# Chapter 4
# Conjoint MNL Models and the "No-Choice" Alternative[5]

## 4.1 Introduction

In conjoint choice experiments respondents choose one profile from each of several choice sets. In order to make the choice more realistic, in many conjoint experiments one of the alternatives in the choice sets is a "no-choice" or "none" option. This option can entail a real no-choice alternative ("None of the above") or an "own-choice" alternative ("I keep my own product"). This base alternative, however, presents the problems of how to include it in the design of the choice experiment, and in what way to accommodate it in the choice model. Regular choice alternatives are most often coded in the design matrix with effects-type or dummy coding. Since the no-choice alternative does not possess any of the attributes in the design, it is often coded simply as a series of zero's, which makes the fixed part of its utility zero in each choice set.

In this chapter we investigate several models that can be used to accommodate the no-choice option and we show that it is best to include a constant in the design matrix of the conjoint experiment for the no-choice alternative when effects-type coding and/or linear coding is used for all attributes. Not including this constant leads to reduced fit and biased estimates of linear attributes. In section 4.2 we first extend the general MNL choice model of chapter 3 to the conjoint situation with more than one choice set for each respondent. In section 4.3 we give some background on the no-choice alternative and we describe several models that potentially can capture the effects of the presence of a no-choice option. Section 4.4 gives the results of an application and section 4.5 the

---

[5] This chapter is based on Haaijer, Johnson, Kamakura and Wedel (1999).

conclusions and discussion. In this chapter we do not consider the MNP model but only use the much simpler MNL model which suffices to make our point.

## 4.2 Conjoint Choice MNL Models

In chapter 3, the general discrete choice model was introduced with one choice observation for each respondent. In a conjoint choice model, however, each respondent has to choose one alternative from each of several choice sets. These choice sets are constructed, as explained in chapter 2, by dividing the total set of profiles over $K$ choice sets. In this and following chapters we assume that each choice set contains the same number of alternatives, without losing generality.

The utility of alternative $m$ in choice set $k$ for individual $j$ is defined as:

$$u_{jkm} = X_{km}\beta + e_{jkm} , \tag{4.1}$$

where $X_{km}$ is a (1x$S$) vector of variables representing characteristics of the $m$th choice alternative in choice set $k$. In most conjoint choice experiments no individual characteristics are present, so $X$ does not depend on $j$. Note, however, that when a individualized design is used, $X$ does depend on $j$, but we omit this index here for convenience. $\beta$ is a ($S$x1) vector of unknown parameters, and $e_{jkm}$ is the error term. The MNL model treats observations coming from the same respondent as independent observations, and falls within the standard random utility approach of chapter 3. In the MNL model 100 respondents choosing from 10 choice sets is therefore computational equal to 1000 respondents choosing from 1 choice set. The MNP model for conjoint choice experiments, with no independence of choice observations from the same respondent, is introduced in chapter 5. The choice probabilities in the conjoint MNL approach can be obtained in a straightforward generalization of formula (3.6). The probability $p_{km}$ that alternative $m$ is chosen from set $k$ is simply equal to:

$$p_{km} = \frac{\exp(X_{km}\beta)}{\sum_{n=1}^{M} \exp(X_{kn}\beta)} \quad . \tag{4.2}$$

The log-likelihood (3.4) in the conjoint context is extended by adding a sum over choice sets:

$$l = \sum_{j=1}^{J} \sum_{k=1}^{K} \sum_{m=1}^{M} y_{jkm} \ln(p_{km}) \quad . \tag{4.3}$$

When the IIA assumption is true, the parameters of the Logit model can be estimated when the sufficient condition is satisfied that the alternatives are independent across choice sets (Louviere and Woodworth 1983). This implies that alternatives must be pairwise independent across choice sets. Such alternatives in a conjoint choice experiment can be obtained by using an orthogonal, fractional factorial main effects design (Louviere and Woodworth 1983; Louviere and Timmermans 1990). A constant base alternative is useful, because it preserves the design orthogonality of the attribute vectors of conjoint alternatives (Louviere 1988; Elrod, Louviere and Davey 1992). However, in the case of the Logit model, design orthogonality does not imply information orthogonality, for which the parameters would be uncorrelated. When similarities across alternatives are incorrectly assumed to be zero, the estimates for the effects of (marketing) variables are incorrect (e.g., Chintagunta 1992).

