# Zawansowane modele wyborów dyskretnych

Jerzy Mycielski

Uniwersytet Warszawski

grudzien 2013

# Model efektów nieobserwowalych dla prób panelowych

- Zmienna ukryta

$$y_i^* = \mathbf{x}_{it}\,\boldsymbol{\beta} + c_i + u_{it}$$

- Zmienna obserwowalna

$$y_i = \begin{cases} 1 & \text{dla} \quad y_i^* \geq 0 \\ 0 & \text{dla} \quad y_i^* < 0 \end{cases}$$

- Kluczowe założenie:
  - $c_i$ zależne od $\mathbf{x}_{it} \implies$ model efektów stałych
  - $c_i$ niezależne od $\mathbf{x}_{it} \implies$ model efektów losowych

# Estymator efektów stałych - problem parametrów pobocznych

- Warunkowa funkcja prawdopodobieństwa:

$$\Pr\left(y_i = 1 | \, c_i\right) = \Pr\left(\mathbf{x}_{it}\,\boldsymbol{\beta} + c_i + u_{it} \geq 0\right) = F\left(\mathbf{x}_{it}\,\boldsymbol{\beta} + c_i\right)$$

- Czy można w tym przypadku zastosować estymator analogiczny do estymatora *LSDV* i oszacować model z dodatkowymi zmiennymi zerojedynkowymi dla każdej jednostki?
- Niestety estymator taki będzie zgodny tylko dla $T \to \infty$ i $N \to \infty$, dla $T$ ustalonego nie będzie zgodny

# Estymator efektów stałych - problem parametrów pobocznych c.d.

- Estymator *LSDV* w liniowym modelu efektów nieobserwowalnych jest o tyle wyjątkowy, że można dla niego pokazać

$$\widehat{\beta}_{FE} \xrightarrow{p} \beta \text{ dla } N \to \infty \text{ i } T \text{ ustalonego}$$

  mimo, że $\widehat{\mathbf{c}}$ nie jest zgodny ($\widehat{\mathbf{c}} \overset{p}{\nrightarrow} \mathbf{c}$ dla $N \to \infty$ i $T$ ustalonego) co wynika z asymptotycznej niezależności obu estymatorów
- Wynik ten generalnie nie jest prawdziwy w przypadku modeli nieliniowych
  - problem parametrów pobocznych (incidental parameters problem)

# Model logitowy z efektami stałymi

- Okazuje się, że w przypadku modelu logitowegoprawdopodobieństwa warunkowego względem ilości sukcesów $n_i$ dla jednostki $i$, nie zależy od $c_i$
- Na przykład dla $T = 2$

$$\Pr\left(y_{i2} = 1 \middle| \mathbf{x}_i, c_i, n_i = 1\right) = \frac{\Pr\left(y_{i2} = 1, n_i = 1 \middle| \mathbf{x}_i, c_i\right)}{\Pr\left(n_i = 1 \middle| \mathbf{x}_i, c_i\right)}$$

$$= \frac{\Pr\left(y_{i2} = 1 \middle| \mathbf{x}_i, c_i\right) \Pr\left(y_{i1} = 0 \middle| \mathbf{x}_i, c_i\right)}{\Pr\left(y_{i1} = 0, y_{i2} = 1 \middle| \mathbf{x}_i, c_i\right) + \Pr\left(y_{i1} = 1, y_{i2} = 0 \middle| \mathbf{x}_i, c_i\right)}$$

$$= \frac{\Lambda\left(\mathbf{x}_{i2}\boldsymbol{\beta} + c_i\right)\left[1 - \Lambda\left(\mathbf{x}_{i1}\boldsymbol{\beta} + c_i\right)\right]}{\left[1 - \Lambda\left(\mathbf{x}_{i1}\boldsymbol{\beta} + c_i\right)\right]\Lambda\left(\mathbf{x}_{i2}\boldsymbol{\beta} + c_i\right) + \Lambda\left(\mathbf{x}_{i1}\boldsymbol{\beta} + c_i\right)\left[1 - \Lambda\left(\mathbf{x}_{i2}\boldsymbol{\beta} + c_i\right)\right]}$$

$$= \Lambda\left[\left(\mathbf{x}_{i2} - \mathbf{x}_{i1}\right)\boldsymbol{\beta}\right]$$

