Chapter 3
Choice Models

3.1 Introduction
This chapter describes the characteristics of random utility choice model in a general setting, specific elements related to the conjoint choice context are given in chapters 4 and 5. In random utility choice models (also known as discrete choice models) a subject \( j \) chooses between \( M \) distinct choice alternatives, and it is assumed that he will choose the alternative that gives maximal utility. The \( (Mx1) \) vector of (unobserved) utility that the \( j \)th individual derives from the \( M \) alternatives, \( U_j \), is equal to:

\[
U_j = X_j\beta + e_j ,
\]

(3.1)

where \( X_j \) is a \((MxS)\) matrix of variables representing characteristics of the \( M \) choice alternatives for the \( j \)th individual, \( \beta \) is a \((Sx1)\) vector of unknown parameters, and \( e_j \) is the \((Mx1)\) error term that also may include effects from attributes not specified in the matrix \( X \). In a more general specification, the parameter vector \( \beta \) can depend on \( j \) or can contain different elements for different alternatives \( m \) or both, and the matrix \( X \) need not depend on \( j \) or may have equal rows for all \( m \). Furthermore, it can contain, for example, quadratic main effects and interaction effects. When the \( X \)-matrix does not depend on \( m \), the model in (3.1) is called an alternative specific model and \( \beta \) then has to depend on \( m \). For each individual \( j \), it is assumed that the alternative with the highest utility is chosen. The variable \( y_{jm} \) describes the observed choices and is defined as:

\[
y_{jm} = \begin{cases} 
1 & \text{when } U_{jm} > U_{jn} \forall n \neq m \\
0 & \text{when } \exists n \neq m : U_{jn} > U_{jm} , n = 1,\ldots,M .
\end{cases}
\]

(3.2)

Let \( p_{jm} \) be the probability that \( y_{jm} \) equals one. Then, when there are \( J \) individuals, the likelihood function is equal to:
Chapter 3

The Multinomial Logit model for conjoint choice experiments is described in section 4.2, the Multinomial Probit model for conjoint choice experiments in section 5.2. In this chapter the MNL and MNP model are described in the standard context with only one choice set for each respondent.

Maximum likelihood estimates of the parameters in (3.1) are obtained by maximizing (3.3). In most cases not the likelihood itself is evaluated but the log-likelihood instead:

\[ l = \sum_{j=1}^{J} \sum_{m=1}^{M} y_{jm} \ln(p_{jm}) . \]  

(3.4)

The form of the probabilities in (3.3) or (3.4) depends on the distribution of the error term in (3.1). The next two sections discuss two specifications, leading to the Multinomial Logit and Multinomial Probit model respectively.

3.2 Logit

3.2.1 The MNL Model

The Multinomial Logit (MNL) model follows when the assumption is made that the error term in (3.1), \( e_j \), is independently and identically distributed with a Weibull density function. A Weibull density function for a random variable \( Y \) is defined as (see, e.g., McFadden 1976):

\[ P(Y \leq y) = \exp^{-e^y} . \]  

(3.5)

This distribution belongs to the class of double negative exponential distributions as are, e.g., the Type I extreme value distribution and the Gumbell distribution, which are sometimes also used to specify the MNL.

---

4 The Multinomial Logit model for conjoint choice experiments is described in section 4.2, the Multinomial Probit model for conjoint choice experiments in section 5.2. In this chapter the MNL and MNP model are described in the standard context with only one choice set for each respondent.
model. In the MNL model the choice probabilities have a simple closed form (see, e.g., Maddala 1983, p. 60-61; Ben-Akiva and Lerman 1985; Swait and Louviere 1993):

\[ p_{jm} = \frac{\exp(X_{jm}\beta \mu)}{\sum_{n=1}^{M} \exp(X_{jn}\beta \mu)}, \tag{3.6} \]

where \( \mu \) is the scale parameter of the MNL model. This parameter scales the variance of the MNL model, but since it cannot be identified it is often set to one. However, when data sets are compared on their MNL results, scale differences between these data sets may give a misleading results and the ratio of the scale parameters of these two data sets has to be estimated. See for an extensive discussion on the scale parameter in the MNL model Swait and Louviere (1993). Since in this thesis no data sets are compared using MNL estimates, this parameter is omitted here for convenience from here on (i.e., it is set to one). The simple and easy to calculate form of the choice probabilities (3.6) in the MNL model has much contributed to its popularity. However, there is a serious limitation to the use of this model which is related to the Independence of Irrelevant Alternatives property.

