

3 Multinomial Response Models

3.1 The Random Utility Model

A decision maker i faces a finite choice set Ω_J of dimension J . The decision maker would obtain a certain level of utility from each alternative. The utility that decision maker i derives from alternative j is U_{ij} , which is known to them but **unknown to the researcher**. Thus alternative j is chosen iff $U_{ij} > U_{ik}$ for all $k \neq j$.

The researcher doesn't observe the decision-maker's utility, only some attributes of the alternatives \mathbf{v}_j (sometimes \mathbf{v}_{ij}) $\forall j$, and some attributes of the decision maker \mathbf{x}_i . The function is denoted $V_{ij} = V(\mathbf{v}_j, \mathbf{x}_i, \theta)$ where θ is a vector of unknown parameters. Since there are aspects of utility that are not observed, $V_{ij} \neq U_{ij}$, and we decompose utility as $U_{ij} = V_{ij} + \varepsilon_{ij}$, where ε_{ij} captures the unobserved aspects of utility. For example:

- $U_{ij} = \mathbf{v}'_j \omega + \varepsilon_{ij}$, where \mathbf{v}_j is an $L \times 1$ vector of attributes for alternative j , and ω is an $L \times 1$ vector of parameters.
- $U_{ij} = \mathbf{x}'_i \beta + \alpha_j + \varepsilon_{ij}$, where \mathbf{x}_i is an $K \times 1$ vector of attributes for decision maker i , β is a $K \times 1$ vector of parameters and α_j is a scalar alternative-specific constant (capturing average utility of alternative j).
- $U_{ij} = \frac{\mathbf{v}'_j \omega}{\mathbf{x}'_i} + \alpha_j + \varepsilon_{ij}$ denoting interactions between attributes and characteristics.

The researcher doesn't know ε_{ij} and therefore treats it as random, with the joint density of random vector $\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{iJ})$ denoted $f(\varepsilon_i)$.

Note:-

How should we interpret ε_{ij} ? It is not defined for a choice situation per se, rather it is defined relative to a researcher's representation of that choice situation. What does it mean for this to have a distribution?

Consider a population of people who face the same observed utility as person i . Among these people, the values of the unobserved factors differ. The density $f(\varepsilon_{ij})$ is the distribution of the unobserved portion of utility within the population of people who face the same observed utility as person i . It can also just be thought of as a subjective probability imposed by the researcher.

With this density, we can make probabilistic statements about choices. The probability that decision maker i chooses alternative j is:

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

This probability is the CDF of $\varepsilon_{ij} - \varepsilon_{ik}$ evaluated at $V_{ik} - V_{ij}$. Using the density of ε_{ij} , we can write this as:

$$\begin{aligned} P_{ij} &= P(\varepsilon_{ij} - \varepsilon_{ik} < V_{ik} - V_{ij} \forall k \neq j) \\ &= \int_{\varepsilon} I(\varepsilon_{ij} - \varepsilon_{ik} < V_{ik} - V_{ij} \forall k \neq j) f(\varepsilon_{ij}) d\varepsilon_i \end{aligned}$$

where $I(\cdot)$ is the indicator function. This is a multidimensional integral over the density of the unobserved portion of utility.

3.1.1 Identification

”Only differences in utility matter”

The absolute level of utility is irrelevant to both the decision maker and the researcher. If a constant is added to all utilities, the ordering of alternatives is unchanged. We can see this from the decomposition of the choice probability:

$$P_{ij} = P(\varepsilon_{ij} - \varepsilon_{ik} < V_{ik} - V_{ij} \forall k \neq j)$$

which only depends on differences. In general this means that the only parameters that can be estimated (identified) are those that capture differences across alternatives. This general statement takes several forms:

