Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 00 (2018) 000–000

Transportation Research Procedia

www.elsevier.com/locate/procedia

World Conference on Transport Research - WCTR 2019 Mumbai 26-31 May 2019

Predicting Lessee Switch Behavior using Logit Models

Jan-Willem Feilzer[a,b,]*, Daan Stroosnier[a], Elenna Dugundji[b]

[a]*PricewaterhouseCoopers Advisory N.V., Thomas R. Malthusstraat 5, 1066 JR Amsterdam, the Netherlands*
Vrije Universiteit, De Boelelaan 1081a, 1081 HV Amsterdam, the Netherlands

Abstract

Modeling individual's mode choice is a challenging task. In this paper, vehicle choice of lessees is discussed. Prediction of vehicle choice occurs by fitting three different logit models: standard, nested and cross-nested multinomial logistic regression. Both nested and cross-nested logit relax error term distribution assumptions and therefore allow for correlations across alternative vehicle choices. It is shown that allowing for correlation across alternatives is the proper way of modeling lessees' vehicle choice: cross-nested logit achieves best prediction results, both on training and test data.

© 2018 The Authors. Published by Elsevier B.V.
Peer-review under responsibility of WORLD CONFERENCE ON TRANSPORT RESEARCH SOCIETY.

Keywords: discrete choice, (cross-)nested logit, mode choice

1. Introduction

Modeling individual's choices in selecting transportation modes has been a large area of research. However, predicting and analyzing individuals' evaluation of mode alternatives, and their corresponding decision of mode among a set of interrelated choices, remains complex. Discrete choice models are the typical family of models used to analyze and predict an individual's choice of one alternative from a set of mutually exclusive and collectively exhaustive alternatives (Koppelman and Bhat, 2006). These types of models are widely discussed in literature, and rose to fame when Daniel McFadden won the Nobel Prize in economics for his development of theory and methods for analyzing discrete choice (McFadden, 2001). Discrete choice models have had considerable influence on the growth of the mode choice modeling field, by trying to accommodate for both observed and unobserved effects on an individual's choice. In such models, it is assumed that an individual's preference for an alternative is captured by a value, called utility, and selects the alternative with highest utility. Concurrently, the assumption is made that the analyst does not have complete information, and therefore a factor of uncertainty is considered (Ben-Akiva and

* Corresponding author. Tel.: +316-81269064
E-mail address: j.feilzer@student.vu.nl

2352-1465 © 2018 The Authors. Published by Elsevier B.V.
Peer-review under responsibility of WORLD CONFERENCE ON TRANSPORT RESEARCH SOCIETY

Bierlaire, 2003). Discrete choice models are widely used due to the extent of literature available, and the relative ease of interpretation of such models.

In this paper several models, all part of the discrete choice family, are applied to a specific use case related to mode choice. Multiple types of discrete choice models have focused on mode choice analysis (Vovsha, 1997; Bhat and Sardesai, 2006; Chu, 2009). This paper revolves around logit models, a branch of the discrete choice family. Logit models are well-represented in literature and most used for modeling mode choice (De Jong et al., 2003; Hess et al., 2012; Ding et al., 2014). Even though logit choice models have historically been most prominent in the field of mode choice modeling, such models preserve certain correlation assumptions. These assumptions might not prove accurate. The second and third model discussed in this paper build on the *previous* model each relaxing correlation assumptions.

Research in this paper is conducted for a leasing company with establishments throughout the world. This company is considered to be one of the leaders in the field of fleet management. The leasing company approached PwC to gain better understanding of its fleet data. The obtained insights should lead to improved alignment of separate entities within the company. For instance, buying, selling, and leasing of vehicles are all related, and should therefore be aligned to maximize profit. To understand customer behavior, and to provide tailored offers to these customers, it could prove greatly advantageous to model switch behavior of the company's lessees. A switch can be thought of as the choice of vehicle, given the customer has had a leasing contract with the company. That is to say, upon termination of the customer's current contract, the leasing company would like to know, what make of vehicle the customer will most likely lease next. Each constructed model should predict what make of vehicle a lessee is most likely to lease post termination of his contract. It is expected that the more assumptions on relations across alternatives are relaxed, the better a model performs in predicting these switches.

