Aula 02

Modelos de Escolha Discreta com Dados Desagregados

Claudio R. Lucinda

FEA-RP/USP

Agenda

1. Modelos de Escolha Discreta
   - Modelo LOGIT Multinomial
   - Elasticidades e o Problema da IIA

Modelos de Escolha Discreta

- Agora, iremos discutir os modelos em que a escolha se dá sobre o "espaço de características"; os produtos derivam utilidade apenas na medida em que eles são agregados de características.
- Esta escolha no espaço de característica possui um elemento inerentemente idiosincrático; este lado idiosincrático é o que permite a estimação dos parâmetros.
- Inicialmente começaremos analisando o processo de estimação quando o analista possui dados sobre a escolha individual dos consumidores; depois discutiremos as situações em que apenas possuimos dados agregados.
- Uma bibliografia muito boa sobre esse assunto é

Modelos de Escolha Discreta

- O analista começará postulando uma função que relaciona estes dados observados com a escolha do consumidor, que chamaremos de $V(x_{nj}, s_{nj})$, sendo que x_{nj} representa as características observadas do produto e s_{nj} as características não observadas.
- Uma vez que alguns aspectos da utilidade do consumidor não são observados, em geral $V \neq U$, em que U é a "verdadeira" utilidade do consumidor. Desta forma, podemos fazer o seguinte ajuste:

$$U_{ij} = V_{ij} + \epsilon_{ij}$$

- Em que i denota o consumidor e j a alternativa. A idéia é que o termo ϵ_{ij} capture os aspectos do produto ou do indivíduo que não são observados pelo econometrista.
- Dada esta definição, as características deste termo dependem fundamentalmente de como o mesmo especifica V_{ij}.

Modelos de Escolha Discreta II

- No entanto, para que possamos estimar os componentes de V_{ij}, precisamos do termo ϵ_{ij}, e de uma distribuição conjunta para os ϵ_{ij} de todos os j. Denominando a distribuição conjunta de $\varepsilon =< \epsilon_{i1}, \epsilon_{i2}, \cdots, \epsilon_{iN} >$, temos:

$$
\begin{aligned}
P_{ij} &= Prob(U_{ij} > U_{ik}, \forall k \neq j) \\
&= Prob(V_{ij} + \epsilon_{ij} > V_{ik} + \epsilon_{ik}, \forall k \neq j) \\
&= Prob(V_{ij} - V_{ik} > \epsilon_{ik} - \epsilon_{ij}, \forall k \neq j) \\
&= Prob(\epsilon_{ik} - \epsilon_{ij} < V_{ij} - V_{ik}, \forall k \neq j)
\end{aligned}
$$

Modelos de Escolha Discreta III

- Esta última igualdade é uma distribuição acumulada, que nos diz a probabilidade que cada um dos termos aleatórios $\epsilon_{ik} - \epsilon_{ij}$ está abaixo das diferenças entre as avaliações observadas $V_{ij} - V_{ik}$. Podemos calcular este negócio, usando a distribuição conjunta dos ε, com a seguinte integral multidimensional:

$$P_{ij} = \int_{\varepsilon} I(\epsilon_{ik} - \epsilon_{ij} < V_{ij} - V_{ik}, \forall k \neq j) f(\varepsilon) d(\varepsilon)$$

- Em português, esta integral nos dá a área da distribuição conjunta de ε tal que as diferenças nos componentes idiosincráticos sejam menores do que as diferenças nos componentes determinísticos.
- Diferentes especificações de modelos de escolha discreta surgem em resposta a diferentes especificações da variável aleatória multidimensional ε. Por exemplo, se ε for uma distribuição $N(0, \Omega)$, isso nos dá o modelo probit multinomial.

Modelos LOGIT Multinomial:

- Se ε seguir uma distribuição de valores extremos I:

$$
\begin{aligned}
f(\epsilon_{ij}) &= e^{-\epsilon_{ij}} e^{-e^{-\epsilon_{ij}}} \\
F(\epsilon_{ij}) &= e^{e^{-\epsilon_{ij}}}
\end{aligned}
$$

- Temos o modelo LOGIT Multinomial. É importante notar que, para o caso dos modelos LOGIT, a integral multidimensional que fizemos anteriormente pode ser resolvida analiticamente.
- O primeiro passo para entendermos isso é uma regrinha que diz que as diferenças entre duas variáveis aleatórias que seguem esta distribuição valores extremos I têm distribuição logística:

$$
\begin{aligned}
\epsilon^*_{ij} &= \epsilon_{ik} - \epsilon_{ij} \\
F(\epsilon^*_{ij}) &= \frac{\epsilon^*_{ij}}{1 + \epsilon^*_{ij}}
\end{aligned}
$$

