4 GEV

4.1 Functional Form of Choice Probabilities in Simple Cases

In some situations, independence from irrelevant alternatives (IIA) holds for some pairs of alternatives but not all. Logit is inappropriate in these situations since it assumes there is IIA between each pair of alternatives; however, a probit approach, even if feasible, might be unduly complex and expensive since it does not exploit the fact that IIA holds for some pairs of alternatives. Another qualitative choice model, called GEV for reasons to be described, is designed to handle situations like these.

A type of GEV model is used when the set of alternatives faced by a decisionmaker can be partitioned into subsets such that the ratio of probabilities for any two alternatives that are in the same subset is independent of the existence or characteristics of other alternatives. An example can best explain how to determine whether a set of alternatives can be so partitioned. Suppose the set of alternatives available to a worker for his commute to work consists of driving an auto alone, carpooling, taking the bus, and taking rail. If any one alternative were removed, the probabilities of the other alternatives would increase (e.g., if the worker became injured and could not drive an auto, then the probability of carpooling, bus, and rail would increase). The relevant question in partitioning these alternatives is, By what proportion would each probability increase when an alternative is removed? Suppose the changes in probabilities occur as set forth in table 4.1. Note that

- When the auto alone alternative is removed, the bus and rail probabilities increase by the same proportion, and consequently the ratio of their probabilities stays constant. That is, the ratio of bus and rail probabilities is independent of the existence of the auto alone alternative. However, the carpool probability increases more, proportionately, than either the bus or rail probability, meaning that the ratio of the carpool probability to either the bus or rail probability when the auto alone alternative is included differs from this ratio when it is removed. Therefore, the ratio of the carpool probability to either the bus or rail alternative is **not** independent of the existence of the auto alternative.
- Similarly, the change in probabilities that occurs when the carpool alternative is removed indicates that the ratio of the bus and rail probabilities is independent of the existence of carpool, but the ratio of the auto alone probability to either the bus or rail probability is not independent of the existence of the carpool alternative.

Table 4.1
Example of IIA holding within subsets of alternatives

Increase in probability of remaining alternatives, as % of original probability	Alternative removed			
	Auto alone	Carpool	Bus	Rail
Auto alone	—	5	7	7
Carpool	40	—	7	7
Bus	10	3	—	50
Rail	10	3	80	—

• With respect to removal of the bus alternative, the ratio of auto alone and carpool probabilities is independent, but not the ratio of the rail probability to either the auto alone or carpool probability.
• Finally, with respect to removal of the rail alternative, the ratio of auto alone and carpool probabilities is independent, but not the ratio of the rail probability to either the auto alone or carpool probability.

These facts suggest a partition of the set of alternatives into two subsets, with auto alone and carpool in one subset and bus and rail in the other subset. Under this partition, all relations among probabilities can be described succinctly: The ratio of probabilities of any two alternatives within the same subset is independent of the existence of other alternatives; however, the ratio of probabilities of two alternatives from different subsets is not independent of the existence of other alternatives. That is, IIA holds within subsets but not across subsets.

A convenient way to picture the choice situation is with a tree diagram. In such a tree, each branch denotes a subset of alternatives within which IIA holds, and every leaf on each branch denotes an alternative. For example, the tree diagram for the worker's choice of mode described above is given in figure 4.1. The (upside down) tree consists of two branches, labeled "auto" and "transit," for the two subsets of alternatives, and each of the branches contains two leaves for the two alternatives within the subset. Note that auto and transit are not themselves alternatives available to the worker, but rather are simply the names of groups of alternatives, designating the common feature among the alternatives within the group.