### 3.2.2 The IIA Property

The *Independence of Irrelevant Alternatives* (IIA) property, which arises from the assumption of independent random errors and equal variances for the choice alternatives, implies that the odds of choosing one alternative over another alternative must be constant regardless of whatever other alternatives are present (e.g., Louviere and Woodworth 1983; Ben-Akiva and Lerman 1985), which may often be too restrictive in practical situations. To illustrate the possible implications of the IIA property, take the famous example that an individual can choose between two kinds of transportation, driving by bus or by car, and suppose that he chooses the car in two-thirds of the cases. Now suppose that a third alternative is available which is also a bus, but in a different color. In this situation one would still
expect that the car is chosen in two-thirds of the cases and the probability for the bus is divided over the two buses with different colors. Because of the IIA property in the Logit model, however, the probability of choosing the car falls to a half in order to keep the odds two-to-one in favor for the car for either bus. So, with one car and one bus, \( P(\text{car}) = \frac{2}{3} \) and \( P(\text{bus}) = \frac{1}{3} \). When an additional bus is present one expects that \( P(\text{car}) \) remains \( \frac{2}{3} \) and \( P(\text{bus-A}) = P(\text{bus-B}) = \frac{1}{6} \), but in the Logit model it holds that \( P(\text{car}) \) drops to \( \frac{1}{2} \) and \( P(\text{bus-A}) = P(\text{bus-B}) = \frac{1}{4} \).

If it is assumed that the IIA property holds and the MNL model is used, predicting the choice probabilities of new alternatives can simply be done by inserting the attribute values of these new alternatives in the closed form expressions for the choice probabilities (3.6). Green and Srinivasan (1978) stated that in consumer behavior contexts the IIA property may not be a realistic assumption, especially when some of the alternatives are close substitutes (cf. McFadden 1976). When covariances across alternatives are incorrectly assumed to be zero, the estimates for the effects of explanatory variables are inconsistent (Hausman and Wise 1978; Chintagunta 1992).

McFadden (1986) stated that there are three ways to deal with the IIA property. First, statistical tests can be performed to test whether IIA holds in a particular application (e.g., Louviere and Woodworth 1983; Hausman and McFadden 1984). When these tests show that IIA holds the MNL model can be used. One way to test the IIA assumption is by estimating (two-way) interaction effects. When those are all nonsignificant, IIA holds (see, e.g., Louviere and Woodworth 1983; McFadden 1986). Higher order interactions are also possible, but Louviere and Woodworth (1983) stated that the estimation of the two-way interactions should be sufficient to test for IIA violations. Second, when the scale value for an alternative is allowed to depend on features of the choice set, the IIA property can be avoided in the MNL model, however, at the costs of losing some structural simplicity (McFadden 1986). Third, when the IIA property does not hold, other models which avoid IIA, should be used instead of the standard MNL model, however, at the cost of computational complexity. For a discussion
on several of these models see, e.g., Wedel et al. (1999). One of the most general of these models is the Multinomial Probit model, which is discussed in the next section.

3.3 Probit

3.3.1 The MNP Model

We start again with the utility function in the situation that there is one choice set. The utilities derived from the $M$ choice alternatives to individual $j$ are contained in the $M$-vector $U_j$, which is assumed to satisfy:

$$U_j = X\beta + e_j,$$  \hspace{1cm} (3.7)

where $X$ is an $(M \times S)$-matrix containing the characteristics of the alternatives, $\beta$ is the $(S \times 1)$ weight vector, and $e_j$ is the vector containing the random component of the utilities. It is assumed for convenience that $X$ does not depend on $j$. In the Multinomial Probit (MNP) model it is assumed that the error term has a Normal density:

$$e_j \sim N_M(0, \Sigma_e),$$ \hspace{1cm} (3.8)

independent between individuals; $\Sigma_e$ is an $(M \times M)$ positive definite covariance matrix, which allows choice alternatives to be correlated. The utility vector of individual $j$ is distributed as:

$$U_j \sim N_M(X\beta, \Sigma_e).$$ \hspace{1cm} (3.9)