The remainder of this paper is structured as follows. In section 2, the data provided by the leasing company are examined in detail. Section 3 discusses the three types of discrete choice models used in this research. Multinomial logistic regression is the main variant of discrete choice model used for prediction. Two direct extensions of this model, nested and cross-nested multinomial logistic regression are addressed. Both these models relax the assumptions made by the multinomial logistic regression model, and should in principle allow for improved prediction accuracy. In section 4, the results of fitting all models on the provided data are discussed. Lastly, the paper is concluded in section 5, with a discussion of the obtained results and the performance of all used methods.

2. Data

This section discusses the data used to compare the performance all constructed models. As discussed in section 1, the data are provided by a vehicle leasing company with establishments throughout the world. By approaching PwC, the lease company would like to better understand customer behavior, and align entities within the company. Predicting customer switch behavior is part of this improved understanding, and the topic of this paper. To predict these switches, a thorough understanding of the provided data is essential and is provided in this section.

The provided data consists of 69,952 matched contracts or *switches* occurring between January 2014 and March 2018. One data point depicts two contracts, one regarding the driver's previous vehicle, whilst the second contract depicts the current vehicle. In addition, both contracts contain vehicle and contract characteristics which are used as modeling variables. The provided contracts originate from ten different branches of the company, with each branch depicting a different *country*. The company considers each client to belong to a particular *client segment*: Corporate, International, Private or SME. By definition, it is not possible to switch client segments. In addition, each vehicle of the company's fleet is placed within different segments. To be more precise, all vehicles are associated with the following segments: *brand classification*, *vehicle segment*, *OEM group*, *make* and *model*. For instance, an Italian driver working for a corporate association, driving a Volkswagen Passat is classified as is portrayed in Table 1. All these labels are predefined by the leasing company and are stated within each provided contract. Note that most segments depend on the higher-level segment. That is, if the variable *model* is known, one knows the variables *make*, *OEM group*, *vehicle segment* and *brand classification* by definition. Recall that it is possible that a vehicle is present in multiple client segments and countries. Additionally, if only the variable *make* is known, one does know the variable *OEM group* by definition. Knowing the make does not necessarily imply the vehicle segment to be known. Since vehicles are classified to be part of different vehicle segments based on the model of vehicle, it is possible for makes to belong to multiple vehicle segments. This property proves extremely convenient when subdividing makes into nests, described in section 3.3, and will be touched upon in this section.

Table 1. Example segmentation of a lessee.

Country	Client Segment	Brand Classification	Vehicle Segment	OEM Group	Make	Model
Italy	Corporate	Mainstream	D	VAG	Volkswagen	Passat

Aside from variables segmenting the vehicles of the fleet, each contract provides the following information: *customer ID*, *vehicle ID*, *fuel type*, *vehicle type*, *body style*, *lease type*, *catalogue price*, *commercial discount amount*, *standard discount percentage, total accessories amount, total options amount, ufwt amount, start mileage, end mileage*, *contract mileage, intro date model, end date model, sale date, sale amount*, *termination info*, *start date contract ,end date contract* and *contract duration*. The variables *mileage per month* and *switch quarter* are extracted by the analyst. Not all variables are used for prediction purposes. Some variables are either too highly correlated, or variables are omitted due to a lack of descriptive quality.

To use variables provided alongside the matched contracts, these need be processed prior modeling. Processing of data can be separated in two parts: processing of numerical variables, and processing of categorical variables. First, processing of numerical variables is discussed. All missing values of numerical variables contained in contracts are replaced with the mean value of the concerned variable after grouping by the variables *country*, *client segment* and *make*. To illustrate the matter, if the variable *catalogue price* is missing for a Volkswagen Golf of a driver stemming from the SME segment of the Spanish branch, it is replaced with the mean value of the catalogue price for that particular vehicle in those segments. In addition, outliers are set to either the determined lower or upper boundary of the concerned variable. Some statistics on the numerical variables used for prediction can be found in Table 2.

Table 2. Summary of all (processed) numerical variables used for modelling purposes. The rows indicate the variable. The columns portray the mean, standard deviation, median value, minimum value and the maximum value of the concerned variable respectively.