For any situation in which the alternatives can be partitioned in the manner described, or more graphically, depicted in a tree diagram with IIA

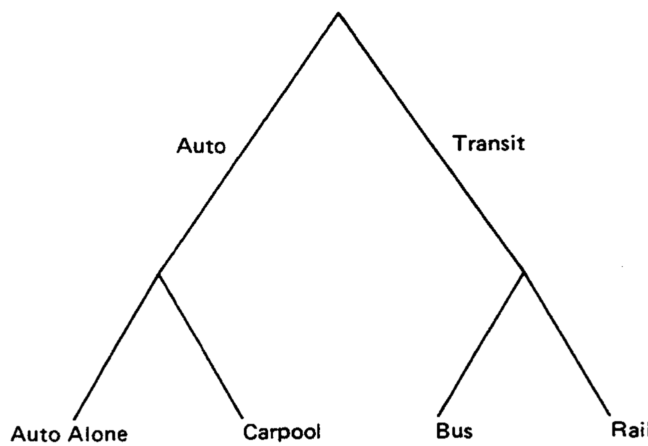

Figure 4.1
Tree diagram for mode choice.

holding for all leaves on each branch, a GEV model can be used to describe the choice situation.

Formally, the GEV model is specified as follows. Let the set of alternatives J_n be partitioned into K subsets denoted $B_n^1, \ldots, B_n^K$. The utility that person n obtains from alternative i in subset B_n^k is denoted, as usual, as $U_{in} = V_{in} + e_{in}$, where V_{in} is observed by the researcher and e_{in} is a random variable whose value is not observed by the researcher. The GEV model is obtained by assuming that e_{in}, for all elements i in J_n, are distributed in accordance with a generalized extreme value (GEV) distribution. That is, the joint cumulative distribution of the random variables e_{in}, for all i in J_n, is assumed to be

$$\exp\left\{-\sum_{k=1}^{K} \alpha_k \left(\sum_{i \in B_n^k} e^{-e_{in}/\lambda_k}\right)^{\lambda_k}\right\}.$$

This distribution, as its name implies, is a generalization of the distribution that gives rise to the logit model. For logit, each e_{in} is independent with a univariate extreme value distribution. For GEV, the **marginal** distribution of each e_{in} is univariate extreme value, but all e_{in} within each subset are correlated with each other. The parameter λ_k is a measure of the correlation of unobserved utility within subset B_n^k. More precisely, $(1 - \lambda_k)$ is a measure of correlation since λ_k itself drops as the correlation rises.[1] For any i and j in different subsets (that is, i in B_n^k and j in B_n^h, where $k \neq h$), there is no correlation between e_{in} and e_{jn}.

McFadden (1978) has shown, using a proof that is complex and, since it is not heuristic, will not be reproduced here, that this distribution for the unobserved components of utility gives rise to the following choice probability for alternative i in subset B_n^k:

$$P_{in} = \frac{e^{V_{in}/\lambda_k}(\sum_{j \in B_n^k} e^{V_{jn}/\lambda_k})^{\lambda_k - 1}}{\sum_l (\sum_{j \in B_n^l} e^{V_{jn}/\lambda_l})^{\lambda_l}}. \tag{4.1}$$

Note that when $\lambda_k = 1$ for all k (and hence $1 - \lambda_k = 0$, indicating no correlation between the unobserved components of utility for alternatives within a subset), the choice probabilities become simply logit. Consequently, the GEV model is a generalization of logit that allows for particular patterns of correlation in unobserved utility.

Expression (4.1) is complex and, aside from the fact that it reduces to logit when all $\lambda_k = 1$, is not very illuminating. However, the choice probabilities can be expressed in an alternative fashion, as follows, that is quite simple and readily interpretable.

Without loss of generality, the observed component of utility can be decomposed into two parts: (1) a part that is constant for all alternatives within a subset, and (2) a part that is not constant within subsets. This can be denoted

$$U_{in} = W_n^k + \lambda_k Y_{in}^k + e_{in}, \qquad \text{for} \quad i \text{ in } B_n^k,$$

where

W_n^k is the mean of V_{in} over all alternatives in subset B_n^k;
Y_{in}^k is the deviation of V_{in} from the mean W_n^k; and
λ_k is a normalizing constant whose meaning will become evident.

Note that W_n^k varies over k (i.e., subsets) but not over i (i.e., alternatives within a subset), while Y_{in}^k varies over both k and i. Note that this decomposition is completely general, since Y_{in}^k is defined simply as $(V_{in} - W_n^k)/\lambda_k$.