The probability that alternative $m$ is chosen can be expressed as a function of the model parameters as follows. Let
\[
A_m = \begin{bmatrix}
1 & -1 \\
\vdots & \vdots \\
1 & -1 \\
-1 & 1 \\
\vdots & \vdots \\
-1 & 1
\end{bmatrix} m
\]

\begin{equation}
\begin{aligned}
1 & \quad m-1 \\
1 & \quad m-1 \\
-1 & \quad m+1 \\
1 & \quad M
\end{aligned}
\end{equation}

So,

\[
\tilde{u}_{jm} = A_m^\prime u_j - N_{M-1}(A_m^\prime X\beta, A_m^\prime \Sigma e A_m) .
\]

The variable $\tilde{u}_{jm}$ gives the differences in utility with respect to the $m$th alternative. The density of $\tilde{u}_{jm}$ is $d_m(.)$ and has the following form in the MNP model:

\[
d_m(t) = (2\pi)^{\frac{1}{2}(M-1)} |A_m^\prime \Sigma e A_m|^{-\frac{1}{2}} e^{-\frac{1}{2}(t-A_m^\prime X\beta)/(A_m^\prime \Sigma e A_m)^{-1}(t-A_m^\prime X\beta)} ,
\]

which does not depend on $j$ since $X$ does not depend on $j$. The argument $t$ of $d_m(.)$ is of dimension $(M-1)$. The assumption that the alternative with the highest utility is chosen implies that the probability of choosing $m$, $p_m$, equals:

\[
p_m = P(\tilde{u}_{jm} \leq 0) = \int_{-\infty}^{0} d_m(t) \, dt ,
\]

where the integral is of dimension $(M-1)$. ML estimates are obtained by maximization of the log-likelihood (3.4) over $\beta$ and the parameters in $\Sigma_e$. When the dimension of the integral is greater than three, the probabilities $p_m$ cannot be evaluated numerically (e.g., McFadden 1976; Maddala 1983; Kamakura 1989; Keane 1992). However, some structure can be imposed on the error terms in the model, for instance, assume that all covariances between alternatives are zero and all variances are equal to one. This leads to the Independent Probit (IP) model (Hausman and Wise 1978) which can be used even when there are many alternatives. However, the IP model is
closely related to the MNL model and also suffers from the IIA property (e.g., Amemiya 1981; Ben-Akiva and Lerman 1985). There is a fixed relation between the variance of the MNL and IP model where it holds that the variance of the MNL model is a factor $\pi^2/6$ times that in the IP model (Ben-Akiva and Lerman 1985, p. 105). This implies that both models give the same estimation results where the estimates in the MNL and IP model differ a factor $\sqrt{\pi^2/6}$. Because the MNL and MNP models are not nested, they cannot be compared with a Likelihood Ratio type test. However, because of the above relation between MNL and IP, the Logit and Probit models can be compared indirectly using the IP model that is nested within the general MNP model. This (indirect) way to compare MNL and MNP is also done in the applications of chapters 5 and 6. When the covariances are zero but the variances differ in the IP model, IIA is alleviated, but such a model does not allow choice alternatives to be correlated. Simulation techniques have to be used to approximate the integrals when a general Probit model is used.

3.3.2 Simulation of Probit Probabilities

There have been a few attempts to overcome the problem that when there are more than three choice alternatives, the integral in (3.13) cannot be evaluated anymore. An early method to approximate the highly-dimensional integrals is known as Clark’s approximation (Clark 1961; Daganzo 1979). However, this method turned out not to perform very well (Horowitz, Sparmann and Daganzo 1982; Langdon 1984) especially in the case of negative correlations and/or unequal covariances (McFadden 1989), with increasing number of alternatives and furthermore, there is no associated asymptotic theory with the method (Chintagunta 1992).