The expression for the choice probabilities (4.2) may be expanded to accommodate ranking data, which is particularly useful in conjoint analysis (McFadden 1986; Kamakura, Wedel and Agrawal 1994). However, the assumptions needed to translate rankings into choices need not hold in practice, especially when individuals use elimination and nesting strategies, the IIA property does not hold (Louviere 1988). Also, the use of brand names in the conjoint design may result in correlations between the utilities of the alternatives, violating the IIA property. In order to be able to test for

IIA, design plans which allow as many relevant two-way interaction effects as possible to be tested can be used (Louviere and Woodworth 1983).

## 4.3 The Base Alternative

### 4.3.1 General Elements

In conjoint choice experiments a base alternative is included in the design of the experiment, among others, to scale the utilities between the various choice sets. A base alternative can be specified in several ways. First, it can be a regular profile that is held constant over all choice sets. Second, it can be specified as “your current brand” and third, as a “none”, “other” or “no-choice” alternative (e.g., Louviere and Woodworth 1983, Batsell and Louviere 1991, Carson et al. 1994). Additional advantages of including a “no-choice” or “own” base alternative are that it may make the choice decision more realistic and may lead to better predictions of market penetrations. A disadvantage of a no-choice alternative is that it may lead respondents to avoid difficult choices, which detracts from the validity to the use of the no-choice probability to estimate market shares. However, Johnson and Orme (1996) claim that this seems not to happen in conjoint choice experiments. In addition, the no-choice alternative gives no information about preferences for attributes of the choice alternatives, which is the main reason for doing a conjoint choice experiment. Furthermore, when the choice model is estimated with an IIA choice model (such as the MNL model) the presence of a no-choice may violate this IIA-assumption, because the utilities of the choice alternatives may be correlated. Dhar (1997) gave an overview of why and when respondents may choose a no-choice option in general. He stated that respondents may choose the no-choice when none of the alternatives appears to be attractive, or when the decision maker expects to find better alternatives by continuing to search. Furthermore, when subjects are uncertain about the range of potential alternatives they may continue to look for better alternatives and choose the "no-choice" in the early stages of a choice process. Dhar (1997) also showed that adding an attractive alternative to an already attractive choice set increases the preference of the no-choice option and adding an unattractive alternative to that choice set decreases the preference of the no-choice. This implies that when alternatives are close to each other in preference people will choose the no-choice more often compared to when there is a clearly dominant or unattractive profile in the choice set, which violates the IIA property. Huber and Zwerina (1996) stated that conjoint choice sets that are utility balanced are more informative and efficient compared to sets with dominant alternatives. However, according to the above, adding a no-choice alternative to such a choice set may influence this negatively, because people may start avoiding to make the difficult choice. This is related with findings in the psychological literature, where it has been found that people prefer consequences that arise of inaction over those arising from action since the decision to stay within a status quo has certain psychological advantages (Baron and Ritov 1994, Dhar 1997). This may make people choose "none" if there is no dominant alternative present in the choice set. The above shows that the reasons to choose the no-choice may be different from choosing any of the other "real" profiles in a conjoint choice experiment. In this view, "no-choice" cannot be seen as just another choice alternative, leading to potential violations of IIA, amongst others.

In this chapter we investigate the no-choice option from a modeling point of view[6]. We start by discussing a number of alternative model formulations. First, simply having a series of zero's describing the attribute values of the no-choice alternative seems a straightforward option, but this formulation may produce misleading results. When "no-choice" is coded as a series of zeros, its fixed part of utility is equal to zero by definition.