Let the probability of choosing alternative i in subset B_n^k be expressed as the product of the probability that an alternative within subset B_n^k is chosen and the probability that alternative i is chosen (given that an alternative in B_n^k is chosen). This is denoted as

$$P_{in} = P_{in|B_n^k} \cdot P_{B_n^k},$$

where

$P_{in|B_n^k}$ is the conditional probability of choosing alternative i given that an alternative in the subset B_n^k is chosen, and

$P_{B_n^k}$ is the marginal probability of choosing an alternative in B_n^k (with the marginality being over all alternatives in B_n^k).

Note that this equality is exact since any probability can be written as the product of a marginal and a conditional probability.

The reason for decomposing P_{in} into a marginal and a conditional probability is that, with the GEV formula for P_{in}, the marginal and conditional probabilities take the form of logits. In particular, the marginal and conditional probabilities can be expressed as[2]

$$P_{in|B_n^k} = \frac{e^{Y_{in}^k}}{\sum_{j \in B_n^k} e^{Y_{jn}^k}},$$

$$P_{B_n^k} = \frac{e^{W_n^k + \lambda_k I_k}}{\sum_{l=1}^K e^{W_n^l + \lambda_l I_l}},$$

where

$$I_k = \ln \sum_{j \in B_n^k} e^{Y_{jn}^k}.$$

Stated in words, the conditional probability of choosing i, given that an alternative in B_n^k is chosen, is expressed as logit with variables that vary over alternatives within each subset entering representative utility in the logit formula. The marginal probability of choosing an alternative in B_n^k is also expressed as logit with the variables that vary over subsets of alternatives (but not over alternatives within each subset) entering representative utility. In addition, the representative utility in the marginal probability includes a term (i.e., I_k) that is the log of the denominator of the conditional probability. This term denotes the average utility that the person can expect from the alternatives within the subset. In recognition of this, the term I_k is called the "inclusive value" or "inclusive utility" of subset k.[3]

With this specification of the choice probabilities, it is clear that IIA holds within each subset but not across subsets. Consider two alternatives, i and m, both of which are in subset B_n^k.

$$P_{in}/P_{mn} = \frac{P_{in|B_n^k} \cdot P_{B_n^k}}{P_{mn|B_n^k} \cdot P_{B_n^k}} = \frac{P_{in|B_n^k}}{P_{mn|B_n^k}}$$

$$= \exp(Y_{in}^k)/\exp(Y_{mn}^k),$$

which is independent of alternatives other than m and i. However, for two alternatives in different subsets, say i in B_n^k and r in B_n^h,

$$P_{in}/P_{rn} = \frac{P_{in|B_n^k} \cdot P_{B_n^k}}{P_{rn|B_n^h} \cdot P_{B_n^h}},$$

which depends on the characteristics of all alternatives in B_n^k and B_n^h.

Example Because of the feature of the GEV model that IIA hold within subsets but not across subsets, it is particularly well-suited for describing situations like that of the worker's choice of mode, presented earlier. Suppose the only observed factors affecting each worker's choice are the total cost, c_{in}, and total time, t_{in}, of traveling on each mode. With linear utility, we have for worker n,

$$U_{in} = \alpha c_{in} + \beta t_{in} + e_{in},$$

where i denotes the mode. Similar unobserved factors enter the utility for auto alone and carpool (i.e., avoidance of strangers), making the e_{in} for these two alternatives correlated. The e_{in} for bus and rail are also correlated. However, there is, by assumption, no correlation between the unobserved utility of either transit mode with that of either auto mode.

An appropriate specification for this situation is a GEV model consisting of three "submodels": (1) a marginal probability submodel of the choice between auto and transit; (2) a conditional probability submodel for the choice of auto alone or carpool given that an auto mode is chosen; and (3) a conditional probability submodel for the choice of bus or rail given that a transit mode is chosen. This specification follows the tree diagram in figure 4.1, with a submodel for each of the three nodes in the tree.