The first idea to use simulations to evaluate (3.13) was by Lerman and Manski (1981), who used Monte Carlo draws of the error term to simulate the latent vector of utilities $U$. The frequency that $u_{jm} = \max(u_{j1}, \ldots, u_{jM})$ was used to approximate $p_{jm}$. This simulator was later called the Crude Frequency Simulator (CFS) (e.g., Hajivassiliou 1993). This simulator
Chapter 3

needs many replications (draws) of the error term before the empirical frequencies are good approximations for the probabilities $p_{jm}$, in particular when $p_{jm}$ is close to 0 or 1. Another disadvantage of this simulator is that it can only take discrete values. Since Lerman and Manski (1981) developed their simulator, several other simulators for MNP choice probabilities have been developed. These simulators differ basically as to the way the draws from the error distribution are obtained. Hajivassiliou, McFadden and Ruud (1993) compared thirteen simulators and concluded that the Smooth Recursive Conditioning (SRC) simulator is one of the best. This simulator is also known as the GHK-simulator (after Geweke, Hajivassiliou and Keane). It has the best (or among the best) mean and median bias, standard deviations, and RMSE (Root Mean Squared Error) characteristics of all considered simulators. The SRC simulator is a so-called smooth simulator, which means that small changes in the parameters do not result in jumps in the simulated probabilities or in the likelihood, i.e., the likelihood is continuous in the parameters. Furthermore, the simulated probabilities do not lie on the boundaries of the unit interval [0, 1]. A potential drawback of using smooth simulators is that the simulated probabilities are not restricted to add up to one over the $M$ choices (McFadden 1989; Mühleisen 1991; Lee 1992). Lee (1992) stated that the adding-up property can always be satisfied by normalizing the original simulators, at the extra cost of simulating choice probabilities for all alternatives. Because the SRC simulator is used in the applications of chapters 5 and 6 to simulate the choice probabilities in Conjoint MNP models, the next sub-section describes in detail how the SRC simulator works.

The SRC Probability Simulator

This sub-section is based on Börsch-Supan and Hajivassiliou (1993), see also, e.g., Hajivassiliou (1993) or Geweke, Keane and Runkle (1994). We start with the following discrete choice model:
Choice Models

\[ Y^* = X\beta + \epsilon = \mu + \Gamma e, \quad \text{with} \]
\[ Y^* \sim N_M(\mu, \Omega), \quad \Omega = \Gamma \Gamma', \]
\[ e \sim N_M(0, I), \quad \epsilon = \Gamma e. \quad (3.14) \]

Observed is the \((Mx1)\)-vector \(Y\), where

\[ y_m = \begin{cases} 1 & \text{if } y^*_m \geq y^*_n \quad \forall \ n = 1, \ldots, M, \ n \neq m \\ 0 & \text{otherwise} \end{cases} \quad (3.15) \]

The probability \(P(y_m=1)\) involves an \((M-1)\)-dimensional integral which cannot be evaluated for \(M>4\). The SRC probability simulator simulates the probability \(\pi_m = P(y_m=1)\) as follows:

Let \(V\) (which is equal to \(A'/m(3.10)\)) be a non-singular matrix and define \(L\) as the lower Choleski-factor of \(V \Omega V' = LL'\). Then from (3.14) and (3.15) it follows that for all \(n \neq m\) it must hold that:

\[ -\infty \leq y^*_n - y^*_m \leq 0 \]
\[ \Rightarrow -\infty \leq VY^* \leq 0 \]
\[ \Rightarrow -\infty \leq V(X\beta + \epsilon) \leq 0 \]
\[ \Rightarrow -\infty \leq V(\mu + \Gamma e) \leq 0 \]
\[ \Rightarrow -\infty \leq V\mu \leq Le \leq 0 - V\mu \]
\[ \Rightarrow a \leq Le \leq b \quad (3.16) \]