- **Alternative-specific constants:** α_j captures the average utility of alternative j . When they are included ε_{ij} has zero mean by construction. However, since only differences in utility matter, only differences in α_j matter, not the absolute level. To reflect this, the researcher must set the level of one constant. **With J alternatives, at most $J - 1$ constants are identified, with one of the constants normalised to zero¹.** It’s irrelevant which constant is normalised to zero, just be careful of the interpretation of the other constants.
- **Number of independent error terms.** The choice probabilities take the form of a J dimensional integral over the density of the J error terms. However, recognising that only differences in utility matter, we can reduce the dimension by defining $\tilde{\varepsilon}_{ijk} = \varepsilon_{ij} - \varepsilon_{ik}$, giving:

$$P_{ij} = \int I(\tilde{\varepsilon}_{ijk} < V_{ik} - V_{ij} \forall k \neq j) g(\tilde{\varepsilon}_{ijk}) d\tilde{\varepsilon}_{ijk}$$

This is a $J - 1$ dimensional integral!²

”The overall scale of utility is irrelevant ”

Just as adding a constant to the utility of all alternatives doesn’t change the choice, neither does multiplying each alternative’s utility by a constant. To account for this, the researcher must normalise the scale of utility.

The standard way to do this is to normalise the variance of the error terms. When utility is multiplied by κ , the variance of ε_{ij} changes by κ^2 : $Var(\kappa\varepsilon_{ij}) = \kappa^2 Var(\varepsilon_{ij})$. Therefore **normalising the variance of the error term is equivalent to normalising the scale of utility**.

When errors are assumed i.i.d., normalising is straightforward. We set the error variance to some number, usually something convenient. Since all errors have the same variance by assumption, normalising the variance of any of them sets the variance for them all. The original model becomes equivalent to $U_{ij}^1 = \mathbf{v}'_j \frac{\omega}{\sigma} + \varepsilon_{ij}^1$, with $Var(\varepsilon_{ij}^1) = 1$. The new coefficients reflect the impact of observed variables relative to standard deviation of the unobserved factors.

Example (Probit). $Var(\varepsilon_{ij}) = \sigma^2$, so we can set variance to 1 by dividing by σ .

Example (Logit). $Var(\varepsilon_{ij}) = \sigma^2 \pi^2 / 6$, so we can normalise by setting variance to $\pi^2 / 6$ (and dividing by σ). Thus

$$Pr(y_i = j | \mathbf{v}) = \frac{e^{\mathbf{v}'_j \omega / \sigma}}{\sum_{j=1}^J e^{\mathbf{v}'_j \omega / \sigma}}$$

¹The researcher could normalize to some value other than zero, of course; however, there would be no point in doing so, since normalizing to zero is easier (the constant is simply left out of the model) and has the same effect.

²Since ε_i has more elements than $\tilde{\varepsilon}_{ij}$, g is consistent with an infinite number of different f ’s. One dimension of f is not identified, and must be normalised.

Question 1

Show that for a logit parametrisation which sets the normalised variance to 1.6, to convert the probit coefficients to the same scale as logit, they should be multiplied by $\sqrt{1.6}$.

Solution:-

$$U_{ij} = \mathbf{v}_{ij}'\omega + \varepsilon_{ij} \text{ where } \text{Var}(\varepsilon_{ij}) = 1.6$$

$$\Rightarrow U_{ij}^1 = \mathbf{v}_{ij}'\frac{\omega}{\sqrt{1.6}} + \varepsilon_{ij}^1 \text{ where } \text{Var}(\varepsilon_{ij}^1) = 1$$

Since probit sets the variance to 1:

$$\omega^p = \frac{\omega^l}{\sqrt{1.6}}$$

$$\Rightarrow \sqrt{1.6}\omega^p = \omega^l$$

3.2 Multinomial Probit

We write the conditional probability of choosing alternative j' as

$$Pr(y_i = j' | \mathbf{v}_i, \theta) = \int_{-\infty}^{V_{j'} - V_1} \cdots \int_{-\infty}^{V_{j'} - V_J} g(\tilde{\varepsilon}_{1j'} \cdots \tilde{\varepsilon}_{Jj'}, \Xi_{\varepsilon J-1}) d\tilde{\varepsilon}_{1j'} \cdots d\tilde{\varepsilon}_{Jj'}$$

g is a multivariate normal density of dimension $J - 1$, $\tilde{\varepsilon}_{ij'} = \varepsilon_{ij'} - \varepsilon_{ij}$ as before, and $\Xi_{\varepsilon J-1}$ is the covariance matrix for the error differences.

