4 The Mixed Logit Model

The mixed logit model is a highly flexible generalisation of the standard logit from last lecture. It resolves the three limitations of standard logit by allowing for random taste variation, unrestricted substitution patterns and correlation in unobserved factors over time.

4.1 Mixed Logit Setup

Mixed logit probabilities are the integral of standard logit probabilities over a density of parameters, i.e. it can be expressed in the form:

$$P_{ij} = \int L_{ij}(\omega) g(\omega) d\omega \tag{4.1}$$

where L_{ij} is the logit probability evaluated at parameters ω:

$$L_{ij}(\omega) = \frac{e^{V_{ij}(\omega)}}{\sum_{k=1}^{J} e^{V_{ik}(\omega)}}$$

and $g(\omega)$ is a density function. As before $V_{ij}(\omega)$ is a portion of utility which depends on the parameters ω ($U_{ij} = V_{ij} + \varepsilon_{ij}$). If utility is linear in ω, then $V_{i}j(\omega) = \omega' x_{ij}$ and the mixed logit probability takes the form:

$$P_{ij} = \int \frac{e^{\omega' x_{ij}}}{\sum_{k=1}^{J} e^{\omega' x_{ik}}} g(\omega) d\omega$$

> **Note:-**
>
> In the statistics literature, a *mixed function* is defined as the weighted average of several functions, and the density providing the weights is known as the mixing distribution. Mixed logit is a mixture of the logit function evaluated at different ω's, and the mixing distribution is $g(\omega)$.

4.1.1 Special Cases

Standard logit is a special case, where the mixing distribution $g(\omega)$ is degenerate at fixed parameters w: $g(\omega) = 1$ for $\omega = w$ and $g(\omega) = 0$ otherwise. I.e. it is a point mass at $\omega = w$. The choice probabilities are then given by the standard logit formula:

$$P_{ij} = \frac{e^{w' x_{ij}}}{\sum_{k=1}^{J} e^{w' x_{ik}}}$$

Discrete mixing distributions are also possible, with ω taking on a finite number of values. Suppose ω can take on M different values $w_1, w_2, \ldots, w_M$ with $Pr(\omega = w_m) = s_m$. This is known as the *latent class model* and the choice probabilities are given by:

$$P_{ij} = \sum_{m=1}^{M} s_m \frac{e^{w'_m x_{ij}}}{\sum_{k=1}^{J} e^{w'_m x_{ik}}}$$

This specification is useful if there are M segments in the population, each with its own choice behaviour. The share of each population in segment m is given by s_m.

4.1.2 Estimation

The researcher specifies a distribution for ω (i.e. a distribution for the taste parameters) and aims to estimate the parameters of this distribution[1]. For example if $\omega \sim N(\bar{\omega}, \Sigma_\omega)$, then the researcher estimates $\bar{\omega}$ and Σ_ω. Denote the parameters of the mixing distribution as θ, we now have choice probabilities $P_{ij} = \int L_{ij}(\omega) g(\omega|\theta) d\omega$ which is a function of θ, the parameters ω are integrated out. In this sense, the ω's are similar to the ε_{ij}'s in that both are random terms we integrate out to obtain the choice probabilities.

Note that the dimension of ω is $L \times 1$ with L attributes, maximisation requires the estimation of an L dimensional integral. This is typically done using simulation methods.

4.2 Elasticities

Mixed logit does not exhibit IIA or the restrictive substitution patterns of logit. The ratio of mixed logit probabilities depends on all the data, including attributes other than j and k:

$$\frac{P_{ij}}{P_{ik}} = \frac{\int \frac{e^{V_{ij}}}{\sum_{k=1}^{J} e^{V_{ik}}} g(\omega) d\omega}{\int \frac{e^{V_{ik}}}{\sum_{k=1}^{J} e^{V_{ik}}} g(\omega) d\omega}$$

The denominators of the logit formula are now inside the integrals, and therefore do not cancel as in standard logit. We can further compute elasticities, denote by $E_{jx_h^\ell}$ the elasticity of P_{ij} with respect to attribute h of alternative ℓ:

$$E_{jx_h^\ell} = \frac{\partial P_{ij}}{\partial x_h^\ell} \frac{x_h^\ell}{P_{ij}}$$

$$= \frac{\partial}{\partial x_h^\ell} \left(\int \frac{e^{V_{ij}}}{\sum_{k=1}^{J} e^{V_{ik}}} g(\omega) d\omega \right) \frac{x_h^\ell}{P_{ij}}$$

