The ordered probit model is widely used in applied econometric work. A simple, graphical exposition of this model is provided by Becker and Kennedy (1992). Like the ordinary probit model, the ordered probit model can be generalized in a number of ways; see, for example, Terza (1985). An interesting application of a generalized version, which allows for heteroskedasticity, is Hausman, Lo, and MacKinlay (1992). They apply the model to price changes on the New York Stock Exchange at the level of individual trades. Because the price change from one trade to the next almost always takes on one of a small number of possible values, an ordered probit model is an appropriate way to model these changes.

The Multinomial Logit Model

The key feature of ordered qualitative response models like the ordered probit model is that all the choices depend on a single index function. This makes sense only when the responses have a natural ordering. A different sort of model is evidently necessary to deal with unordered responses. The most popular of these is the **multinomial logit model**, sometimes called the **multiple logit model**, which has been widely used in applied work.

The multinomial logit model is designed to handle $J+1$ responses, for $J \geq 1$. According to this model, the probability that any one of them is observed is

$$\Pr(y_t = l) = \frac{\exp(\boldsymbol{W}_{tl}\boldsymbol{\beta}^l)}{\sum_{j=0}^{J}\exp(\boldsymbol{W}_{tj}\boldsymbol{\beta}^j)} \quad \text{for } l = 0, \ldots, J. \tag{11.34}$$

Here $\boldsymbol{W}_{tj}$ is a row vector of dimension k_j of observations on variables that belong to the information set of interest, and $\boldsymbol{\beta}^j$ is a k_j–vector of parameters, usually different for each $j = 0, \ldots, J$.

Estimation of the multinomial logit model is reasonably straightforward. The loglikelihood function can be written as

$$\sum_{t=1}^{n}\bigg(\sum_{j=0}^{J} I(y_t = j)\boldsymbol{W}_{tj}\boldsymbol{\beta}^j - \log\Big(\sum_{j=0}^{J}\exp(\boldsymbol{W}_{tj}\boldsymbol{\beta}^j)\Big)\bigg), \tag{11.35}$$

where $I(\cdot)$ is the indicator function. Thus each observation contributes two terms to the loglikelihood function. The first is $\boldsymbol{W}_{tj}\boldsymbol{\beta}^j$, where $y_t = j$, and the second is minus the logarithm of the denominator that appears in (11.34). It is generally not difficult to maximize (11.35) by using some sort of modified Newton method, provided there are no perfect classifiers, since the loglikelihood function (11.35) is globally concave with respect to the entire vector of parameters, $[\boldsymbol{\beta}^0 \vdots \ldots \vdots \boldsymbol{\beta}^J]$; see Exercise 11.16.

Some special cases of the multinomial logit model are of interest. One of these arises when the explanatory variables $\boldsymbol{W}_{tj}$ are the same for each choice j. If a model is intended to explain which of an unordered set of outcomes applies