There is a curse of dimensionality here, there are no closed form expressions for such high dimensional integrals, and for $J > 2$ we have to use simulation methods (SMLE).

Lemma 3.2.1 (Cholesky decomposition). Let A be a $K \times K$ matrix. We say that A possesses a Cholesky decomposition if and only if there exists a lower triangular $K \times K$ matrix L such that its diagonal entries are strictly positive real numbers and

$$A = LL'$$

Example (Trinomial probit).

$$y_j^* = \mathbf{x}'\beta + \varepsilon_j \quad j = 1, 2, 3$$

The deterministic component of utility (for alternative j) is given by $V_j = \mathbf{x}'\beta$ and $\varepsilon = (\varepsilon_1, \varepsilon_2, \varepsilon_3)' \sim N(0, \Sigma)$. We can redefine using error difference to reduce dimensionality to 2:

$$Pr(y = 1 | \mathbf{x}) = \int_{-\infty}^{V_1 - V_2} \int_{-\infty}^{V_1 - V_3} g(\tilde{\varepsilon}_{21}, \tilde{\varepsilon}_{31}, \rho) d\tilde{\varepsilon}_{21} d\tilde{\varepsilon}_{31}$$

where g is bivariate normal and ρ is the correlation between $\tilde{\varepsilon}_{21}$ and $\tilde{\varepsilon}_{31}$.

We can write the probability above as the product of conditionals as below:

$$Pr(y = 1 | \mathbf{x}) = Pr(\tilde{\varepsilon}_{21} < V_1 - V_2, \tilde{\varepsilon}_{31} < V_1 - V_3 | \mathbf{x})$$

$$= Pr(\tilde{\varepsilon}_{21} < V_1 - V_2 | \mathbf{x}) Pr(\tilde{\varepsilon}_{31} < V_1 - V_3 | \mathbf{x}, \tilde{\varepsilon}_{21} < V_1 - V_2)$$

The errors $\tilde{\varepsilon}$ are bivariate normal with covariance matrix Σ . Since Σ is positive definite, it

can be decomposed as $\Sigma = LL'$, where L is a lower triangular matrix.

$$L = \begin{bmatrix} \ell_{11} & 0 \\ \ell_{21} & \ell_{22} \end{bmatrix}$$

Using this decomposition, we can construct $\tilde{\varepsilon}$ using *independent* standard normal random variables e_1 and e_2 :

$$\begin{bmatrix} \tilde{\varepsilon}_{21} \\ \tilde{\varepsilon}_{31} \end{bmatrix} = \begin{bmatrix} \ell_{11} & 0 \\ \ell_{21} & \ell_{22} \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

Thus we can represent the probability $Pr(y = 1|\mathbf{x})$ as:

$$Pr \left(\underbrace{\begin{bmatrix} -\infty \\ -\infty \end{bmatrix}}_A \leq \underbrace{\begin{bmatrix} \ell_{11} & 0 \\ \ell_{21} & \ell_{22} \end{bmatrix}}_L \underbrace{\begin{bmatrix} e_1 \\ e_2 \end{bmatrix}}_e \leq \underbrace{\begin{bmatrix} V_1 - V_2 \\ V_1 - V_3 \end{bmatrix}}_B \middle| \mathbf{x} \right)$$

where A and B are the respective lower and upper limits of the integrals given above. The two components of $Pr(y = 1|\mathbf{x})$ are:

