Conditional Logit

Anders Munk-Nielsen

October, 2021

The Conditional Logit (CLOGIT) Model

The Conditional Logit model is a powerful tool for analyzing descrete choices, which is heavily used in e.g. empirical industrial organization. In this model, decision makers have to choose between J different discrete options, $\{1, 2, ..., J\}$. An option might be a car. For each option, we observe a vector of characteristics, $x_j \in \mathbb{R}^K$, and for each individual $i = 1, ..., N$, we observe the chosen alternative, $y_i \in \{1, ..., J\}$. If individuals face different alternatives, e.g. if the prices or characteristics of cars available were different, then characteristics also vary across individuals, $x_{ij} \in \mathbb{R}^K$.

Our model assumes that individual i chose the alternative that maximized utility,

$$y_i = \arg\max_{j \in \{1,...,J\}} u_{ij}.$$

Our model for utility takes the form

$$u_{ij} = x_{ij}\beta_o + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \text{IID Extreme Value Type I}.$$

That is, utility is composed of a part that depends on observables, $x_{ij}\beta$, and an idiosyncratic error term, ε_{ij}, observed to the individual but not to the econometrician. The problem of estimation is to recover β without knowledge on ε_{ij}.

It turns out that the distributional form for ε_{ij} implies that

$$\Pr(y_i = j|\mathbf{X}_i) = \frac{\exp(x_{ij}\beta_o)}{\sum_{k=1}^{J} \exp(x_{ik}\beta_o)},$$

where $\mathbf{X}_i = (x_{i1}, x_{i2}, ..., x_{iJ})$. This remarkable result was first introduced to economists by nobel laureate Daniel McFadden. Taking logarithms, we obtain a particularly parsimonious form for the log-likelihood contribution for generic $\boldsymbol{\beta}$:

$$\ell_i(\beta) = x_{ij}\beta - \log \sum_{k=1}^{J} \exp(x_{ik}\beta).$$

Max rescaling for numerical stability

A particularly important numerical trick that one is typically forced to use with logit models is what is called *max rescaling*. For simplicity, let $v_{ij} \equiv x'_{ij}\beta$. Now, note that for any $K_i \in \mathbb{R}$,

$$\begin{aligned}\frac{\exp(v_{ij})}{\sum_{k=1}^{J}\exp(v_{ik})} &= \frac{\exp(v_{ij})}{\sum_{k=1}^{J}\exp(v_{ik})}\frac{\exp(-K_i)}{\exp(-K_i)}\\ &= \frac{\exp(v_{ij}-K_i)}{\sum_{k=1}^{J}\exp(v_{ik}-K_i)}.\end{aligned}$$

This means that we can subtract any scalar from all utilities for an individual. This is very useful because the exponential function is a highly unstable numerical object on any computer: for large or small values of z, $\exp(z)$ will result in round-up or round-down errors, respectively. Since round-up errors are particularly bad for estimation, it turns out to be useful to choose K_i so that we avoid them, even at the cost of encountering more round-down errors. Thus, we choose

$$K_i = \max_{j\in\{1,\ldots,J\}} v_{ij},$$

and subtract K_i from all utilities before taking any exponential values.

Price elasticity

A key feature of interest in most applied work using the conditional logit model is the price elasticity of demand. A surprisingly fun and energizing exercise in calculus is to compute the derivative of the logit choice probabilities with respect to one of the regressors, $x_{ij\ell}$. For convenience, let us write the choice probability as $s_{ij} \equiv \frac{\exp(\mathbf{x}_{ij}\boldsymbol{\beta})}{\sum_{k=1}^{J}\exp(\mathbf{x}_{ik}\boldsymbol{\beta})}$. Then the derivative is

$$\nabla s_{ij} = s_{ij}\left(\nabla v_{ij} - \sum_{k=1}^{J} s_{jk}\nabla v_{ik}\right).$$

So since we have a linear model of utility, $v_{ij} = \mathbf{x}_{ij}\boldsymbol{\beta}$, the "inner" derivatives, ∇v_{ij} take particularly simple forms. For example, $\frac{\partial v_{ij}}{\partial x_{ik\ell}} = \mathbf{1}(k=j)\beta_\ell$, and $\frac{\partial v_{ij}}{\partial \theta_k} = x_{ijk}$.

Suppose the log of the price, $\log p_{ij}$, is one of variables in $\mathbf{x}_{ij}$. Then the derivative is

$$\frac{\partial s_{ij}}{\partial \log p_{ik}} = \begin{cases} s_{ij}(1-s_{ij})\beta_\ell & \text{if } j=k,\\ -s_{ij}s_{ik}\beta_\ell & \text{if } j\neq k,\end{cases}$$

with the notation that the ℓth coefficient in $\boldsymbol{\beta}$ is the coefficietn on the log price. Finally, if we are interested in the elasticity of the jth market share wrt. the kth price, $\mathcal{E}_{jk}$, we can use that

$$\mathcal{E}_{jk} \equiv \frac{\partial s_{ij}}{\partial p_{ik}}\frac{p_{ik}}{s_{ij}} = \frac{\partial s_{ij}}{\partial \log p_{ik}}\frac{1}{s_{ij}}.$$

Compensating variation in logit models

Note: This is not needed for the exercise, but is very relevant for anyone using logit models.

A useful feature of logit models is that they provide a neat welfare measure in the form of what is commonly referred to as the "log sum." This is because of the fact that

$$\mathbb{E}_{\varepsilon_{i1},\ldots,\varepsilon_{iJ}}\left[\max(v_{ij}+\varepsilon_{ij})\right] = \log\left[\sum_{j=1}^{J}\exp(v_{ij})\right].$$

Because the left-hand side is the expected utility, prior to knowing the error terms, of the choice instance, it can be thought of the "value" of the choice instance. If one of the variables, say the first, is a price variable, then β_1 is the marginal utility of price, converting money into utils. Thus, economists tend to divide the welfare measure with β_1 to get a money-metric utility measure. Again, the level of that measure in itself may not be useful, but differences are. Suppose we change something about the utilities, from v_{ij} to $\tilde{v}_{ij}$, and want to compute the compensating variation: that is, how much we would have to pay the agent (regardless of the chosen alternative) to make the agent indifferent between being placed in the first or the second choice instance. We can compute that as

$$CV = \frac{1}{\beta_1}\log\left[\sum_{j=1}^{J}\exp(\tilde{v}_{ij})\right] - \frac{1}{\beta_1}\log\left[\sum_{j=1}^{J}\exp(v_{ij})\right].$$

If we add the monetary amount, CV, to all utilities in the baseline, v_{ij}, then the expected maximum utility would be the same in the two choice instances, and the agent would thus be indifferent between the two.

Policy makers can use the CV measure to compare how much better or worse an individual is from a change in taxation, an introduction or reduction in the choiceset, or a change in one or more of the attributes of the alternatives.