	μ	σ	Median	Min. Value	Max. Value
Catalogue price	45,051.68	83,363.76	23,829.43	5,575.00	1,440,000.00
commercial discount percentage	6,765.62	7,835.73	4,956.86	0.00	125,760.00
standard discount percentage	10.68	7.96	10.00	0.00	35.00
total accessories amount	270.93	712.85	0.00	0.00	9,877.00
Total options amount	3299.08	6,525.14	1,000.86	0.00	49,984.00
ufwt amount	613.26	1,115.05	250.00	0.00	10,000.00
mileage per month	2639.57	1,248.30	2500.00	500.00	10,000.00
contract duration	42.10	10.70	42.00	6.00	96.00

To use numerical variables for prediction, these need be scaled. Prior to this procedure, all variables depicting a monetary value are transformed using the natural logarithmic function. The variables *catalogue price*, *commercial discount amount*, *total accessories amount*, *total options amount*, and *ufwt amount* are transformed by taking the natural log of the original value. In addition, all zero values for which the natural log is not defined, are replaced with the minimum value for which the natural log is defined. The idea of using the natural logarithmic function for variable transformation, is to push the variable towards being normally distributed. Post logarithmic transform, the monetary variables are treated as any other numerical variable. Next, all numerical variables are standardized to have zero mean, and standard deviation of 1. This transformation refrains functions present in discrete choice models from saturating.

Processing of categorical variables occurs in slight different fashion. No missing values occur in the data. The provided data does contain values such as *unknown*, or *country did not supply a value*. These values rarely occur and therefore do not form a significant problem. States of categorical variables that rarely occur, are either set to the state *other*, or are merged with an already-existing state. For instance, the variable *client segment* is reduced to contain three states, since Private is merged with SME. A brief summary of the categorical variables can be found in Table 3.

Lastly, the provided vehicle leasing data are split into training and test data. These data are identical for all models used in this thesis. The most recent 10% of data are considered to be test data. Data are classified as most recent based on the date a switch occurred. Providing test data allows for models to predict on data that are seen as most presentable of the current situation. Prior to splitting data into train and test data, all data are shuffled to avoid dis-balanced data.

Table 1. Summary of all categorical variables used for modelling purposes. The rows indicate the categorical variable. The first column portrays the number of unique states per variable, with the number in parenthesis stating the number of unique variables prior processing. The remaining columns depict the most occurring state per variable, the least occurring state per variable, and the least occurring state per variable prior to processing.

	#States	**Max State**	**Min State**	**Min State (orig)**
country	8 (10)	France	Other	Austria
client segment	3 (4)	Corporate	SME	Private
fuel type	2 (6)	Diesel	Petrol	Unknown
vehicle type	2 (2)	Vehicle	Van	Van
vehicle segment	8 (12)	D	Other	F
body style	8 (17)	Stationwagon	Vehicle/Van	Unknown
lease type	2 (2)	Operational Lease	Financial Lease	Financial Lease
switch quarter	4 (4)	1	2	2
make	13 (41)	Volkswagen	Nissan	Chrysler

3. Discrete Choice Models

This chapter discusses the family of models used to model the vehicle leasing data, discrete choice models. More precisely, this section discusses logit models, a branch of the discrete choice model family. These types of models are widely used to model mode choice. Section 3.1 describes the properties common to all discrete choice models. The general framework of such models is introduced and elaborated on. The subsequent sections dive into the three types of logit models used in this thesis. The models addressed in sections 3.3 and 3.4 relax the assumptions made for standard multinomial logistic regression models (section 3.2). The chapter is concluded with a section describing the applicability of these models on the provided vehicle leasing data. Note that almost all mathematical derivations and formulae are derived from Train (2009). Text, derivations and formulae not originating from this source are cited accordingly.

3.1 General Properties

Discrete choice models analyze and predict individuals' choices among a set of alternatives. This set of alternatives, the choice set, needs to contain three characteristics. The alternatives of the choice set need be mutually exclusive, and the choice set must be exhaustive and finite. The first two properties are not restrictive, since data can be modeled in such ways that these properties are met. However, finity of the choice set is a restrictive property and the defining characteristic of discrete choice models. Typically, these models are derived under the assumption of utility maximization. Models derived under this assumption are referred to as random utility models. In random utility models, it is assumed that the decision maker assigns a preference value, called utility, to each alternative in the choice set. The decision maker is assumed to have perfect discrimination capability and therefore chooses the alternative possessing the highest utility value. In addition, the analyst is assumed to have incomplete information, hence a factor of uncertainty needs to be considered. Such models can be defined as follows. If decision maker n chooses an alternative from a choice set C of size J elements, each element is assigned a utility value U. Then alternative i is chosen if and only if $U_{ni} > U_{nj}\ \forall j \neq i$. Each utility value is composed of a deterministic and random part: $U_{ni} = V_{ni} + \varepsilon_{ni}$. The deterministic part V_{ni}, or the representative utility, portrays the attributes of both the decision maker and the alternative, and is often specified to be linear in parameters. The random part, or error term of utility, ε_{ni}, captures the factors that affect utility but are not included in V_{ni}. The value of $\varepsilon_{ni}\ \forall i$ is not known a priori, hence these terms are treated as random. Assuming a joint density function $f(\varepsilon_n)$ of the vector containing all random terms, the probability of individual n choosing alternative i is defined as