[6] In the reminder of the chapter we only mention the "no-choice" (or "none"), but the results also apply to the "own" or "other" base alternative when nothing is known about their characteristics.

However, when there are linear attributes present, these zero values of the "no-choice" alternative act as "real" levels of the linear attributes. When for instance "price" is a linear attribute in the design, the zero value for no-choice will correspond to a zero price. We hypothesize that this can lead to a biased estimate of the parameter of the linear attribute.

Second, when all attributes are modeled with effects-type coding the bias discussed above does not arise, because all part-worths are now specified relative to the zero-utility of the no-choice alternative. However, even when all attributes are coded with effects-type dummies, adding such a constant may improve model fit. This can be explained because the no-choice option in fact adds one level to all attributes. If there are, for example, $S$ attributes each with $L$ levels, then there are $L^S$ possible alternatives and $L$ part-worths to be estimates for each of the $S$ attributes in the analysis. The inclusion of a no-choice option would produce a design with $S$ attributes each with $L+1$ levels. However, the resulting full $(L+1)^S$ set of profiles cannot be constructed, since the $(L+1)th$ level of all attributes is confounded with the no-choice alternative. So, instead of $(L+1)^S$ alternatives only $L^S+1$ alternatives are possible. This is accommodated in the MNL analysis by estimating the $L$ part-worths (through $L-1$ effects-type dummies) for each of the $S$ attributes plus one part-worth for the no-choice option. Thus adding the no-choice constant is hypothesized to increase model fit[7]. This change in the design of the choice experiment is hypothesized to reduce the potential bias in the estimates for linear attributes. Although this additional constant increases the number of parameters by one, it sets the utility level of the no-choice alternative, as explained above, and therefore compensates for the bias in the estimates for linear attributes in the design.

---

[7] Note that the same can be accomplished by coding one of the attributes (e.g. Brand) with dummy coding instead of effects-type coding. This also increases the number of parameters with one and sets the level of the no-choice. Another possibility is to add a constant for the "real" alternatives instead and keep a series of zero's for the no-choice, in that case the constant is equal to $-c_{nc}$.

Finally, another way to model the presence of a no-choice option is by specifying a Nested Logit model. When two nests are specified, one containing the no-choice and the other the "real" profiles, the no-choice alternative is no longer treated as "just another alternative". The idea is that respondents first decide to choose a real profile or not and only when they decide to choose a real profile they select one of them, leading to a nested choice decision. This way of modeling the no-choice potentially also removes the effects of linear attributes because the zeros of the no-choice are no longer treated as real levels, because they are now captured in a different nest. The Nested Logit model can also be tested for the situation with or without the presence of linear attributes next to effects-type coded attributes.

### 4.3.2 No-Choice Models

We use three models, (1) the Multinomial Logit model (MNL), (2) the Nested Multinomial Logit model (NMNL), and (3) the No-choice Multinomial Logit model. The difference between the Logit model and the No-choice Logit model is the extra constant ($c_{nc}$) added for the no-choice option in the design, but both models fall within the standard Multinomial Logit context for conjoint experiments described in section 4.2. In the Nested Logit model there is one extra parameter ($\lambda$) called the *dissimilarity coefficient* (Börsch-Supan 1990). When its value is equal to 1, the Logit and Nested Logit model are equal. In the Nested Logit model we assume that there are two nests, one containing the no-choice alternative and the second containing the remaining "real" alternatives.