$$\begin{aligned} Pr(\tilde{\varepsilon}_{21} < V_1 - V_2 | \mathbf{x}) &= Pr(\ell_{11}e_1 < V_1 - V_2 | \mathbf{x}) \\ &= Pr(e_1 < \frac{V_1 - V_2}{\ell_{11}} | \mathbf{x}) \\ Pr(\tilde{\varepsilon}_{31} < V_1 - V_3 | \mathbf{x}, \tilde{\varepsilon}_{21} < V_1 - V_2) &= Pr(\ell_{21}e_1 + \ell_{22}e_2 < V_1 - V_3 | \mathbf{x}, e_1 < \frac{V_1 - V_2}{\ell_{11}}) \\ &= Pr(e_2 < \frac{V_1 - V_3 - \ell_{21}e_1}{\ell_{22}} | \mathbf{x}, e_1 < \frac{V_1 - V_2}{\ell_{11}}) \end{aligned}$$

Going forward all probabilities are conditional on \mathbf{x} and I define $b_1 = V_1 - V_2$ and $b_2 = V_1 - V_3$. From before:

$$Pr(y = 1) = Pr \left(e_1 < \frac{b_1}{\ell_{11}} \right) Pr \left(e_2 < \frac{b_2 - \ell_{21}e_1}{\ell_{22}} \middle| e_1 < \frac{b_1}{\ell_{11}} \right)$$

Suppose now we draw a random variable e_1^* from a truncated standard normal density with upper truncation point of b_1/ℓ_{11} . Then the probability can be rewritten as:

$$Pr(y = 1) = Pr \left(e_1 < \frac{b_1}{\ell_{11}} \right) Pr \left(e_2 < \frac{b_2 - \ell_{21}e_1^*}{\ell_{22}} \right)$$

The GHK simulator is the arithmetic mean of the probabilities given by the above for R random draws of e_1^* :

$$Pr^{GHK}(y = 1 | \mathbf{x}) = \frac{1}{R} \sum_{r=1}^R \Phi \left\{ \frac{b_1}{\ell_{11}} \right\} \Phi \left\{ \frac{b_2 - \ell_{21}e_1^*}{\ell_{22}} \right\}$$

where Φ is the standard normal CDF.

If $\ell_{21} = 0$ then simulation methods are pointless, the errors are already uncorrelated

The advantage of this expression is the fact that the e 's are independent normal distributed random variables and hence the probability can be equivalently expressed as a product of independent but conditioned univariate cumulative density functions.

3.3 Logit

In Multinomial Probit, for large J we need to solve a high dimensional integral. Logit is a highly tractable alternative, where the joint density $\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{iJ})$ is assumed to be i.i.d. extreme value type I. The density for each unobserved component is

$$f(\varepsilon_{ij}) = e^{-\varepsilon_{ij}} e^{-e^{-\varepsilon_{ij}}}, \quad F(\varepsilon_{ij}) = e^{-e^{-\varepsilon_{ij}}}.$$

As we used before, $\text{Var}(\varepsilon_{ij}) = \pi^2/6$ ³.

The difference between two extreme value type I random variables is a logistic random variable, then $\tilde{\varepsilon}_{ijk} = \varepsilon_{ij} - \varepsilon_{ik}$ has density

$$F(\tilde{\varepsilon}_{ijk}) = \frac{e^{\tilde{\varepsilon}_{ijk}}}{1 + e^{\tilde{\varepsilon}_{ijk}}}$$

Using this density, we can write the choice probability as:

$$Pr(y_i = j | \mathbf{v}_i) = \frac{e^{V_{ij}}}{\sum_{k=1}^J e^{V_{ik}}}$$

Clearly all probabilities are positive and sum to 1.