To specify the variables entering each model, calculate the average time and cost of travel by transit and auto, that is, calculate

$$\bar{t}_n^{\mathrm{t}} = (t_{\mathrm{b}n} + t_{\mathrm{r}n})/2;$$

$$\bar{t}_n^{\mathrm{a}} = (t_{\mathrm{a}n} + t_{\mathrm{c}n})/2;$$

$$\bar{c}_n^{\mathrm{t}} = (c_{\mathrm{b}n} + c_{\mathrm{r}n})/2;$$

$$\bar{c}_n^{\mathrm{a}} = (c_{\mathrm{a}n} + c_{\mathrm{c}n})/2;$$

where subscripts a, c, b, and r denote auto alone, carpool, bus, and rail, respectively, and superscripts a and t denote auto and transit, respectively (or, more precisely, superscripts a and t denote the subsets of alternative labeled auto and transit). The time and cost of each mode is decomposed into the average for the mode's subset, just given, and a deviation from this average (denoted with a tilde over the letter):

$$t_{in} = \bar{t}_n^k + \tilde{t}_{in};$$

$$c_{in} = \bar{c}_n^k + \tilde{c}_{in};$$

for $i = \mathrm{a, c, b, r}$ and $k = \mathrm{a}$ or t as appropriate. Then utility for each mode can be decomposed into a portion that varies over subsets (i.e., over auto versus transit) but not modes within the subset and another component that varies over modes within a subset:

$$U_{in} = W_n^k + \lambda_k Y_{in}^k + e_{in},$$

where

$$W_n^k = \alpha \bar{c}_n^k + \beta \bar{t}_n^k;$$

$$Y_{in}^k = (\alpha \tilde{c}_{in} + \beta \tilde{t}_{in})/\lambda_k.$$

The three submodels can now be written explicitly. The submodel for the choice between auto alone and carpool, given that an auto mode is chosen, is logit with the conditional choice probabilities being

$$P_{in|\mathrm{a}} = \frac{e^{Y_{in}^{\mathrm{a}}}}{e^{Y_{\mathrm{a}n}^{\mathrm{a}}} + e^{Y_{\mathrm{c}n}^{\mathrm{a}}}}, \qquad \text{for} \quad i = \mathrm{a, c}.$$

Two explanatory variables enter the representative utility (Y_{in}^{a}) of each alternative. For the auto mode the variables are the cost of travel by auto alone expressed as a deviation from the average cost of travel by auto alone and carpool, and the time of travel by auto alone expressed as a deviation from the average time of travel by auto alone and carpool. For the carpool mode, similar variables enter. The coefficients of these two variables are $\alpha/\lambda_{\mathrm{a}}$ and $\beta/\lambda_{\mathrm{a}}$, respectively. (An estimate of λ_{a} is obtained from the marginal submodel of auto versus transit, and so estimates of the coefficients of cost and time in this conditional submodel provide estimates of α and β.)

The submodel for the choice of bus or rail, given that a transit mode is chosen, is also logit with two explanatory variables entering the model: the deviation of cost and time of each mode from the average cost and time for both transit modes.[4]

The submodel for the choice between the auto and transit subsets is also logit with the marginal choice probabilities taking the form

$$P_n^k = \frac{e^{W_n^k + \lambda_k I_k}}{e^{W_n^a + \lambda_a I_a} + e^{W_n^t + \lambda_t I_t}}, \qquad \text{for} \quad k = \text{a or t}.$$

Three explanatory variables enter the "representative" utility of the auto and transit "modes" (quotation marks are used around these terms to indicate that auto and transit are not actually modes but groups of modes). For the auto "mode," the average cost and time of travel by auto enter (averaged over auto alone and carpool). In addition, the inclusive value of the auto modes (calculated as $I_a = \log(\exp(Y_{an}^a) + \exp(Y_{cn}^a))$ or, more simply, as the log of the denominator on the conditional submodel for auto versus carpool) enters as an explanatory variable. The coefficient of the cost and time variables are α and β, respectively, and the coefficient of the inclusive value term is λ_a. For the transit "mode," three similar variables enter, with averages being over the bus and rail modes. The coefficient of the inclusive value term for transit is λ_t.

