Mixed Logit
or
Random Parameter Logit Model

Mixed Logit Model

- Very flexible model that can approximate any random utility model.
- This model when compared to standard logit model overcomes the
  - Taste variation issues and
  - Does not exhibit IIA property
- Mixed logit probabilities are the integrals of standard logit probabilities over a density of parameters

Functional Form

a mixed logit model is any model whose choice probabilities can be expressed in the form

$$Pni = \int L_{ni}(\beta)\, f(\beta)\, d\beta$$

where $L_{ni}(\beta)$ is the logit probability evaluated at parameters β:

$$L_{ni}(\beta) = \frac{e^{V_{ni}(\beta)}}{\sum_{j=1}^{J} e^{V_{nj}(\beta)}}$$

and $f(\beta)$ is a density function.

$V_{ni}(\beta)$ is the observed portion of the utility, which depends on the parameters β. If utility is linear in β, then

$V_{ni}(\beta) = \beta` x_{ni}$.

In this case, the mixed logit probability takes its usual form:

$$P_{ni} = \int \left(\frac{e^{\beta' x_{ni}}}{\sum_j e^{\beta' x_{nj}}} \right) f(\beta)\, d\beta$$

The mixed logit probability is a weighted average of the logit formula evaluated at different values of β, with the weights given by the density $f(\beta)$.

- the weighted average of several functions is called a mixed function, and the density that provides the weights is called the mixing distribution.
- Mixed logit is a mixture of the logit function evaluated at different β's with f (β) as the mixing distribution.
- Standard logit is a special case where the mixing distribution f (β) is degenerate at fixed parameters b:

  f (β) = 1 for β = b and 0 for β ≠ b.

  The choice probability then becomes the simple logit formula

$$P_{ni} = \frac{e^{b' x_{ni}}}{\sum_j e^{b' x_{nj}}}.$$

- The mixing distribution f (β) can be discrete, with β taking a finite set of distinct values. Suppose β takes M possible values labeled $b_1, \ldots, b_M$, with probability s_m that $\beta = b_m$.
- In this case the choice probability is

$$P_{ni} = \sum_{m=1}^{M} s_m \left(\frac{e^{b_m' x_{ni}}}{\sum_j e^{b_m' x_{nj}}} \right)$$

- The above can be interpreted as

  there are M segments in the population, the share of the population in segment m is s_m, which the researcher can estimate within the model along with the b's for each segment.

Parameter Distributions

- In mixed logit, $f(\beta)$ is generally specified to be continuous.
- Normal, lognormal, uniform, triangular, gamma, or any other distribution can be used as a density function for β
- By denoting the parameters that describe the density of β as θ, the more appropriate way to denote this density is $f(\beta \mid \theta)$.
- The mixed logit choice probabilities do not depend on the values of β. These probabilities are $P_{ni} = \int L_{ni}(\beta) f(\beta \mid \theta)\, d\beta$, which are functions of θ.
- The parameters β are integrated out. Thus, the β's are similar to the ε_{nj}s

- f (β) can be specified to be normal or lognormal: β ~ N(b,W) or ln β ~ N(b,W) with parameters b and W that are estimated.
- The lognormal distribution is useful when the coefficient is known to have the same sign for every decision maker, such as cost and time coefficient that are known to be negative for everyone in a mode choice situation.
- Quite a few researchers have used triangular and uniform distributions.
- With the uniform density, β is distributed uniformly between b – s and b + s, where the mean b and spread s are estimated.

- The triangular distribution has positive density that starts at b – s, rises linearly to b, and then drops linearly to b + s, taking the form of a triangle.
- The mean b and spread s are estimated, as with the uniform, but the density is peaked instead of flat.
- These densities have the advantage of being bounded on both sides, thereby avoiding the problem that can arise with normals and lognormals having unreasonably large coefficients for some share of decision makers.
- By constraining s = b, one can assure that the coefficients have the same sign for all decision makers.

Estimation by Simulation

- Mixed logit is well suited to simulation methods for estimation. Utility is $U_{nj} = \beta`_n x_{nj} + \varepsilon_{nj}$, where the coefficients β_n are distributed with density $f(\beta \mid \theta)$, where θ refers collectively to the parameters of this distribution.
- The researcher specifies the functional form $f(\cdot)$ and wants to estimate the parameters θ.
- The choice probabilities are

$$P_{ni} = \int L_{ni}(\beta) f(\beta \mid \theta)\, d\beta,$$

where

$$L_{ni}(\beta) = \frac{e^{\beta' x_{ni}}}{\sum_{j=1}^{J} e^{\beta' x_{nj}}}.$$

- The probabilities are approximated through simulation for any given value of θ:
  - (1) Draw a value of β from f (β | θ), and label it β^r with the superscript r = 1 referring to the first draw.
  - (2) Calculate the logit formula $L_{ni}(\beta^r)$ with this draw.
  - (3) Repeat steps 1 and 2 many times, and average the results.
  - This average is the simulated probability:

$$\check{P}_{ni} = \frac{1}{R} \sum_{r=1}^{R} L_{ni}(\beta^r),$$

  - where R is the number of draws.

- The simulated probabilities are inserted into the log-likelihood function to give a simulated log likelihood:

$$\mathrm{SLL} = \sum_{n=1}^{N} \sum_{j=1}^{J} d_{nj} \ln \check{P}_{nj},$$

- where $d_{nj} = 1$ if n chose j and zero otherwise. The maximum simulated likelihood estimator (MSLE) is the value of θ that maximizes SLL.