3.3.1 Power and limitations of logit

- **Taste variation** - Logit can represent systematic taste variation (that is, taste variation that relates to observed characteristics of the decision-maker) but not random taste variation (differences in tastes that cannot be linked to observed characteristics).
- **Substitution patterns** - The logit model implies proportional substitution across alternatives, given the researcher's specification of representative utility (see later sections on IIA).

Independence of irrelevant alternatives (IIA)

For any two alternatives j and k , the ratio of the logit probabilities is:

$$\begin{aligned} \frac{P_{ij}}{P_{ik}} &= \frac{e^{V_{ij}}}{\sum_{k=1}^J e^{V_{ik}}} \frac{\sum_{k=1}^J e^{V_{ik}}}{e^{V_{ik}}} \\ &= \frac{e^{V_{ij}}}{e^{V_{ik}}} \\ &= e^{V_{ij} - V_{ik}} \end{aligned}$$

This ratio does not depend on any alternatives other than j and k . That is, the relative odds of choosing j over k is the same no matter what other alternatives are available or what the attributes of the other alternatives are. Since the ratio is independent from alternatives other than j and k , it is said to be independent from “irrelevant” alternatives.

Example (IIA and Three Buses). A car and blue bus are initially available. Assume $Pr_c = Pr_{bB} = 0.5$, thus $Pr_c/Pr_{bB} = 1$.

Now add a red bus to the choice set Ω_J , and assume the buses are perfect substitutes ($P_{rB} = P_{bB}$). Given IIA, $Pr_c/Pr_{bB} = 1$ holds before and after the introduction of the red bus.

$$\Rightarrow Pr_c/Pr_{bB} = Pr_{rB}/Pr_{bB} = 1$$

$$\Rightarrow Pr_c = Pr_{rB} = Pr_{bB} = 1/3$$

We expect the original probability Pr_c not to change when we introduce the red bus, such that $Pr_c = 0.5$ and $Pr_{bB} = 0.25$. We overestimate Pr_{rB} and Pr_{bB} , and underestimate Pr_c .

³The mean of this distribution is not zero, but since only differences in utility matter this is not a problem

Question 2

Let U_{ij} denote the unobserved utility for individual $i \in 1, \dots, n$ who chooses from large gas cars (lgc), small gas cars (sgc) and small electric cars (sec) $j \in \text{lgc, sgc, sec}$. A model which is additive in a linear index and error term is given by:

$$U_{ij} = \mathbf{x}'_{ij}\beta + \varepsilon_{ij}$$

where ε_{ij} is an unobserved error term. The error terms are assumed to be i.i.d. extreme value type I. The choice probabilities are:

$$P_{lgc} = 0.66, \quad P_{sgc} = 0.33, \quad P_{sec} = 0.01$$

Suppose a subsidy for electric cars is introduced, and this subsidy raises P_{sec} to 0.1. What are the new values of P_{lgc} and P_{sgc} ?

Solution:-

By the logit model, the probability for each of the gas cars would be predicted to drop by the same percentage:

The probability for the large gas car would drop by 10% of its original value, from 0.66 to 0.594 and the probability for the small gas car would drop by 10% of its original value, from 0.33 to 0.297.

In terms of absolute numbers, the increase in demand for the small electric car (0.09) is predicted to come twice as much from the large gas car (0.06) as from the small gas car (0.03).

This is precisely the notion of proportional substitution.

3.3.2 Logit variants

Conditional Logit

This is the normal logit we've seen before, *conditional* on the observed characteristics/attributes we get a choice probability:

Definition 3.3.1: Conditional Logit

$$\begin{aligned} U_{ij} &= V_{ij} + \varepsilon_{ij} \quad \text{where } \varepsilon_{ij} \sim \text{EV1} \\ V_{ij} &= \mathbf{v}'_{ij}\omega \\ Pr(y_i = j|\mathbf{v}) &= \frac{e^{V_{ij}}}{\sum_{k=1}^J e^{V_{ik}}} \\ &= \frac{e^{\mathbf{v}'_{ij}\omega}}{\sum_{k=1}^J e^{\mathbf{v}'_{ik}\omega}} \end{aligned}$$

