Aula 02

Bibliografia: Apresentação

Cláudio R. Lucinda

FEA-RP/USP

Objetivos da Aula

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.

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:

$$f(\epsilon_{ij}) = e^{-\epsilon_{ij}} e^{-e^{-\epsilon_{ij}}}$$
$$F(\epsilon_{ij}) = e^{e^{-\epsilon_{ij}}}$$

- 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:

$$\epsilon_{ij}^{*} = \epsilon_{ik} - \epsilon_{ij}$$
$$F(\epsilon_{ij}^{*}) = \frac{\epsilon_{ij}^{*}}{1 + \epsilon_{ij}^{*}}$$

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 $\epsilon_{ij} \to \infty$, $t \to 0$, e quando $\epsilon_{ij} \to -\infty$, $t \to -\infty$, 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 dias afirmações. A primeira é que "apenas diferenças de utilidade são importantes", e a segunda é que "a escala da utilidade é arbitrária".

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 dias afirmações. A primeira é que "apenas diferenças de utilidade são importantes", e a segunda é que "a escala da utilidade é arbitrária".

1. Outro ponto importante é que variáveis socio-demográficas, que afetam uniformemente as alternativas, não podem ser incluídas no lado determinista da função utilidade em todas as alternativas. Pelo menos um dos coeficientes associado com uma das alternativas precisa ser normalizado em zero. Isso não ocorre quando temos esta variável entrando como interação com as características específicas das alternativas.

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 (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:

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

1. McFadden (1974) provou que, para V_{ij} linear nos parâmetros, esta função objetivo é côncava, o que garante que os algoritmos numéricos tradicionais convirjam rapidamente para um mínimo.

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:

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

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:

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

1. O que vai dentro de um parênteses seria análogo ao resíduo da regressão – a diferença entre o valor observado (que são as escolhas e estão caputradas na variável y_{ij}) e o valor previsto, que são as probabilidades de escolha. O x_{ij} que o está multiplicando seriam os regressores do modelo. Ou seja, podemos pensar que estimativas por Máxima Verossimilhança são também estimativas de GMM, só que considerando as derivadas – também conhecidas como *scores* – como condições de momento.