When there are $M$ alternatives in a choice set, the choice probabilities of choosing alternative $m$ in the MNL model have the form as specified in (4.2). The probabilities in a Nested Logit model are different from the probabilities in the standard MNL model. Assume the case where there are $N$ nests and each nest $n$ contains $M_n$ alternatives (the index representing the choice sets is suppressed for the moment), then the probability that alternative *m* is chosen from nest *n* is calculated as (cf., e.g., McFadden 1981, Kamakura, Kim and Lee 1996):

$$P(n,m) = P(m \mid n) \cdot P(n) \ , \tag{4.4}$$

where

$$P(m \mid n) = \exp(X_{nm}\beta) \ / \sum_{m'=1}^{M_n} \exp(X_{nm'}\beta) \tag{4.5}$$

is the standard MNL probability *within* a nest, and

$$P(n) = \exp(\lambda V_n) \ / \sum_{n'=1}^{N} \exp(\lambda V_{n'}) \tag{4.6}$$

is the probability of choosing nest *n*, $\lambda$ is the above mentioned dissimilarity coefficient, and $V_n = \ln(\exp(\sum_{m=1}^{M_n} X_{nm}\beta))$ is the *inclusive value* for nest *n*. When $\lambda = 1$ it is easy to show that the Nested Logit is equal to the Logit model. When there is only one nest (*N*=1) (4.5) reduces to the standard Logit formula (4.2) for conjoint experiments. In the case of the Nested Logit model, when the no-choice is coded as a series of zero's for the attributes, the inclusive value ($V_n$) of the nest containing only the no-choice is equal to zero in the Nested Logit model. Furthermore, $P(m \mid n) = 1$ for the nest containing the no-choice because it contains only one alternative.

## 4.4 Empirical Investigation of Modeling Options

We describe the results of the models described in the previous section, using a conjoint choice experiment on a technological product with six attributes[8]: Brand (6 levels), Speed (4 levels), Technology Type (6 levels), Digitizing Option (no and 2 yes-levels), Facsimile Capable (y/n), and Price

---

[8] We thank Rich Johnson from Sawtooth Software for allowing us to analyze this data. Because the data is confidential we cannot give more details about the attributes and levels.

(4 levels). For all models we use three versions, in the first situation the Price and Speed attributes are coded linear with {1, 2, 3, 4} for the four levels respectively (Speed ascending, Price descending) and the other attributes are coded using effects-type coding. In the second situation we use as linear codes {-3, -1, 1, 3} for the linear levels instead to investigate whether meancentering the linear attributes solves (part of) the problem[9]. In the third situation all attributes are coded with effects-type coding. We use 200 respondents who each had to choose from 20 choice sets with four alternatives, where the last alternative is the "no-choice" option. We used the first 12 choice sets for estimation and the last 8 for prediction purposes. Each respondent had to choose from individualized choice sets. We compare the results of the models on the log-likelihood value, AIC (Akaike 1973) and BIC (Schwarz 1978) statistics and the Pseudo $R^2$ value (e.g., McFadden 1976) relative to a null-model in which all probabilities in a choice set are equal to $1/M$. The AIC criterium is defined as: $AIC = -2\ln L + 2n$, where $n$ is the total number of estimated parameters in the model and the BIC criterium is defined as: $BIC = -2\ln L + n\ln(O)$, where $O$ is the number of independent observations in the conjoint choice experiment. We test differences in the log-likelihood values for models that are nested with the likelihood ratio (LR) test. To obtain the estimates the CONPRO computer program developed by the author is used. For details see the appendix of this thesis.

In Table 4.1 the estimation results are listed for all models with linear attributes, the left-hand side gives the results for the models with linear levels {1, 2, 3, 4}, the right-hand side with linear levels {-3, -1, 1, 3}. Table 4.2 gives the results for the models where all attributes are coded with effects-type dummies. Note that when effects-type coding is used for an attribute, the part-worth for the last level of that attribute can be obtained by taking the sum of the estimates of the other levels of that

---

9 Note that we could have used different linear coding instead, e.g., actual values for the prices.

attribute and change the sign. Note also that the Nested Logit and the No-choice Logit models are not nested, but both are nested within the standard Logit model. In the Nested Logit model we do not estimate $\lambda$ itself but estimate $(1-\lambda)$ instead, to have a direct test on $\lambda=1$.