Remark A final note is required concerning terminology. Since, in GEV models, the subsets can be considered "nests" of alternatives, and since the choices of nests and alternatives within nests are described by logit formulas, the GEV model is often called "nested logit." Other commonly used terms for the GEV model are structured, or ordered, logit (to emphasize that the set of alternatives has a particular structure, or order, as represented by the partitioning) and sequential logit (since GEV probabilities are a sequence of marginal and conditional probabilities which are logit in form). Two of these terms, however, should be avoided. The term "ordered logit" has been used to denote models other than GEV and consequently can cause confusion. The term "sequential logit" can be misunderstood to suggest that the **decisionmaker** makes a sequence of choices, each of which is described by logit, whereas the GEV model is derived by assuming the decisionmaker makes one choice, namely, one alternative out of the available set. The sequence of probabilities in the GEV model is simply a method for the **researcher** to represent the lack of IIA among the choice probabilities.

4.2 More Complex GEV Models

The GEV model just described is called a two-level GEV model because there are, in a sense, two levels of modeling: the marginal probabilities and the conditional probabilities. In the case of the mode choice, the two levels are the marginal model of auto versus transit and the conditional model of **type** of auto or transit (auto alone or carpool given auto, and bus or rail given transit).

In some situations, however, three- or higher level GEV models are appropriate. Three-level GEV models are obtained by partitioning the set of alternatives into subsets and then partitioning the subsets into subsubsets. One logit model is used to describe the choice of subset; another logit model is used to describe the choice of subsubset; and a third describes the choice of alternative within the subsubset. The first of these models includes an inclusive value term that represents the average utility that the decisionmaker can expect from the subsubsets with each subset. This is defined as the log of the denominator of the second model. Similarly, the second model includes an inclusive value term that represents the average utility that the decisionmaker can expect from the alternatives within each subsubset. It is defined as the log of the denominator of the third model.

As an example, a household's choice of housing unit can perhaps be described as a three-level GEV model. The household has a choice among all the available housing units in the household's area of residence. The housing units can be grouped according to neighborhood within the city, and then by the number of bedrooms in the unit. Using San Francisco, a tree diagram depicting this situation is given in figure 4.2. Following this tree diagram, the set of housing units are partitioned into subsets on the basis of neighborhood and into subsubsets on the basis of the number of bedrooms. A GEV model on this partitioning assumes that (1) the ratio of probabilities of two housing units in the same neighborhood **and** with the same number of bedrooms is independent of other alternatives, (2) the ratio of probabilities of two housing units in the same neighborhood but with different numbers of bedrooms is independent of the characteristics of housing in other neighborhoods but not independent of the characteristics of housing units in the same neighborhood, and (3) the ratio of probabilities of two housing units in different neighborhoods is not independent of the characteristics of any other housing units.

More complex GEV models, with, for example, overlapping "nests," can also be constructed (see Ben-Akiva and Lerman, 1985).

4.3 Estimation

The parameters of a GEV model can be estimated by standard maximum likelihood techniques. Substituting the choice probabilities of expression