$$= \int \frac{\partial}{\partial x_h^\ell} \left(\frac{e^{V_{ij}}}{\sum_{k=1}^{J} e^{V_{ik}}} \right) g(\omega) d\omega \frac{x_h^\ell}{P_{ij}}$$

$$\text{when } j \neq h: = \int -\frac{e^{V_{ij}} \frac{\partial V_{ih} e^{V_{ih}}}{\partial x_h^\ell}}{\left(\sum_{k=1}^{J} e^{V_{ik}} \right)^2} g(\omega) d\omega \frac{x_h^\ell}{P_{ij}}$$

$$= -\frac{x_h^\ell}{P_{ij}} \int \frac{\partial V_{ih}}{\partial x_h^\ell} L_{ij}(\omega) L_{ih}(\omega) g(\omega) d\omega$$

$$\text{when } j = h: = \int \frac{\left(\sum_{k=1}^{J} e^{V_{ik}} \right) \frac{\partial V_{ij}}{\partial x_h^\ell} e^{V_{ij}} - e^{V_{ij}} \frac{\partial V_{ij}}{\partial x_h^\ell} e^{V_{ij}}}{\left(\sum_{k=1}^{J} e^{V_{ik}} \right)^2} g(\omega) d\omega \frac{x_h^\ell}{P_{ij}}$$

$$= \frac{x_h^\ell}{P_{ij}} \int \frac{\partial V_{ij}}{\partial x_h^\ell} L_{ij}(\omega)(1 - L_{ij}(\omega)) g(\omega) d\omega$$

When $V_{ij} = \omega' x_{ij}$, $\frac{\partial V_{ij}}{\partial x_h^\ell} = \omega_h$ in the above. This elasticity is different for each alternative j. A 10% reduction for one alternative need not imply (as with logit) a 10% reduction for all other alternatives. Note that the percent change in probability depends on the correlation between $L_{ij}(\omega)$ and $L_{ih}(\omega)$, which is determined by the researchers specification of $g(\omega)$. For example, to represent a situation where an improvement in alternative j draws more proportionately from alternative i than alternative k, the researcher can specify an element of x that is positively correlated between i and j but uncorrelated or negatively correlated between k and j, with a mixing distribution that allows the coefficient of this variable to vary.

---

[1]There are 2 sets of parameters in the mixed logit model: the ω's which enter the logit formula and the parameters of the mixing distribution $g(\omega)$. Typically we care about estimating the latter.

4.3 Random coefficients

The mixed logit probability can be derived from utility-maximising behaviour in several ways that are are formally equivalent, but provide different interpretations. One interpretation is based on random coefficients, the decision-maker faces a choice among J alternatives and has a utility function of the form:

$$U_{ij} = \omega_i' x_{ij} + \varepsilon_{ij}$$

where ω_i now varies across decision-makers in the population with density $g(\omega)$. This density is a function of parameters θ that represent, for example, the mean and variance of the ω's in population. This specification is the same as for standard logit, except that ω is no longer fixed but varies across decision-makers. Of course the decision maker knows the value of their ω_i and ε_{ij}'s for all j, but the researcher only observes the x_{ij}'s. If the decision maker observed the ω_i's, then the model would be a standard logit. Thus the choice probability conditional on ω_i is the standard logit probability:

$$L_i j(\omega_i) = \frac{e^{\omega_i' x_{ij}}}{\sum_{k=1}^{J} e^{\omega_i' x_{ik}}}$$

However the researcher does not observe ω_i so cannot condition on it. To find the unconditional choice probability we integrate over the density of ω_i:

$$P_{ij} = \int L_{ij}(\omega) g(\omega) d\omega$$

which is exactly the mixed logit probability from equation 4.1.

Note:-

The researcher imposes $g(\omega)$ as before, and estimates its parameters θ. Typically we specify $g(\omega)$ as normal or lognormal: $\omega \sim N(\bar{\omega}, \Sigma_\omega)$ or $\ln \omega \sim N(\bar{\omega}, \Sigma_\omega)$. The researcher then estimates $\bar{\omega}$ and Σ_ω. The lognormal distribution is useful when the coefficient is known to have the same sign for every decision-maker, such as price that is known to be negative for everyone.

4.4 Error-components

Another interpretation of mixed logit is based on error-components that create correlations among the utilities for different alternatives. Utility is specified as:

$$U_i j = \omega' x_{ij} + \mu_i' z_{ij} + \varepsilon_{ij}$$

where x_{ij} and z_{ij} are vectors of observed variables relating to alternative j, ω represents fixed coefficients and μ_i represents a vector of random terms with zero mean. That is, the unobserved (random) portion of utility is $\eta_{ij} = \mu_i' z_{ij} + \varepsilon_{ij}$ which can be correlated across alternatives depending on the specification of z_{ij} [2].