# Model logitowy z efektami stałymi c.d.

- Ogólnie

$$\Pr\left(\mathbf{y}_i \middle| \mathbf{x}_i, c_i, n_i = n\right) = \frac{\exp\left(\sum_{i=1}^{T} y_{it}\mathbf{x}_{it}\boldsymbol{\beta}\right)}{\exp\left(\sum_{\mathbf{a}\in R_i} a_t y_{it}\mathbf{x}_{it}\boldsymbol{\beta}\right)}$$

where $R_i \in R^T$ oraz $R_i = \left\{\mathbf{a} : a_t \in \{0,1\}, \sum_{i=1}^{T} a_i = n_i\right\}$

- Funkcję wiarygodności budujemy przy użyciu tej prawdopodobieństwa

# Model probitowy z efektami losowymi

- Obserwacje dla poszczególnych jednostek nie są niezależne
- Obserwacje dla różnych jednostek są od siebie niezależne
  - zgodnym ale nieefektywnym estymatorem jest w takim przypadku estymator $pseudo - ML$ (identyczny w sensie obliczeniowym do estymatora $ML$).
  - macierz wariancji dla oszacowań powinniśmy w tym przypadku uzyskać za pomocą estymatora warstwowego (cluster )

- Warunkowa funkcja prawdopodobieństwa dla obserwacji dla jednostki $i$:

$$\Pr\left(\mathbf{y}_i \middle| c_i\right) = \prod_{t=1}^{T} \left[\Phi\left(\mathbf{x}_{it}\,\boldsymbol{\beta} + c_i\right)\right]^{y_{it}} \left[1 - \Phi\left(\mathbf{x}_{it}\,\boldsymbol{\beta} + c_i\right)\right]^{1-y_{it}}$$

- Załóżmy, że efekty losowe są pochodzą z wielowymiarowego rozkładu normalnego i są niezależne:

$$\mathbf{c} \sim N\left(\mathbf{0}, \sigma_u^2 \mathbf{I}\right)$$

- Bezwarunkowa funkcja prawdopodbieństwa ma więc następującą postać:

$$\Pr\left(\mathbf{y}_i\right) = \int_{-\infty}^{\infty} \prod_{t=1}^{T} \left[\Phi\left(\mathbf{x}_{it}\,\boldsymbol{\beta} + c_i\right)\right]^{y_{it}} \left[1 - \Phi\left(\mathbf{x}_{it}\,\boldsymbol{\beta} + c_i\right)\right]^{1-y_{it}} \times \times \frac{1}{\sigma_c} \boldsymbol{\phi}\left(\frac{c_i}{\sigma_c}\right) dc_i$$

- Tego rodzaju całki nie da się policzyć analitycznie - zamiast tego stosujemy metody numeryczne (kwadratury)
- W modelu probitowym z efektami losowymi obserwacje są niezależne pomiędzy jednostkami, więc funkcja wiarygodności ma standardową postać:

$$\ell\left(\boldsymbol{\beta}\right) = \sum_{i=1}^{N} \ln \Pr\left(\mathbf{y}_i\right)$$

# Kwadratury

- Kwadratura polega na zastąpieniu całki sumą policzoną dla pewnego zbioru punktów. W najprostszym przypadku:

$$\int_a^b f(x)\,dx \approx \sum_{s=0}^{S-1} f(x_i^*)\,d^*$$

gdzie $d^* = \frac{a-b}{S}$, $x_i^* = a + (i + 0.5)\,d^*$

# Model probitowy i logitowy na estymowane rangach

- charakterystyka danych
  - mamy $K$ alternatyw
  - dysponujemy uszeregowaniem alternatyw pod względem ich atrakcyjności

$$y_{k_i} > y_{k_i} > \ldots > y_{k_i}$$

  - uszeregowanie nie musi być kompletne
- Metoda szacowania: oparta na estymatorze dla modelu proporcjonalnego hazardu Cox-a

# Model probitowy z endogenicznymi zmiennymi objaśniającymi c.d.

- Model

$$
\begin{aligned}
y_{i1}^{*} &= \mathbf{y}_{2i}\boldsymbol{\beta} + \mathbf{x}_{1i}\gamma + u_{1i} = \mathbf{z}_i\delta + u_{1i} \\
y_{2i} &= \mathbf{x}_{1i}\boldsymbol{\alpha}_1 + \mathbf{x}_{2i}\boldsymbol{\alpha}_2 + u_{2i} = \mathbf{x}_i\boldsymbol{\alpha} + u_{2i}
\end{aligned}
$$