For example, when there are two alternatives, the conditional logit model is:

$$\begin{aligned} Pr(y_i = 1|\mathbf{v}) &= \frac{e^{\mathbf{v}'_{i1}\omega}}{e^{\mathbf{v}'_{i1}\omega} + e^{\mathbf{v}'_{i2}\omega}} = \frac{e^{V_{i1}}}{e^{V_{i1}} + e^{V_{i2}}} \\ &= \frac{1}{1 + e^{\mathbf{v}'_{i2}\omega - \mathbf{v}'_{i1}\omega}} \end{aligned}$$

The analyst may observe both alternative specific and alternative invariant characteristics, models including both are referred to as **hybrid models**. This could be done by specifying an additively separable model: $V_{ij} = \mathbf{v}'_{ij}\omega + \mathbf{x}'_i\beta$, or alternatively allowing a bunch of interactions: $V_{ij} = \mathbf{v}'_{ij}\omega + \sum_k^K \mathbf{v}_{ij} \times \mathbf{x}_{ik}\delta_{jk}$.

3.3.3 Identification - Aside

A consumer facing the choice between J alternatives will choose the alternative in choice set Ω_j with the highest utility. As before, this implies only differences in utility matter.

Example.

$$U_{i1} = \alpha_1 + \mathbf{x}'_i\beta_1 + \varepsilon_{i1} \quad U_{i2} = \alpha_2 + \mathbf{x}'_i\beta_2 + \varepsilon_{i2}$$

With J alternatives, only J-1 ASC's are identified (as earlier). Similarly only $\beta_{12} = \beta_1 - \beta_2$ is identified. Normalisations: set $\alpha_1 = 0$ and $\beta_1 = 0$.

3.3.4 Elasticities

Below we derive own and cross elasticities for the logit model:

$$U_{ij} = V_{ij} + \varepsilon_{ij}, \quad P_{ij} = \frac{e^{V_{ij}}}{\sum_l e^{V_{il}}}$$

Own Partial effects - We first calculate the partial effect of a change in attribute k of alternative j for individual i on the probability of choosing alternative j :

$$\begin{aligned} \frac{\partial P_{ij}}{\partial v_{ij}^k} &= \frac{\partial}{\partial v_{ij}^k} \left(\frac{e^{V_{ij}}}{\sum_l e^{V_{il}}} \right) \\ &= \frac{\sum_l e^{V_{il}} \frac{\partial V_{ij}}{\partial v_{ij}^k} e^{V_{ij}} - e^{V_{ij}} \frac{\partial V_{ij}}{\partial v_{ij}^k} e^{V_{ij}}}{(\sum_l e^{V_{il}})^2} \\ &= \frac{\partial V_{ij}}{\partial v_{ij}^k} \left(\frac{e^{V_{ij}}}{\sum_l e^{V_{il}}} - \left(\frac{e^{V_{ij}}}{\sum_l e^{V_{il}}} \right)^2 \right) \\ &= \frac{\partial V_{ij}}{\partial v_{ij}^k} (P_{ij} - P_{ij}^2) \\ &= \frac{\partial V_{ij}}{\partial v_{ij}^k} P_{ij} (1 - P_{ij}) \end{aligned}$$

If representative utility is linear in \mathbf{v}_{ij} , then $\frac{\partial V_{ij}}{\partial v_{ij}^k} = \omega_k$. Thus:

$$\frac{\partial P_{ij}}{\partial v_{ij}^k} = \omega_k P_{ij} (1 - P_{ij})$$

Own elasticity - Denote the elasticity of P_{ij} with respect to v_{ij}^k as $E_{jv_{ij}^k}$.