With non-zero error-components, utility is correlated over alternatives:

$$Cov(\eta_{ij}, \eta_{ik}) = \mathbb{E}(\mu_i' z_{ij} + \varepsilon_{ij})(\mu_i' z_{ik} + \varepsilon_{ik}) = z_{ij}' \Sigma_\mu z_{ik}$$

where Σ_μ is the covariance of μ_i. Utility is correlated over alternatives even when the error components are independent such that W is diagonal.

Example. An analog to nested logit is obtained by specifying a dummy for each nest that equals for each alternative in the nest, and zero otherwise. With K non-overlapping nests, the error components are $\mu_i' z_{ij} = \sum_{k=1}^{K} \mu_{ik} d_{kj}$ where $d_{kj} = -1$ if alternative j is in nest k. For convenience let $\mu_{ik} \sim NID(0, \sigma_k^2)$. The random term μ_{ik} enters the utility of each nest,

---

[2]In the standard logit model $z_{ij} = 0$ for all i and j, such that there is no correlation in utility across alternatives. This lack of correlation is what results in IIA.

inducing correlation among these alternatives. It does not enter any od the alternatives in other nests, so there is no correlation across nests. The variance σ_k^2 captures the magnitude of the correlation within each nest.

The correlation within nests is given by:

$$
\begin{aligned}
Corr(\eta_{ij}, \eta_{ik}) &= \frac{Cov(\eta_{ij}, \eta_{ik})}{\sqrt{Var(\eta_{ij})Var(\eta_{ik})}} \\
&= \frac{Cov(\mu_k + \varepsilon_{ij}, \mu_k + \varepsilon_{ik})}{\sqrt{Var(\mu_k + \varepsilon_{ij})Var(\mu_k + \varepsilon_{ik})}} \\
&= \frac{Var(\mu_k)}{Var(\mu_k) + \sigma_\varepsilon^2} \\
&= \frac{\sigma_k^2}{\sigma_k^2 + \frac{\pi^2}{6}}
\end{aligned}
$$

Comparing error-components and random coefficients

Both specifications are formally equivalent, but provide different interpretations.

Random coefficients $\to$ Error-components: Random-coefficients can be written as $U_{ij} = \bar{\omega}'x_{ij} + \tilde{\omega}_i'x_{ij} + \varepsilon_{ij}$ where $\bar{\omega}$ is the mean of ω_i and $\tilde{\omega}_i$ the deviation. We can see this has error components defined by $z_{ij} = x_{ij}$ with $\mu_i = \tilde{\omega}_i$.

Error-components $\to$ Random coefficients: Error-components is denoted by $U_{ij} = \omega'x_{ij} + \mu_i'z_{ij} + \varepsilon_{ij}$, which is equivalent to random coefficients with fixed coefficients for variables x_{ij} and random coefficients with zero means for variables z_{ij}.

4.5 Variability in population

How can we learn about what proportion of the population have positive preferences for a particular attribute? I.e. given estimates of $\bar{\omega}$ and Σ_ω, how can we compute $Pr(\omega_j > 0)$?

Suppose $\omega \sim N(\bar{\omega}, \Sigma_\omega)$, then $\omega_j \sim N(\mathbb{E}[\omega_j], Var(\omega_j))$. We can replace $\mathbb{E}[\omega_j]$ and $Var(\omega_j)$ with their estimates $\hat{\omega}_j$ and $\hat{\sigma}_j^2$, then compute $Pr(\omega_j > 0)$ using the standard normal distribution:

$$
z_j = \frac{\omega_j - \mathbb{E}[\omega_j]}{\sqrt{Var(\omega_j)}} \sim N(0,1)
$$

$$
\begin{aligned}
Pr(\omega_j > 0) &= Pr\left(\frac{\omega_j - \mathbb{E}[\omega_j]}{\sqrt{Var(\omega_j)}} > -\frac{\mathbb{E}[\omega_j]}{\sqrt{Var(\omega_j)}}\right) \\
&= Pr\left(z_j > -\frac{\mathbb{E}[\omega_j]}{\sqrt{Var(\omega_j)}}\right) \\
&= Pr\left(z_j > -\frac{\hat{\omega}_j}{\hat{\sigma}_j}\right) \\
&= \Phi\left(\frac{\hat{\omega}_j}{\hat{\sigma}_j}\right)
\end{aligned}
$$