Table 4.1: Estimation Results, Price and Speed linear

| | Levels linear attributes: {1,2,3,4} | | | Levels linear attributes: {-3,-1,1,3} | | |
|---|---|---|---|---|---|---|
| *Model* | MNL | NMNL | No-choice MNL | MNL | NMNL | No-choice MNL |
| *Parameter* | | | | | | |
| Brand A $\beta_{01}$ | 0.386 (.061)* | 0.522 (.077)* | 0.472 (.066)* | 0.350 (.064)* | 0.524 (.076)* | 0.472 (.066)* |
| Brand B $\beta_{02}$ | -0.013 (.067) | -0.009 (.078) | -0.011 (.070) | -0.012 (.069) | -0.010 (.078) | -0.011 (.070) |
| Brand C $\beta_{03}$ | 0.052 (.067) | -0.006 (.080) | 0.037 (.071) | 0.024 (.070) | 0.001 (.080) | 0.037 (.071) |
| Brand D $\beta_{04}$ | -0.150 (.071)* | -0.166 (.087)* | -0.187 (.075)* | -0.142 (.074) | -0.176 (.086)* | -0.187 (.075)* |
| Brand E $\beta_{05}$ | 0.011 (.067) | -0.018 (.081) | 0.020 (.071) | 0.013 (.070)* | -0.011 (.080) | 0.021 (.071) |
| Speed $\beta_{06}$ | -0.237 (.021)* | 0.118 (.031)* | 0.129 (.028)* | 0.046 (.014)* | 0.067 (.016)* | 0.064 (.014)* |
| Tech. Type A $\beta_{07}$ | -0.531 (.081)* | -0.678 (.096)* | -0.633 (.086)* | -0.451 (.084)* | -0.682 (.095)* | -0.634 (.086)* |
| Tech. Type B $\beta_{08}$ | 0.505 (.060)* | 0.613 (.074)* | 0.575 (.064)* | 0.432 (.062)* | 0.621 (.073)* | 0.575 (.064)* |
| Tech. Type C $\beta_{09}$ | -0.321 (.075)* | -0.366 (.086)* | -0.368 (.078)* | -0.281 (.077)* | -0.377 (.086)* | -0.368 (.078)* |
| Tech. Type D $\beta_{10}$ | 0.514 (.060)* | 0.635 (.074)* | 0.628 (.064)* | 0.471 (.062)* | 0.651 (.073)* | 0.628 (.064)* |
| Tech. Type E $\beta_{11}$ | -0.132 (.071) | -0.149 (.084) | -0.173 (.075) | -0.132 (.073) | -0.159 (.083) | -0.173 (.075) |
| Dig. Opt (n) $\beta_{12}$ | -0.586 (.049)* | -0.714 (.055)* | -0.732 (.052)* | -0.488 (.050)* | -0.721 (.055)* | -0.732 (.052)* |
| Dig. Opt (y1) $\beta_{13}$ | 0.128 (.041)* | 0.180 (.046)* | 0.172 (.043)* | 0.099 (.043)* | 0.178 (.046)* | 0.172 (.043)* |
| Facsimile $\beta_{14}$ | -0.445 (.030)* | -0.528 (.035)* | -0.543 (.032)* | -0.386 (.031)* | -0.539 (.035)* | -0.543 (.032)* |
| Price $\beta_{15}$ | -0.013 (.020) | 0.385 (.031)* | 0.396 (.028)* | 0.144 (.014)* | 0.203 (.015)* | 0.198 (.014)* |
| Nested Logit 1-$\lambda$ | - | 0.924 (.009)* | - | - | 0.840 (.017)* | - |
| No-choice $c_{nc}$ | - | - | 2.461 (.121)* | - | - | 1.150 (.048)* |
| *Statistic* | | | | | | |
| Ln-Likelihood | -2906.738 | -2715.413 | -2663.017 | -2947.716 | -2701.927 | -2663.017 |
| AIC | 5843.476 | 5462.826 | 5358.035 | 5925.431 | 5435.854 | 5358.035 |
| BIC | 5930.224 | 5555.358 | 5450.566 | 6012.180 | 5528.386 | 5450.567 |
| Pseudo R² | 0.126 | 0.184 | 0.200 | 0.114 | 0.184 | 0.200 |

*: p<0.05, standard errors between parentheses.