Modelo LOGIT Multinomial:

- A segunda parada é que os componentes idiosincráticos das utilidades são i.i.d.; mas antes, vamos reescrever a última das probabilidades antes da integral da seguinte forma:

$$P_{ij} = Prob(\epsilon_{ik} < \epsilon_{ij} + V_{ij} - V_{ik}, \forall k \neq j)$$

- Se o ϵ_{ij} é considerado como dado, esta função nos dá a função de distribuição acumulada para cada ϵ_{ik} avaliada em $\epsilon_{ij} + V_{ij} - V_{ik}$, o que, de acordo com a distribuição valores extremos I é igual a $\exp[-\exp[-(\epsilon_{ij} + V_{ij} - V_{ik})]]$. Como os elementos do vetor ε são independentes, isto significa que esta probabilidade conjunta – afinal de contas, vale para todos os elementos de ε exceto j – é igual a um produto das distribuições individuais:

$$P_{ij}|\epsilon_{ij} = \prod_{k \neq j} e^{-e^{-(\epsilon_{ij}+V_{ij}-V_{ik})}}$$

Modelo LOGIT Multinomial (II):

- Evidentemente, ϵ_{ij} não é dado, desta forma a probabilidade conjunta é a integral desta parada com respeito a todos os valores de ϵ_{ij}:

$$P_{ij} = \int_{\epsilon_{ij}=-\infty}^{\infty} \left(\prod_{k \neq j} e^{-e^{-(\epsilon_{ij}+V_{ij}-V_{ik})}} \right) e^{-\epsilon_{ij}} e^{-e^{-\epsilon_{ij}}} d\epsilon_{ij}$$

- Vamos cozinhar um pouco esta equação; lembrando que, para o produto j , $V_{ij} - V_{ij} = 0$, temos que a integral acima pode ser reconstruída da seguinte forma:

$$P_{ij} = \int_{\epsilon_{ij}=-\infty}^{\infty} \left(\prod_{k} e^{-e^{-(\epsilon_{ij}+V_{ij}-V_{ik})}} \right) e^{-\epsilon_{ij}} d\epsilon_{ij}$$

Modelo LOGIT Multinomial (III):

- Podemos transformar este produtório em soma, uma vez que as bases são iguais:

$$\begin{aligned} P_{ij} &= \int_{\epsilon_{ij}=-\infty}^{\infty} \exp\left(-\sum_k e^{-(\epsilon_{ij}+V_{ij}-V_{ik})}\right) e^{-\epsilon_{ij}} d\epsilon_{ij} \\ &= \int_{\epsilon_{ij}=-\infty}^{\infty} \exp\left(-e^{-\epsilon_{ij}} \sum_k e^{-(V_{ij}-V_{ik})}\right) e^{-\epsilon_{ij}} d\epsilon_{ij} \end{aligned}$$

- Redefinindo as variáveis de integração, tal que $e^{-\epsilon_{ij}} = t$, tal que $dt = -e^{-\epsilon_{ij}} d\epsilon_{ij}$. Note que, quando $\epsilon_{ij} \rightarrow \infty$, $t \rightarrow 0$, e quando $\epsilon_{ij} \rightarrow -\infty$, $t \rightarrow -\infty$, o que faz com que os limites de integração agora sejam 0 e ∞.

Modelo LOGIT Multinomial

- Usando este novo termo:

$$\begin{aligned}
P_{ij} &= \int_{t=\infty}^{0} \exp\left(-t\sum_k e^{-(V_{ij}-V_{ik})}\right)(-dt) \\
&= \int_{t=0}^{\infty} \exp\left(-t\sum_k e^{-(V_{ij}-V_{ik})}\right) dt \\
&= \left.\frac{\exp\left(-t\sum_k e^{-(V_{ij}-V_{ik})}\right)}{\sum_k e^{-(V_{ij}-V_{ik})}}\right|_0^{\infty} \\
&= \frac{1}{\sum_k e^{-(V_{ij}-V_{ik})}} = \frac{e^{V_{ij}}}{\sum_k e^{V_{ik}}}
\end{aligned}$$

- Podemos resumir os cuidados que temos na estimação dos modelos de escolha discreta em duas afirmações, "apenas diferenças de utilidade são importantes" e "a escala da utilidade é arbitrária".

