can be written as

$$\ell(\boldsymbol{\beta}^1, \ldots, \boldsymbol{\beta}^J) = \sum_{j=1}^{J} \sum_{y_t=j} \boldsymbol{X}_t\boldsymbol{\beta}^j - \sum_{t=1}^{n} \log\Big(1 + \sum_{j=1}^{J} \exp(\boldsymbol{X}_t\boldsymbol{\beta}^j)\Big).$$

This function is a sum of contributions from each observation. Each contribution has two terms: The first is $\boldsymbol{X}_t\boldsymbol{\beta}^j$, where the index j is that for which $y_t = j$ (or zero if $j = 0$), and the second is minus the logarithm of the denominator that appears in (15.35) and (15.36).

One important property of the multinomial logit model is that

$$\frac{\Pr(y_t = l)}{\Pr(y_t = j)} = \frac{\exp(\boldsymbol{X}_t\boldsymbol{\beta}^l)}{\exp(\boldsymbol{X}_t\boldsymbol{\beta}^j)} = \exp\big(\boldsymbol{X}_t(\boldsymbol{\beta}^l - \boldsymbol{\beta}^j)\big) \tag{15.38}$$

for any two responses l and j (including response zero if we interpret $\boldsymbol{\beta}^0$ as a vector of zeros). Thus the odds between any two responses depend solely on $\boldsymbol{X}_t$ and on the parameter vectors associated with those two responses. They do not depend on the parameter vectors associated with any of the other responses. In fact, we see from (15.38) that the log of the odds between responses l and j is simply $\boldsymbol{X}_t\boldsymbol{\beta}^*$, where $\boldsymbol{\beta}^* \equiv (\boldsymbol{\beta}^l - \boldsymbol{\beta}^j)$. Thus, conditional on either j or l being chosen, the choice between them is determined by an ordinary logit model with parameter vector $\boldsymbol{\beta}^*$.

Closely related to the multinomial logit model is the **conditional logit model** pioneered by McFadden (1974a, 1974b). See Domencich and McFadden (1975), McFadden (1984), and Greene (1990a, Chapter 20) for detailed treatments. The conditional logit model is designed to handle consumer choice among J (*not* $J + 1$) discrete alternatives, where one and only one of the alternatives can be chosen. Suppose that when the i^{th} consumer chooses alternative j, he or she obtains utility

$$U_{ij} = \boldsymbol{W}_{ij}\boldsymbol{\beta} + \varepsilon_{ij},$$

where $\boldsymbol{W}_{ij}$ is a row vector of characteristics of alternative j as they apply to consumer i. Let y_i denote the choice made by the i^{th} consumer. Presumably $y_i = l$ if U_{il} is at least as great as U_{ij} for all $j \neq l$. Then if the disturbances ε_{ij} for $j = 1, \ldots, J$ are independent and identically distributed according to the Weibull distribution, it can be shown that

$$\Pr(y_i = l) = \frac{\exp(\boldsymbol{W}_{il}\boldsymbol{\beta})}{\sum_{j=1}^{J} \exp(\boldsymbol{W}_{ij}\boldsymbol{\beta})}. \tag{15.39}$$

This closely resembles (15.37), and it is easy to see that the probabilities must add to unity.

There are two key differences between the multinomial logit and conditional logit models. In the former, there is a single vector of independent variables for each observation, and there are J different vectors of parameters.

In the latter, the values of the independent variables vary across alternatives, but there is just a single parameter vector $\boldsymbol{\beta}$. The multinomial logit model is a straightforward generalization of the logit model that can be used to deal with any situation involving three or more unordered qualitative responses. In contrast, the conditional logit model is specifically designed to handle consumer choices among discrete alternatives based on the characteristics of those alternatives.

Depending on the nature of the explanatory variables, there can be a number of subtleties associated with the specification and interpretation of conditional logit models. There is not enough space in this book to treat these adequately, and so readers who intend to estimate such models are urged to consult the references mentioned above. One important property of conditional logit models is the analog of (15.38):

$$\frac{\Pr(y_i = l)}{\Pr(y_i = j)} = \frac{\exp(\boldsymbol{W}_{il}\boldsymbol{\beta})}{\exp(\boldsymbol{W}_{ij}\boldsymbol{\beta})}. \tag{15.40}$$

This property is called the **independence of irrelevant alternatives**, or **IIA**, property. It implies that adding another alternative to the model, or changing the characteristics of another alternative that is already included, will not change the odds between alternatives l and j.

The IIA property can be extremely implausible in certain circumstances. Suppose that there are initially two alternatives for traveling between two cities: flying Monopoly Airways and driving. Suppose further that half of all travelers fly and the other half drive. Then Upstart Airways enters the market and creates a third alternative. If Upstart offers a service identical to that of Monopoly, it must gain the same market share. Thus, according to the IIA property, one third of the travelers must take each of the airlines and one third must drive. So the automobile has lost just as much market share from the entry of Upstart Airways as Monopoly Airways has! This seems very implausible.[6] As a result, a number of papers have been devoted to the problem of testing the independence of irrelevant alternatives property and finding tractable models that do not embody it. See, in particular, Hausman and Wise (1978), Manski and McFadden (1981), Hausman and McFadden (1984), and McFadden (1987).

This concludes our discussion of qualitative response models. More detailed treatments may be found in surveys by Maddala (1983), McFadden (1984), Amemiya (1981; 1985, Chapter 9), and Greene (1990a, Chapter 20), among others. In the next three sections, we turn to the subject of limited dependent variables.

[6] One might object that a price war between Monopoly and Upstart would convince some drivers to fly instead. So it would. But if the two airlines offered lower prices, that would change one or more elements of the $\boldsymbol{W}_{ij}$'s associated with them. The above analysis assumes that all the $\boldsymbol{W}_{ij}$'s remain unchanged.