Since \(V \Gamma = L\), random draws are needed such that:

\[ e_1 \sim N(0,1) \ s.t. \quad a_1 \leq l_{11}e_1 \leq b_1 \]
\[ e_2 \sim N(0,1) \ s.t. \quad a_2 \leq l_{21}e_1 + l_{22}e_2 \leq b_2 \]
\[ : \]
\[ e_M \sim N(0,1) \ s.t. \quad a_M \leq l_{M1}e_1 + \cdots + l_{MM}e_M \leq b_M \]
\[ \Rightarrow a_M \leq L_{M,<M}e_M + l_{MM}e_M \leq b_M \quad . \quad (3.17) \]

These \(e_1, \ldots, e_M\) is a series of uncorrelated, truncated standard normal random variates and these draws can be obtained by drawing a vector \(U \sim \)
Uniform(0,1) and by applying to this vector $U$ the inverse distribution function method:

$$e = \Phi^{-1} \left[ (\Phi(b^*) - \Phi(a^*)) U + \Phi(a^*) \right], \quad (3.18)$$

where $\Phi$ is the normal cumulative distribution function. Then it holds that $e$ follows a truncated normal distribution (Hajivassiliou and McFadden 1990):

$$e \sim N(0,1) \text{ s.t. } a^* \leq e \leq b^*, \text{ with } -\infty \leq a^* < b^* \leq \infty. \quad (3.19)$$

Now define $Q_m (m=1, \ldots, M)$ as:

$$Q_1 = P \left( \frac{a_1}{l_{11}} \leq e_1 \leq \frac{b_1}{l_{11}} \right)$$

$$\vdots$$

$$Q_m (e_1, \ldots, e_{m-1}) = P \left( \frac{a_m - L_{m,m} e_{m-1}}{l_{mm}} \leq e_m \leq \frac{b_m - L_{m,m} e_{m-1}}{l_{mm}} \right), \quad (3.20)$$

where $L_{m,m} e_{m-1} = l_{m1} e_1 + \cdots + l_{m(m-1)} e_{m-1}$ and where $P(\cdot)$ represents a probability. The SRC probability simulator for $\pi_m$ then is for random draws $e_r = (e_{1r}, \ldots, e_{Mr})$, and with $R$ replications ($r=1, \ldots, R$), defined as:

$$\pi_m^{SRC} = f_m(\beta, \Omega) = \frac{1}{R} \sum_{r=1}^{R} \prod_{m=1}^{M} Q_m (e_{1r}, \ldots, e_{m-1,r}) \quad (3.21)$$

The simulator is unbiased and smooth. That is, $\pi_m^{SRC}$ is a continuous and differentiable function of $\beta$ and the parameters in $\Omega$.

### 3.3.3 Simulation Techniques

The breakthrough in Probit estimation came with McFadden’s (1989) article, simultaneously with Pakes and Pollard (1989), on the Method of Simulated Moments (MSM). McFadden showed that when an unbiased simulator is used to simulate the choice probabilities, the functions to be simulated appear linearly in the conditions defining the estimator, and the
same set of random draws is used to simulate the model at different trial parameter values in the process of iteratively searching for the estimator, only a small number of random draws are necessary to approximate the choice probabilities. Since then other methods have been developed including the Method of Simulated Scores (MSS) (Hajivassiliou and McFadden 1990) and Simulated ML (SML). Hajivassiliou (1993) gave an extensive overview of these and other simulation techniques. The next two sub-sections describe these simulation techniques. SML is the technique that will be used to obtain the Probit estimates in the applications of chapters 5 and 6.