A Escala da Utilidade é Arbitrária

- Se somarmos uma constante a cada um dos termos V_{ik}, a fórmula da probabilidade do slide anterior não se altera.
- Isso implica que os únicos parâmetros que podem ser estimados são aqueles que capturam diferenças entre as alternativas.
- Como fazer com variáveis que são constantes entre as alternativas:
  - Assumir diferentes coeficientes para cada alternativa
- Como só as diferenças de utilidade importam, na verdade o modelo de utilidade aleatória é expresso em termo de $J-1$ diferenças.

Apenas diferenças de utilidade são importantes

- Podemos notar que se multiplicarmos todos os termos V_{ik} por uma constante, a fórmula do slide anterior não se altera.
- Para lidar com isso, você precisa normalizar a escala dos termos erro, usualmente normalizando a variância dos ε
- No caso do Logit, a variância é $\frac{\pi^2}{6}$, ou aproximadamente 1.6. No Probit, a variância é 1.
- Por isso tem que tomar o cuidado em comparar os coeficientes do Probit e do Logit (os do logit são mais ou menos $\sqrt{1.6}$ o do Probit).
- Os coeficientes são $\frac{\beta}{\sigma}$, com σ sendo o DP do ε.

Estimação dos Modelos de Escolha Discreta:

- Em geral, os procedimentos de estimação do modelo Logit Multinomial está baseado no princípio da Máxima Verossimilhança. Inicialmente, vamos supor que a amostra seja aleatória, e que tenhamos dados sobre N tomadores de decisão. A probabilidade de um indivíduo i escolher a alternativa que ele efetivamente escolheu é igual a:

$$\prod_{j \in J} (P_{ij})^{y_{ij}}$$

- Em que $y_{ij} = 1$ se o indivíduo i escolheu o produto e $y_{ij} = 0$, caso contrário.

Estimação dos Modelos de Escolha Discreta (II):

- Supondo independência das escolhas dos indivíduos, a probabilidade de observação de uma amostra igual à que temos é:

$$L(\beta) = \prod_{i \in N} \prod_{j \in J} (P_{ij})^{y_{ij}}$$

- Em geral, os algoritmos numéricos maximizam o logaritmo desta probabilidade conjunta, o que dá:

$$\ln(L(\beta)) = LL(\beta) = \sum_{i \in N} \sum_{j \in J} y_{ij} \ln P_{ij}$$

Estimação dos Modelos de Escolha Discreta (III):

- Em geral, também podemos dar uma interpretação de GMM ao método de estimação utilizado da seguinte forma. O vetor de parâmetros que minimiza esta função deve atender à seguinte condição de primeira ordem:

$$\frac{\partial LL(\beta)}{\partial \beta} = 0$$

- Para facilitar, vamos supor que $V_{ij} = x_{ij}\beta$. Neste caso, temos:

$$\sum_{i \in N} \sum_{j \in J} (y_{ij} - P_{ij}) x_{ij} = 0$$

Efeitos Marginais

- Efeito Marginal sobre a Probabilidade de escolha do Produto j de uma alteração no atributo r do produto j (efeito marginal próprio):

$$\begin{aligned}\frac{\partial P_j}{\partial x_j^r} &= \frac{\partial(e^{V_{ij}} / \sum_{k \in J} e^{V_{ik}})}{\partial x_j^r} \\ &= \frac{\partial V_j}{\partial x_j^r} P_j(1 - P_j)\end{aligned}$$

- Efeito Marginal sobre a Probabilidade de escolha do Produto j de uma alteração no atributo r de um produto $k \neq j$ (efeito marginal Cruzado):

$$\begin{aligned}\frac{\partial P_j}{\partial x_k^r} &= \frac{\partial(e^{V_j} / \sum_{c \in J} e^{V_{ic}})}{\partial x_k^r} \\ &= - \frac{\partial V_{ik}}{\partial x_k^r} P_j P_k\end{aligned}$$

Elasticidades

- Elasticidade Própria:

$$\epsilon = \frac{\partial P_j}{\partial x_j^r} \times \frac{x_j^r}{P_j} = \frac{\partial V_j}{\partial x_j^r}(1 - P_j)x_j^r$$

- Elasticidade Cruzada:

$$\epsilon_{ikr} = \frac{\partial P_j}{\partial x_k^r} \times \frac{x_k^r}{P_j} = -\frac{\partial V_{ik}}{\partial x_k^r} P_k x_k^r$$

- Esse último termo só depende de uma derivada parcial e de coisas relacionadas a k – e não a j