The first conclusion that can be drawn from Table 4.1 is that the No-choice Logit model gives the best overall fit, in both situations, and converged to the same point. The log-likelihood is significantly better than the standard MNL model (LR(1) tests, p<0.01). The No-choice Logit model and the Nested Logit model are not nested, so these models cannot be compared with an LR test. The *AIC* and *BIC* values show, however, that the No-choice Logit model fits better than the Nested Logit model, which itself is significantly better than the Logit model (LR(1) tests, p<0.01) again in both situations. Table 4.1 also shows that the estimates for the dissimilarity coefficients ($\lambda$) are significantly different from 1 for the Nested Logit model, hence the Nested Logit differs significantly from the MNL model.

When the $\beta$-estimates are compared, Table 4.1 shows that for the parameters representing the attributes with a dummy coding ($\beta_1 \cdots \beta_5, \beta_7 \cdots \beta_{14}$) the estimates do not differ much. However, in the left-hand side of Table 4.1, the coefficients of the linear attributes ($\beta_6$, $\beta_{15}$) in the standard MNL model differ strongly from the other two models. Whereas the estimate for speed is negative (a high level is unattractive) and significant for the MNL model, it is positive (a high level is attractive) and significant for the other models. The price parameter shows a similar effect; it is negative but not significant in one situation and positive in the other for the MNL model and positive (lower price is more attractive) and significant in the other two models. Clearly, both estimates for the linear attributes show a strong negative bias. Note, however, that there are also differences in the other part-worth estimates across the models. The right-hand side of Table 4.1 shows that when the Speed and Price variables are coded with values such that the mean of the levels is zero (the same can of course be obtained by meancentering the linear levels in the left-hand side of Table 4.1), the estimates for Speed and Price do not longer show the wrong sign, but are still biased downwards compared to the other models. In the No-choice MNL model the estimates for all parameters are equal in both situations, except for the linear parameters which have in the right-hand side of Table 4.1 exactly half the value of those in the left-hand side of Table 4.1, which is the result of the doubled step-length of the linear levels.

When the attributes Price and Speed are also coded with effects-type coding (Table 4.2), the $\beta$-estimates are more similar across the three models, having the right signs. Nevertheless, (almost) all estimates in the MNL model are smaller in absolute sense compared to the other two models. As shown above, the No-choice MNL model shows superior fit compared to the other two models, where MNL has the worst fit. Furthermore, the No-choice MNL model has higher t-values for the β-estimates in almost all cases compared to both other models, showing that there is more information in the data in this particular parameterization.