The Method of Simulated Moments
In (3.4) the log-likelihood of a multinomial choice model was given, which is repeated in (3.22):

$$l = \sum_{j=1}^{J} \sum_{m=1}^{M} y_{jm} \ln (p_{jm}) .$$  \hfill (3.22)

In order to estimate the parameters in an MNP model, in the ML context, the first order derivative of the (log-) likelihood must be calculated. Assume that all parameters are contained in the ($T \times 1$) vector $\theta$, then estimates for $\theta$ are obtained by solving $\theta$ from:

$$\frac{\partial \ln l}{\partial \theta} = \sum_{j=1}^{J} \sum_{m=1}^{M} y_{jm} \frac{\partial \ln p_{jm}}{\partial \theta} = \sum_{j=1}^{J} \sum_{m=1}^{M} y_{jm} \frac{\partial p_{jm}}{\partial \theta} = 0 .$$  \hfill (3.23)

Note that $p$ depends on $\theta$, but this has been suppressed for convenience. Because the sum of the probabilities is equal to one, it holds that:

$$\sum_{m=1}^{M} \frac{\partial p_{jm}}{\partial \theta} = 0 .$$  \hfill (3.24)

From (3.24) it follows that (3.23) can be written as:
Now define $W$ as:

$$W = \frac{1}{p_{jm}} \frac{\partial p_{jm}}{\partial \theta} \quad (3.26)$$

then (3.25) can be written in matrix notation as:

$$W(\theta)'(Y - P(\theta)) = 0 \quad (3.27)$$

where $W(\theta)$ is a $(JM \times T)$ matrix, $P(\theta)$ is the $(JM \times 1)$ vector of choice probabilities, and $Y$ is the vector of observed choices. Equation (3.27) can be interpreted as the moments equations of the MNP model. Solving $\theta$ from (3.27) is equivalent to the minimization with respect to $\theta$ of:

$$(Y - P(\theta))'W(\theta)W(\theta)'(Y - P(\theta)) \quad (3.28)$$

Because for the MNP model the probabilities $P(\theta)$ in (3.28) are multidimensional integrals (see (3.13)), they are for more than a few alternatives practically impossible to calculate. In the method of simulated moments, $P(\theta)$ is replaced by a vector of unbiased simulated probabilities $F(\theta)$. So, with MSM, the parameter vector $\theta$ is estimated by minimizing:

$$(Y - F(\theta))'W(\theta)W(\theta)'(Y - F(\theta)) \quad (3.29)$$

The estimators for $\theta$ calculated in this way are consistent and asymptotically normal (CAN), when the simulators $F(\theta)$ are unbiased, the random numbers for the construction of $F(\theta)$ are independent of the random numbers for the construction of $W(\theta)$, $F(\theta)$ and $W(\theta)$ are smooth simulators, and the random numbers are not redrawn during the iteration process (McFadden 1989; Mühleisen 1991). The estimators are efficient when optimal instruments $W(\theta)$ are used (see below). The asymptotic covariance matrix is equal to (McFadden 1989):

$$\sum_{j=1}^{J} \sum_{m=1}^{M} \frac{y_{jm}}{p_{jm}} \frac{\partial p_{jm}}{\partial \theta} = \sum_{j=1}^{J} \sum_{m=1}^{M} (y_{jm} - p_{jm}) \frac{1}{p_{jm}} \frac{\partial p_{jm}}{\partial \theta} = 0 \quad (3.25)$$
\[ \Sigma_\theta = (D'D)^{-1}D'GD(D'D)^{-1}, \quad (3.30) \]

and can consistently be estimated by using (Mühleisen 1991):

\[
\hat{G} = \frac{1}{J} \sum_{j=1}^{J} \sum_{m,n=1}^{M} W_{jm}(y_{jm} - f_{jm})(y_{jn} - f_{jn}) W_{jn},
\]

\[
\hat{D} = \frac{1}{J} W' \frac{\partial P(\hat{\theta}_{msm})}{\partial \hat{\theta}_{msm}}.
\]

With the method of simulated scores, instead of evaluating (3.29), the parameter vector \( \hat{\theta} \) is estimated by solving \( \hat{\theta} \) from:

\[
\frac{1}{J} \sum_{j} s_{j}(\theta) = \frac{1}{J} \frac{\partial \ln l}{\partial \theta} = 0,
\]

where the score, \( s_{j}(\theta) \), is directly simulated. This method also yields CAN estimates, when the SRC simulator is used.