- Zmienna $\mathbf{y}_{2i}$ jest bezpośrednio obserwowalna i endogeniczna w równaniu dla zmiennej ukrytej $y_{i1}^{*}$, ponieważ $Cov\left(u_{1i}, u_{2i}\right) \neq 0$.
- Zakładamy, że wektor błędów losowych $\mathbf{u}_i \sim N\left(0, \boldsymbol{\Sigma}\right)$ przy czym przyjmujemy założenie identyfikujące parametry, że $\Sigma_{11} = 1$
- Powyższy model ma strukturę rekursywną $y_{2i}$ pojawia się w równaniu dla $y_{i1}^{*}$ ale $y_{i1}^{*}$ nie pojawia się w równaniu dla $y_{2i}$.
- Równanie dla $y_{2i}$ może być traktowane jako forma zredukowana

- Zmienna obserwowalna

$$y_{1i} = \begin{cases} 1 & \text{dla} \quad y_i^* \geq 0 \\ 0 & \text{dla} \quad y_i^* < 0 \end{cases}$$

- Funkcja prawdopodobieństwa dla zdarzenia $i$

$$\begin{aligned} f\left(y_{1i}, y_{2i} \middle| \mathbf{x}_{1i}, \mathbf{x}_{2i}\right) &= \Pr\left(y_{1i} \middle| y_{2i}, \mathbf{x}_{1i}, \mathbf{x}_{2i}\right) f\left(y_{2i} \middle| \mathbf{x}_{1i}, \mathbf{x}_{2i}\right) \\ &= \left[\Phi\left(m_i\right)\right]^{y_{1i}} \left[1 - \Phi\left(m_i\right)\right]^{1-y_{1i}} \frac{1}{\sigma} \boldsymbol{\phi} \left(\frac{y_{2i} - \mathbf{x}_{1i}\boldsymbol{\alpha}}{\sigma}\right) \end{aligned}$$

gdzie

$$m_i = \frac{\mathbf{z}_i \,\delta + \boldsymbol{\rho}\left(y_{2i} - \mathbf{x}_i \boldsymbol{\alpha}\right)}{\left(1 - \boldsymbol{\rho}^2\right)^{\frac{1}{2}}}$$

a $\boldsymbol{\rho}$ jest współczynikiem korelacji między $u_{1i}$ i $u_{2i}$

- Z użyciem tej funkcji prawdopodbieństwa możemy stworzyć funkcję wiarygodności

- Założenie o symetrii rozkładu w mówi, że

$$F(z) = 1 - F(z)$$

- Założenie to implikuje, że efekt cząstkowy jest najwyższy dla $F(\mathbf{x}_i \boldsymbol{\beta}) = \frac{1}{2}$
- Istotnie efekt cząstkowy

$$\frac{\partial \mathrm{E}(y_i | \mathbf{x}_i)}{\partial \mathbf{x}_i} = f(\mathbf{x}_i \boldsymbol{\beta}) \boldsymbol{\beta}_k$$

  a dla rozkładu symetrycznego wartość modalna jest zawsze równa medianie
- Jeśli chcemy to założenie uchylić powinniśmy użyć modelu o niesymetrycznym rozkładzie błędu loswego dla zmiennej ukrytej

# Założenie o symetrii rozkładu

## Asymetryczny logit (skewed logit)

- Model: prawdopodbieństwo sukcesu

$$\Pr\left(y_i = 1\right) = F\left(\mathbf{x}_i \boldsymbol{\beta}\right)$$

dla

$$F\left(z\right) = 1 - \frac{1}{\left[1 + \exp\left(x\right)\right]^{\boldsymbol{\alpha}}}, \ \boldsymbol{\alpha} > 0$$

- Dla $\boldsymbol{\alpha} = 1$ otrzymujemy logita
- W innych przypadkach rozkład zmiennej losowej jest niesymetryczny

# Założenie o symetrii rozkładu

## loglog model

- Model: prawdopodbieństwo sukcesu

$$\Pr\left(y_i = 1\right) = F\left(\mathbf{x}_i \boldsymbol{\beta}\right)$$

dla

$$F\left(z\right) = 1 - \exp\left[-\exp\left(z\right)\right]$$

- W tym przypadku dystrybuanta jest także niesymetryczna
- Model ten używamy gdy liczba sukcesów (bądź porażek jest mała)