$$\begin{aligned} \mathbb{P}_{ni} &= \mathbb{P}(U_{ni} > U_{nj}\ \forall j \neq i) = \\ &\mathbb{P}(V_{ni} + \epsilon_{ni} > V_{nj} + \epsilon_{nj}\ \forall j \neq i) = \\ &\mathbb{P}(\epsilon_{nj} - \epsilon_{ni} < V_{ni} - V_{nj}\ \forall j \neq i) = \\ &F(\epsilon_{nj} - \epsilon_{ni} < V_{ni} - V_{nj}\ \forall j \neq i) = \\ &\int_{\epsilon} I(\epsilon_{nj} - \epsilon_{ni} < V_{ni} - V_{nj}\ \forall j \neq i) f(\epsilon_n) d\epsilon_n \end{aligned} \quad (1)$$

Note that $I(\cdot)$ denotes the indicator function, equaling one when true, and zero otherwise. Observe from equation 1 that different types of discrete choice models arise from different distributions of $f(\varepsilon_n)$. The models used in this paper cause the integral of equation 1 to be of closed form due to the specification of $f(\varepsilon_n)$; hence these models do not need to be evaluated numerically. The choice of distribution of the random terms, and the motivation for these different assumptions will be discussed in the following sections.

In addition, from equation 1 note that only the signs of the differences of utilities matter in choosing an alternative, rather than their absolute values. Consequently, this means that the only parameters able to be estimated are those capturing differences across alternatives. Due to this fact, the deterministic part of utility is often specified to be linear in parameters with a constant added. This constant captures the average effect of all factors not included in the model, on utility, and is referred to as the alternative specific constant. Including these constants produces the convenient property that the mean of the error terms can be assumed to equal any constant, typically zero. The deterministic part of utility is then defined as $V_{ni} = x_{ni}\beta + k_i$, where vector x_{ni} depicts the attributes of alternative i and individual n, β is the vector of coefficients of these variables to be estimated, and k_i denotes the alternative specific constant. Including alternative specific constants results in the random part of utility having zero mean by construction. If ε_{ni} has a nonzero mean, adding the alternative specific constants result in the remaining error term having zero mean. Therefore, without loss of generality, it can be assumed that the mean of the error terms is equal to zero by including alternative specific constants in the deterministic part of utility.

Another consequence of the fact that only the signs of differences between utilities matter, is that this property also holds for the alternative specific constants. A direct result is that it is impossible to estimate all alternative specific constants, since there are infinitely many possibilities for a and b when $a - b$ is equal to some constant. Hence one of the constants is typically normalized to zero. It does not matter which alternative specific constants is normalized, since all other constants are interpreted as being relative to whichever constant is normalized.

In addition to normalizing one of the alternative specific constants, the scale of utility must be normalized too. The necessity of this normalization can be observed from the fact that the alternative with highest utility does not change regardless of the scale of utility: the models $U_{ni} = V_{ni} + \varepsilon_{ni}\ \forall i$, and $U_{ni} = \lambda V_{ni} + \lambda\varepsilon_{ni}\ \forall i$ are equivalent. Generally, normalizing the scale of utility corresponds to normalizing the variance of the error terms. Observe that the scale of utility and the variance of the error terms are related by definition, since $Var(aX) = a^2 Var(X)$. Therefore, multiplying utility by λ corresponds to the variance of each ε_{ni} changing by a factor λ^2. The models explained in the subsequent sections assume that the error terms are independently, identically distributed (i.i.d). When the i.i.d assumption is imposed, normalization is quite simple; the error variance is normalized to some convenient value. Since the i.i.d assumption causes all error terms to have equal variance, normalizing the variance of any of the error terms sets the variance for all error terms. In addition, note that the i.i.d. assumption causes the integral of equation 1 to be of closed form.