The MSM of McFadden has the drawbacks that all probabilities in a particular model have to be simulated and that the estimators are only efficient when optimal instruments depending on the true parameter values, which are unknown, are used (McFadden 1989; Hajivassiliou 1993). One way out of the problem with MSM is to run the method several times (McFadden 1989). In this case a start vector of parameters \( \theta_0 \) is used to calculate \( W(\theta_0) \). With this instruments matrix the parameters are estimated, then the estimated parameters are used to calculate a new \( W(\theta) \), and so on. It is essential for the asymptotic statistics of the MSM estimator that the simulators of the response probabilities and their derivatives used to construct the instruments must be independent of the simulator \( F(\hat{\theta}) \) used in (3.29) (McFadden 1989). The procedure described here should be done at least 2-3 times, thereby increasing the total computing time considerably. With MSS no instruments are needed, but Mühleisen (1991) compared MSS using SRC as the simulation method with MSM using Importance Sampling (IS) as the simulation method, which is a generalization of a frequency simulator such as CFS. The findings of Mühleisen were rather disappointing. For the model he used, the MSS
method never resulted in estimates because of non-convergence, and the MSM method resulted only in a few cases in estimates. He also reported that other authors have had problems using MSS and MSM when gradient methods are used for the iterative estimation process. Hajivassiliou (1993) also reports speed problems when using MSS.

The Method of Simulated Likelihood

With the method of Simulated Maximum Likelihood (SML), only the probabilities of the selected alternatives have to be simulated which is computational more efficient. This method is known as Smooth SML (SSML) when it is applied with a smooth choice simulator as the SRC simulator. With SSML, asymptotical efficiency requires that $R/J \rightarrow \infty$ where $R$ is the number of replications. However, several studies show that SSML is efficient even when the number of replications is rather low, say 10 to 20 (Mühleisen 1991; Lee 1992; Börsch-Supan and Hajivassiliou 1993; Geweke, Keane and Runkle 1994). The simulated probabilities replace the probabilities $p_{jm}$ in the likelihood of (3.4), which is very straightforward and with this method one stays therefore within the standard ML context which is very often used for estimating discrete choice models. So, all asymptotic ML properties are still valid for the estimates for $\theta$ in the SSML context. The log-likelihood now is equal to:

$$l = \sum_{j=1}^{J} \sum_{m=1}^{M} y_{jm} \ln (f_{jm}),$$

(3.33)

where the $f_{jm}$ are the simulated choice probabilities. As noted above, the advantage of SSML over MSM with respect to the number of probabilities to simulate disappears when Lee’s (1992) approach has to be used to ensure that the probabilities sum up to one over all alternatives in a choice set.
3.3.4 Identification

A potential problem of the MNP model is that it is difficult to identify the parameters in the model. Identification means that there is only one set of estimates that maximizes the likelihood. When different parameter estimates give the same results, interpretation of the estimates becomes difficult. Bunch and Kitamura (1989) demonstrated that nearly half of the published applications of MNP are based on non-identified models (see also the next section). It is easy to see that when the covariance matrix (3.8) of the (unrestricted) MNP model is multiplied with a factor $\alpha$ and all $\beta$-estimates with a factor $\sqrt{\alpha}$, that this leads to the same results. So, at least one parameter in the MNP model must be fixed to scale the model and to identify the other parameters. Often, one of the variance parameters is used for this purpose, but this is not sufficient, however. In the MNP model, only $M(M-1)/2 - 1$ of the $M(M+1)/2$ covariance parameters in $\Omega$ are identified (Dansie 1985; Bunch 1991; Keane 1992). So, $M+1$ restrictions must be imposed on the $\Omega$-matrix. One solution is to set all covariance parameters of one alternative equal to zero. Also, the utility of one alternative can freely be set to zero, because in discrete choice models only differences in utilities are important and not absolute values (Keane 1992). In a general (not conjoint) MNP context these identification rules are often fulfilled by fixing one (variance) parameter in the differenced covariance matrix $\Omega_m = A_m' \Sigma_x A_m$. However, even then the MNP model is not identified in many situations. Heckman and Sedlacek (1985) stated that a necessary condition for identification of the general MNP model is that there is at least one single regressor that varies across individuals. Keane (1992) showed that even when MNP models are formally identified, identification can still be “fragile” in the absence of so-called exclusion restrictions, implying that the exogenous variables in the model should not be included in the utility levels of all alternatives. To avoid identification problems, it is necessary to have one exclusion from each utility index (Keane 1992). In other words, in each row of the $X$-matrix of exogenous
variables there should be a zero. Additional identification issues for conjoint choice experiments are discussed in section 5.2.2.

