to the different individuals in a sample, then the probabilities of all of these outcomes can be expected to depend on the same set of characteristics for each individual. For instance, a student wondering how to spend Saturday night may be able to choose among studying, partying, visiting parents, or going to the movies. In choosing, the student takes into account things like grades on the previous midterm, the length of time since the last visit home, the interest of what is being shown at the local movie theater, and so on. All these variables affect the probability of each possible outcome.

For models of this sort, it is not possible to identify $J+1$ parameter vectors $\boldsymbol{\beta}^j$, $j = 0, \ldots, J$. To see this, let $\boldsymbol{X}_t$ denote the common set of explanatory variables for observation t, and define $\boldsymbol{\gamma}^j \equiv \boldsymbol{\beta}^j - \boldsymbol{\beta}^0$ for $j = 1, \ldots, J$. On replacing the $\boldsymbol{W}_{tj}$ by $\boldsymbol{X}_t$ for all j, the probabilities defined in (11.34) become, for $l = 1, \ldots, J$,

$$\Pr(y_t = l) = \frac{\exp(\boldsymbol{X}_t\boldsymbol{\beta}^l)}{\sum_{j=0}^{J}\exp(\boldsymbol{X}_t\boldsymbol{\beta}^j)} = \frac{\exp(\boldsymbol{X}_t\boldsymbol{\gamma}^l)}{1+\sum_{j=1}^{J}\exp(\boldsymbol{X}_t\boldsymbol{\gamma}^j)},$$

where the second equality is obtained by dividing both the numerator and the denominator by $\exp(\boldsymbol{X}_t\boldsymbol{\beta}^0)$. For outcome 0, the probability is just

$$\Pr(y_t = 0) = \frac{1}{1+\sum_{j=1}^{J}\exp(\boldsymbol{X}_t\boldsymbol{\gamma}^j)}.$$

It follows that all $J+1$ probabilities can be expressed in terms of the parameters $\boldsymbol{\gamma}^j$, $j = 1, \ldots, J$, independently of $\boldsymbol{\beta}^0$. In practice, it is easiest to impose the restriction that $\boldsymbol{\beta}^0 = \mathbf{0}$, which is then enough to identify the parameters $\boldsymbol{\beta}^j$, $j = 1, \ldots, J$. When $J = 1$, it is easy to see that this model reduces to the ordinary logit model with a single index function $\boldsymbol{X}_t\boldsymbol{\beta}^1$.

In certain cases, some but not all of the explanatory variables are common to all outcomes. In that event, for the common variables, a separate parameter cannot be identified for each outcome, for the same reason as above. In order to set up a model for which all the parameters are identified, it is necessary to set to zero those components of $\boldsymbol{\beta}^0$ that correspond to the common variables. Thus, for instance, at most J of the $\boldsymbol{W}_{tj}$ vectors can include a constant.

Another special case of interest is the so-called **conditional logit model**. For this model, the probability that agent t makes choice l is

$$\Pr(y_t = l) = \frac{\exp(\boldsymbol{W}_{tl}\boldsymbol{\beta})}{\sum_{j=0}^{J}\exp(\boldsymbol{W}_{tj}\boldsymbol{\beta})}. \tag{11.36}$$

where $\boldsymbol{W}_{tj}$ is a row vector with k components for each $j = 0, \ldots, J$, and $\boldsymbol{\beta}$ is a k--vector of parameters, the same for each j. This model has been extensively used to model the choice among competing modes of transportation. The usual interpretation is that the elements of $\boldsymbol{W}_{tj}$ are the characteristics of