Table 4.2: Estimation Results, All Attributes Effects-Type

| *Model* | Logit | Nested Logit | No-choice Logit |
|---|---|---|---|
| *Parameter* | | | |
| Brand A $\beta_{01}$ | 0.352 (.065)$^*$ | 0.524 (.076)$^*$ | 0.474 (.066)$^*$ |
| Brand B $\beta_{02}$ | -0.013 (.069) | -0.011 (.078) | -0.015 (.070) |
| Brand C $\beta_{03}$ | 0.023 (.070) | -0.001 (.080) | 0.035 (.071) |
| Brand D $\beta_{04}$ | -0.141 (.074) | -0.177 (.086)$^*$ | -0.183 (.075)$^*$ |
| Brand E $\beta_{05}$ | 0.014 (.070) | -0.011 (.081) | 0.021 (.071) |
| Speed 1 $\beta_{06}$ | -0.186 (.056)$^*$ | -0.260 (.063)$^*$ | -0.260 (.057)$^*$ |
| Speed 2 $\beta_{07}$ | -0.024 (.054) | -0.049 (.060) | -0.034 (.055) |
| Speed 3 $\beta_{08}$ | 0.133 (.052)$^*$ | 0.191 (.058)$^*$ | 0.180 (.053)$^*$ |
| Tech. Type A $\beta_{09}$ | -0.452 (.084)$^*$ | -0.683 (.095)$^*$ | -0.635 (.086)$^*$ |
| Tech. Type B $\beta_{10}$ | 0.433 (.063)$^*$ | 0.621 (.073)$^*$ | 0.577 (.064)$^*$ |
| Tech. Type C $\beta_{11}$ | -0.280 (.077)$^*$ | -0.376 (.086)$^*$ | -0.367 (.078)$^*$ |
| Tech. Type D $\beta_{12}$ | 0.473 (.063)$^*$ | 0.655 (.074)$^*$ | 0.631 (.064)$^*$ |
| Tech. Type E $\beta_{13}$ | -0.136 (.074) | -0.164 (.084) | -0.179 (.075) |
| Dig. Opt (n) $\beta_{14}$ | -0.487 (.050)$^*$ | -0.720 (.055)$^*$ | -0.732 (.053)$^*$ |
| Dig. Opt (y1) $\beta_{15}$ | 0.099 (.043)$^*$ | 0.178 (.046)$^*$ | 0.172 (.043)$^*$ |
| Facsimile $\beta_{16}$ | -0.387 (.031)$^*$ | -0.541 (.035)$^*$ | -0.545 (.032)$^*$ |
| Price 1 $\beta_{17}$ | -0.389 (.060)$^*$ | -0.566 (.066)$^*$ | -0.551 (.061)$^*$ |
| Price 2 $\beta_{18}$ | -0.142 (.057)$^*$ | -0.194 (.064)$^*$ | -0.180 (.058)$^*$ |
| Price 3 $\beta_{19}$ | 0.046 (.054) | 0.088 (.060) | 0.079 (.054) |
| Nested Logit 1-$\lambda$ | - | 0.840 (.017)$^*$ | - |
| No-choice $c_{nc}$ | - | - | 1.153 (.048)$^*$ |
| *Statistic* | | | |
| Ln-Likelihood | -2942.382 | -2696.803 | -2656.854 |
| AIC | 5922.764 | 5433.605 | 5353.708 |
| BIC | 6032.645 | 5549.270 | 5469.373 |
| Pseudo $R^2$ | 0.116 | 0.189 | 0.202 |

*: p<0.05, standard errors between parentheses.

The conclusion that can be drawn from the above results is that the presence of a no-choice alternative and linearly coded attributes can give very misleading results, in particular for the parameters of those linear attributes when the conjoint choice data is estimated with a standard Logit model. However, the parameters of attributes coded with effects-type dummies are also affected, be it less severely. When all attributes are coded with effects-type coding the bias seems less strong, but still coefficients estimates are highly attenuated. Overall fit can be improved substantially by specifying a Nested Logit or by adding a No-choice constant to the design. When we compare the Nested Logit and the No-choice Logit results we see that both are capable of compensating for the no-choice zero level for the linear attributes, but there are some differences in the magnitudes of the estimated coefficients, some in the range of 5-10%, which may be important in interpreting them in practice. However, the fit of the No-choice Logit model is much better than the Nested Logit model. Note that in Table 4.1 and 4.2 the estimates for the no-choice constant are relatively big and positive. This means that the no-choice has a high overall utility, which is also shown by the high number of times the no-choice alternative was actually chosen (in *43,1%* of all choice sets).

The estimates in Table 4.1 and 4.2 were used to predict the 8 holdout choice sets. Table 4.3 gives the values of the statistics for the predictive fit of the three models for the three different designs options considered.