$$\begin{aligned} E_{jv_{ij}^k} &= \frac{\partial P_{ij}}{\partial v_{ij}^k} \frac{v_{ij}^k}{P_{ij}} \\ &= \frac{\partial V_{ij}}{\partial v_{ij}^k} P_{ij} (1 - P_{ij}) \frac{v_{ij}^k}{P_{ij}} \\ &= \frac{\partial V_{ij}}{\partial v_{ij}^k} v_{ij}^k (1 - P_{ij}) \end{aligned}$$

If representative utility is linear in \mathbf{v}_{ij} , then:

$$E_{jv_{ij}^k} = \omega_k v_{ij}^k (1 - P_{ij})$$

Cross partial effects - We now calculate the partial effect of a change in attribute h of alternative j for individual i on the probability of choosing alternative l :

$$\begin{aligned} \frac{\partial P_{ij}}{\partial v_{ih}^k} &= \frac{\partial}{\partial v_{ih}^k} \left(\frac{e^{V_{ij}}}{\sum_l e^{V_{il}}} \right) \\ &= \frac{-e^{V_{ij}} e^{V_{ih}} \frac{\partial V_{ih}}{\partial v_{ih}^k}}{(\sum_l e^{V_{il}})^2} \\ &= -\frac{\partial V_{ih}}{\partial v_{ih}^k} \left(\frac{e^{V_{ij}}}{\sum_l e^{V_{il}}} \right) \left(\frac{e^{V_{ih}}}{\sum_l e^{V_{il}}} \right) \\ &= -\frac{\partial V_{ih}}{\partial v_{ih}^k} P_{ij} P_{ih} \end{aligned}$$

If representative utility is linear in \mathbf{v}_{ij} , then:

$$\frac{\partial P_{ij}}{\partial v_{ih}^k} = -\omega_k P_{ij} P_{ih}$$

Cross elasticity - Denote the elasticity of P_{ij} with respect to v_{ih}^k as $E_{jv_{ih}^k}$.

$$\begin{aligned} E_{jv_{ih}^k} &= \frac{\partial P_{ij}}{\partial v_{ih}^k} \frac{v_{ih}^k}{P_{ij}} \\ &= -\frac{\partial V_{ih}}{\partial v_{ih}^k} P_{ij} P_{ih} \frac{v_{ih}^k}{P_{ij}} \\ &= -\frac{\partial V_{ih}}{\partial v_{ih}^k} v_{ih}^k P_{ih} \end{aligned}$$

If representative utility is linear in \mathbf{v}_{ij} , then:

$$E_{jv_{ih}^k} = -\omega_k v_{ih}^k P_{ih}$$

Interpretation of cross elasticity

The cross-elasticity of demand is the same for all choices, i.e. the formula for the cross-elasticity doesn't contain anything relating to the j^{th} alternative.

If the k^{th} attribute of h changes, then the effect on substitution probabilities to other alternatives (here alternative j) is independent of the share originally with alternative j .

- An improvement in one alternative draws proportionately from all other alternatives.
- A worsening in one alternative propels proportionately to all other alternatives, i.e. increase in price, lost demand is redistributed in equal proportions.

Example. Assume there is a single attribute denoting cost v^c per alternative, $V_{ij} = \omega v_j^c$. Then for 3 alternatives we can display $E_{jv_h^c}^L$ (the cross elasticity for alternative j - rows - with respect to the cost of alternative h - columns) as:

$$E_{jv_h^c}^L = \begin{bmatrix} (1 - P_{i1})\omega v_1^c & -P_{i2}\omega v_2^c & -P_{i3}\omega v_3^c \\ -P_{i1}\omega v_1^c & (1 - P_{i2})\omega v_2^c & -P_{i3}\omega v_3^c \\ -P_{i1}\omega v_1^c & -P_{i2}\omega v_2^c & (1 - P_{i3})\omega v_3^c \end{bmatrix}$$

For $\omega < 1$ all own price (cross) elasticities are negative (positive).