3.2 Multinomial Logistic Regression

Multinomial logistic regression is the most straightforward and widely used discrete choice model since modeling is quite straightforward, and results of the model are easily interpretable. The derivation of the model is based on the framework specified in section 3.1, and the choice of distribution of unobserved utility $f(\varepsilon_n)$. The multinomial logistic regression model assumes that the random parts of utility are independently, identically extreme value distributed. This distribution is generally referred to as Gumbel. The density function and cumulative distribution of the Gumbel distribution are stated in equations 2 and 3 respectively.

$$f(\epsilon_{ni}) = e^{-\epsilon_{ni}} e^{-e^{-\epsilon_{ni}}} \tag{2}$$

$$F(\epsilon_{ni}) = e^{-e^{-\epsilon_{ni}}} \tag{3}$$

The variance of this distribution is equal to $\pi^2/6$. Recall from section 3.1 that assuming a variance value implies normalizing the scale of utility. The mean of the Gumbel distribution is not equal to zero. However, since only the differences in utility values matter, this is irrelevant. Note that the mean of the difference of two random terms with equal mean is equal to zero by definition, hence all prerequisites are met. To derive the choice probabilities of the multinomial logistic regression model, the assumption of independent error terms becomes significant. Using the property that the cumulative distribution becomes the product of individual cumulative distributions for independent error terms, equation 1 can be written as:

$$\mathbb{P}_{ni} = \int \Big(\prod_{j \neq i} e^{-e^{-(\epsilon_{ni} + V_{ni} - V_{nj})}}\Big) e^{-\epsilon_{ni}} e^{-e^{-\epsilon_{ni}}} \, d\epsilon_{ni} \tag{4}$$

Some algebraic alterations of equation 4, result in the final specification of the choice probabilities of the multinomial logistic regression model, stated in equation 5. For the derivation of this equation, the reader is referred to Train (2009).

$$\mathbb{P}_{ni} = \frac{e^{V_{ni}}}{\sum_j e^{V_{nj}}} \tag{5}$$

Two important properties arise from these choice probabilities. Firstly, McFadden et al. (1973) showed that the log-likelihood function of these probabilities has a global maximum, guaranteeing convergence of the maximization procedures. This property becomes extremely convenient when estimating the models discussed in sections 3.3 and 3.4, and will be touched upon in these sections. Secondly, the assumption of independence of error terms creates the notion of *independence from irrelevant alternatives* (IIA). This property states that for any two alternatives *i* and *k*, the ratio of probabilities remains equal when alternatives are added to the choice set.

$$\frac{\mathbb{P}_{ni}}{\mathbb{P}_{nk}} = \frac{\dfrac{e^{V_{ni}}}{\sum_j e^{V_{nj}}}}{\dfrac{e^{V_{nk}}}{\sum_j e^{V_{nj}}}} = \frac{e^{V_{ni}}}{e^{V_{nk}}} = e^{V_{ni} - V_{nk}} \tag{6}$$

Note from equation 6 that the ratio of probabilities only depends on the alternatives *i* and *k*, and is therefore independent of any other alternatives present in the choice set. This property can impose severe limitations to the multinomial logistic regression model. These limitations are best illustrated with the famous red bus, blue bus paradox. Imagine that the decision maker has two alternatives to choose from when commuting to work: go by car, or take a blue bus. Assuming $P_{car} = P_{blue\ bus} = 1/2$, equation 6 becomes equal to one. Adding a red bus to the choice set should intuitively not matter to the decision maker. That is, taking a red bus or blue bus is most likely irrelevant, and therefore the assumption that $P_{\text{blue bus}} = P_{\text{red bus}} = 1/4$, and $P_{\text{car}} = 1/2$, is reasonable. However, the IIA property states that the ratio of probabilities of the alternatives car and blue bus remains the same. This implies that $P_{\text{blue bus}} = P_{\text{red bus}} = P_{\text{car}} = 1/3$. In other words, in such a model correlation across alternatives is not possible. The data used in this thesis could potentially inhabit such correlations. To overcome the limitation of standard multinomial logistic regression, nested multinomial logistic regression models are introduced next.