Table 4.3: Prediction Results

| | Levels linear attributes: {1,2,3,4} | | | Levels linear attributes: {-3,-1,1,3} | | |
|---|---|---|---|---|---|---|
| *Model* | MNL | NMNL | No-choice MNL | MNL | NMNL | No-choice MNL |
| *Statistic* | | | | | | |
| Ln-Likelihood | -1883.069 | -1741.723 | -1706.087 | -1960.978 | -1735.014 | -1706.087 |
| AIC | 3796.139 | 3515.448 | 3444.175 | 3951.956 | 3502.028 | 3444.175 |
| BIC | 3876.805 | 3601.492 | 3530.219 | 4032.622 | 3588.072 | 3530.219 |
| Pseudo $R^2$ | 0.151 | 0.215 | 0.231 | 0.116 | 0.218 | 0.231 |

| | All attributes effects-type | | |
|---|---|---|---|
| *Statistic* | | | |
| Ln-Likelihood | -1961.798 | -1735.747 | -1707.982 |
| AIC | 3961.596 | 3511.494 | 3455.963 |
| BIC | 4063.774 | 3619.049 | 3563.518 |
| Pseudo $R^2$ | 0.116 | 0.218 | 0.230 |

Table 4.3 shows a similar pattern as Table 4.1 and 4.2: the No-choice Logit model gives the best predictions which are significantly better than the Logit model (LR(1) tests, $p<0.01$) and which are also better than the Nested Logit model (*AIC*, *BIC*) in all situations. The Nested Logit model also predicts significantly better than the standard Logit model (LR(1) tests, $p<0.01$). Thus, the predictive validity results confirm the results on model fit (Table 4.1).

Note that although the MNL model with linear levels {1, 2, 3, 4} is clearly misspecified (as could be seen from the Speed and Price estimates in Table 4.1), the log-likelihood, both in estimation and prediction, is better than those of the MNL models with the two other ways of coding. However, in all situations the Nested Logit and No-choice Logit models show superior fit.

## 4.5 Conclusions and Discussion

In this chapter, we discussed several issues that play a role when there is a no-choice base alternative present in a conjoint choice experiment. From a psychological point of view, the reasons to choose a "none" may be different from the reasons to choose a regular profile. Furthermore, the presence of highly attractive or unattractive profiles in choice sets has consequences for the number of "none" choices. The "none" is chosen less often when such an (un)attractive alternative is present and more when the other profiles are more utility balanced.

We focused on the modeling of the no-choice alternative. Such an alternative is often coded as a series of zeros, because nothing is known by definition of its attributes and levels. However, when all attributes in the choice design are modeled using effects-type coding and/or linear coding, the presence of such a no-choice can give misleading estimates of the parameters of interest, because the zeros of the no-choice act as real levels of attributes in the model. We showed that this effect can especially be important when there are linear attributes in the design. The parameter estimates of the linear attributes are insignificant or even have a significant wrong sign. This problem can be solved in two ways. First, adding a constant to the design for the no-choice alternative results in better estimates, model fit and predictive fit. Second, specifying a Nested Logit model, where the no-choice is in the first nest and the other profiles in the other nest, has the same effects. However, the first solution is preferred, because the MNL model with the no-choice constant is easier to estimate, and it outperforms the Nested Logit model on estimation and predictive fit. In the situation that there are brand-dummies (or dummies on another attribute) present in the design of the experiment, modeled as real dummies and not with effects-type coding, the no-choice constant is implicitly included in the design through these dummies. However, with effects-type coding and/or linear coding on *all* attributes, a no-choice constant, or similar a constant for the "real" alternatives, is needed to set the level of the no-choice utility. Based on the findings in this chapter, we will add in the applications of the next chapters one parameter to the conjoint choice design to set the utility of the no-choice alternative, when such an alternative is present in the choice task.