany zakup ziemidla tych farmerów, którzy zdecydowali się na jej zakup o 0.862 akra. Zwiększenie wielkości farmy o jeden akr powodwało wzros zakupów ziemi o 0.974 w przypadku przeciętnego farmera w Georgi i wzrost o 0.895 w przypadku tych farmerów, którzy nabyli ziemię w badanym okresie.