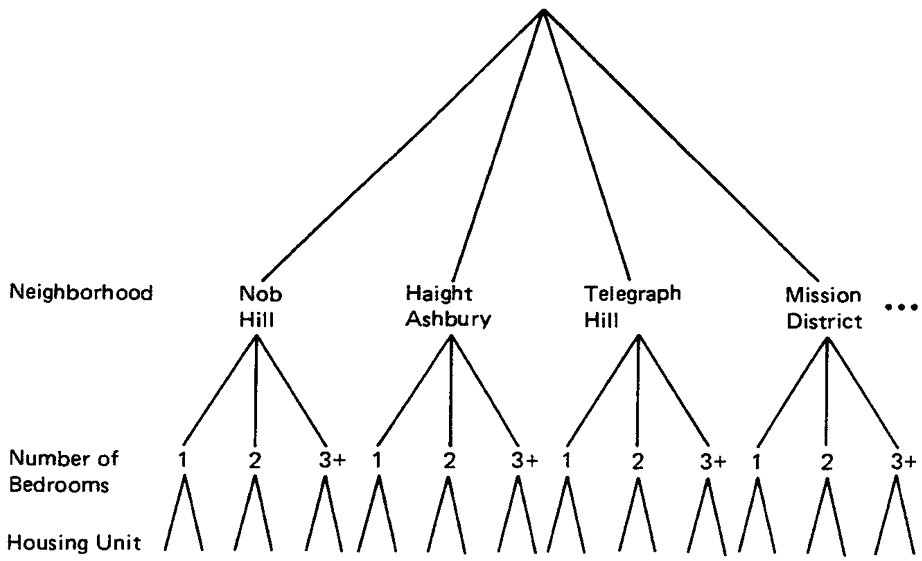

Figure 4.2
Tree diagram for choice of housing unit. (There are numerous housing units in each area with each number of bedrooms; only two "leaves" are drawn for simplicity.)

(4.1) into the log likelihood function defined in section 2.6 gives the log likelihood as an explicit function of the parameters of this GEV model. The value of the parameters that maximizes this function is, under fairly general conditions, consistent and efficient.

Since the GEV choice probabilities are fairly complex, estimation by this standard maximum likelihood method is somewhat difficult. Computer programs are now available for estimating GEV models in this way. However, they are not widely available, and the procedure tends to be relatively expensive.

For these reasons, researchers often estimate GEV models in a sequential fashion, exploiting the fact that the GEV choice probabilities can be decomposed into marginal and conditional probabilities that are logit. This sequential estimation is performed "bottom up," in that the submodels for the lower nodes of a tree diagram are estimated first, followed by the submodels for the higher nodes.

For a simple two-level GEV model, the procedure is the following. First, the logit models for the conditional probabilities are estimated using standard logit estimation routines. In the example of mode choice, described in section 4.1, these are the models of auto alone versus carpool and bus versus rail. Next, the inclusive value terms are calculated by taking the log of the denominator of the models estimated in the first step. Last, the logit model for the marginal probabilities is estimated with each of the inclusive value terms included as an explanatory variable. For the choice of mode, this is the model of auto versus transit with an inclusive value term from the model of auto alone versus carpool and another one from the model of bus versus rail included as explanatory variables. The estimated coefficients of the inclusive value terms are the estimates of the λ_k.

This all sounds very straightforward. There are, however two complications. First, in estimating the marginal submodel, an estimate of each I_k (based on the previously estimated conditional submodels) is entered rather than the "true" I_k. The estimate of I_k is consistent (since the conditional submodel is estimated consistently) and so the parameters of the marginal submodel are still estimated consistently. However, the standard errors of these parameters will be biased. In particular, the "true" standard errors will be larger than those estimated under the incorrect assumption that the inclusive value terms entering the submodel are without error. With downwardly biased standard errors, smaller confidence bounds and larger t-statistics are estimated for the parameters than are "true," and the submodel will appear to be better than it actually is.

Second, it is often the case that some parameters will be common to both the conditional and the marginal submodels. In the example of mode choice in section 4.1, the coefficients of the cost and time of travel (that is, α and β) appear in the conditional submodels of mode choice given auto or transit and the marginal submodel of auto versus transit. Estimating the conditional and marginal submodels sequentially results in two separate estimates of these parameters. It is always possible to specify a GEV model in such a way that different parameters enter each submodel (e.g., by letting the coefficients vary over alternatives, with the average coefficient estimated in the marginal submodel and the deviation from average for each alternative being estimated in the conditional submodel). However, the researcher might not think that such a specification truly describes the choice process being modeled.

These two complications are symptoms of a more general circumstance, namely, that sequential estimation of GEV models, while consistent, is not as efficient as simultaneous estimation (that is, standard maximum likelihood estimation of the complete GEV model). With simultaneous estimation, all information is utilized in the estimation of each parameter, and parameters that are common across submodels are necessarily constrained to be equal. Consequently, if a computer routine for maximum likelihood estimation of GEV models can be obtained, the addition expense[5] is probably warranted.