3.3.5 Marketing Applications

In the marketing literature several overview articles have appeared in the past years describing choice theory. They all agree that the IIA property of the MNL model may be a too restrictive assumption in (marketing) applications (McFadden 1986; Batsell and Louviere 1991; McFadden 1991; Carson et al. 1994; Horowitz et al. 1994; Ben-Akiva et al. 1997; Keane 1997). As stated before, the main reason why MNL is used very often in (marketing) applications is its computational simplicity while non-IIA models, like the general MNP model, are much more difficult to apply. However, several applications of the MNP model have occurred in the marketing literature, some of which made use of simulation techniques. Several of these papers are described briefly below.

One of the first marketing applications of the MNP model was by Currim (1982). He developed a generalization of the Hausman and Wise (1978) conditional Probit model. He showed that the Probit model can approximate other non-IIA models, like the negative exponential and external value models. In his travel mode choice application he used five choice alternatives and used the Clark approximation to obtain choice probabilities. The conclusion of this study was that the generalized Probit model outperforms the IIA models on market share predictions. However, as Bunch and Kitamura (1989) pointed out, the model used by Currim seems not identified. In addition, the reported likelihood for the IP model is better then for the generalized Probit model, which should not be the case because these models are nested.

Kamakura and Srivastava (1984) developed an MNP model with a covariance matrix constructed from a Euclidean distance similarity measure between choice alternatives. They also used the Clark approximation to obtain the choice probabilities. In their application, they used choice sets with three choice alternatives and they compared their
model with the IP model and with the Hausman and Wise (1978) model. Their proposed model outperforms the other two models with respect to estimation and predictive fit, although the difference with the Hausman and Wise model was not big in both situations.

A dynamic Probit model was used by Papatla and Krishnamurthi (1992) who used household panel data on a frequently purchased product class with five choice alternatives (brands). They also used the Clark approximation and estimated a Covariance Probit and IP model. Their Probit model outperformed the IP model on model fit and on several predictive fit measures, although the IP model was better on other predictive fit measures.

Chintagunta (1992) was the first to use a simulation method for a Probit model in marketing. He also estimated a scanner panel data set and used the Method of Simulated Moments to obtain choice probabilities in the Probit model. The results from the MSM simulation (using 20 draws and choice sets with three alternatives) were compared in a simulation study with those from the Clark approximation. Furthermore, the MSM results were compared with the IP and Logit results. The conclusion was that MSM performed better as the Clark approximation and also better as the IP and Logit model. In an application with choice sets containing six choice alternatives (brands) MSM was compared with two versions of the Logit model. Because MSM is not an ML method, no comparisons could be made with respect to the likelihood values of the various models, but an investigation of the results let Chintagunta to conclude that: “Our results indicate that ignoring similarities among choice alternatives while estimating the effects of marketing variables can have serious consequences for the accuracy of estimates for these variables.”

In another panel data application of the Probit model, Elrod and Keane (1995) used the SRC/GHK simulator to obtain the MNP choice probabilities in the MSM context. They developed a factor-analytic structure for their covariance matrix. They had eight choice alternatives (brands) and found that the factor-analytic Probit model outperformed
alternative market structure and brand choice models, in terms of goodness of fit, predictive validity, and face validity.

The above studies showed that a Probit specification on several marketing choice models, although using very different structures for the covariance matrix in the MNP model, performed better with respect to model fit and predictions compared to IIA-type models, like the MNL model. The MNP model has not yet been applied to conjoint choice experiments. In chapter 5 we will develop such an MNP model for conjoint choice models. Before this model is developed, the next chapter investigates how the no-choice alternative should be included in a conjoint choice design matrix.