3.3 Nested Multinomial Regression

As discussed in section 3.2, using the multinomial logistic regression model implies that correlations across alternatives cannot be modeled. This section introduces a methodology that partly overcomes this limitation: nested multinomial logistic regression. These types of models can be placed in a more general framework of models: generalized extreme value models (GEV). The main property defining these models is that the error terms of utility for all alternatives are jointly distributed as generalized extreme value. This property allows for correlation across alternatives. It will be shown that the multinomial logistic regression model provided in section 3.2 is an instance of this family of models as well. When all correlations across alternatives are equal to zero, the GEV distribution becomes the product of independent extreme value distributions, as is the case for multinomial logistic regression. Nested multinomial logistic regression is deemed an appropriate modeling structure when the choice set can be divided into subsets of alternatives: nests. Two properties hold when dividing the choice set into nests. Firstly, the IIA assumption holds for alternatives within the same nest. That is, for two alternatives in the same nest, the ratio of probabilities is independent of the remaining alternatives within that nest. Secondly, the IIA assumption does not hold for alternatives across nests, allowing for correlations across alternatives of different nests. In general, nests are visualized using a tree structure, in which each branch denotes a nest of alternatives. Within those nests, the IIA assumption holds. The leaves of the tree depict the alternatives of the choice set.

Concerning the derivation of the probabilities for the nested logistic regression model, McFadden (1978) showed that the model is consistent with the utility maximization theory provided in section 3.1. To derive these probabilities, suppose the choice set C consisting of J alternatives is to be divided into K nests B_k such that $\boldsymbol{B} = \bigcup_{k=1}^{K} \boldsymbol{B}_k$ and $\boldsymbol{B}_k \cap \boldsymbol{B}_{k'} = \emptyset \forall \boldsymbol{k} \neq \boldsymbol{k'}$. Then, the nested model is obtained by assuming that error vector ε_n has a type of GEV cumulative distribution stated below.

$$F(\varepsilon_n) = \exp\left(-\sum_{k=1}^{K}\left(\sum_{i\in B_k} e^{-\varepsilon_{ni}/\lambda_k}\right)^{\lambda_k}\right) \tag{7}$$

It can immediately be observed that equation 7 collapses to the product of independent extreme value distributions provided in equation 3 when $\lambda_k = 1\ \forall k$; hence the model is reduced to the standard multinomial logistic regression model. This parameter measures the degree of independence of the error terms in utility among the alternatives in nest k. The higher the value of λ_k the less the correlation across alternatives in nest k. In equation 7, the marginal distribution of each ε_{nj} is still univariate extreme value, however the ε_{nj}'s are correlated within nests. For any two variables in different nests, the error terms are still uncorrelated.

McFadden (1978) showed that the probability of individual n choosing alternative $i \in B_k$ is defined as

$$\mathbb{P}_{ni} = \frac{e^{V_{ni}/\lambda_k}\left(\sum_{j\in B_k} e^{V_{nj}/\lambda_k}\right)^{\lambda_k-1}}{\sum_{l=1}^{K}\left(\sum_{j\in B_l} e^{V_{nj}/\lambda_l}\right)^{\lambda_l}} \tag{8}$$

From equation 8 it can be shown that the IIA assumption still holds for alternatives sharing a nest. The fraction of probabilities of two alternatives $i \in B_k$ and $m \in B_l$ is solely defined by the numerator of the equation, since the denominator remains equal for all alternatives:

$$\frac{\mathbb{P}_{ni}}{\mathbb{P}_{nm}} = \frac{e^{V_{ni}/\lambda_k}\left(\sum_{j \in B_k} e^{V_{nj}/\lambda_k}\right)^{\lambda_k - 1}}{e^{V_{nm}/\lambda_l}\left(\sum_{j \subset B_l} e^{V_{nj}/\lambda_l}\right)^{\lambda_l - 1}} \tag{9}$$

Observe that the terms in parentheses in equation 9 cancel out when k = l, resulting in the fraction of probabilities only depending on the attributes of i and m. Hence, the IIA assumption holds for alternatives sharing the same nest. Interestingly, note that some form of IIA still holds if $k \neq l$. In this case, the probability ratio only depends on all alternatives of nests k and l: independence from irrelevant nests.

To better grasp the notion of nested logistic regression models, it is possible to decompose equation 8 into two separate logistic regression models. To be precise, the probability of individual n choosing alternative i in nest B_k can be expressed as the product of two probabilities. Specifically, the probability of alternative $i \in B_k$ being chosen times the probability of nest B_k being chosen: $P_{ni} = P_{ni|B_k} \cdot P_{nB_k}$. This notation allows for splitting utility into a part depending on attributes of the nest, and a part depending on attributes describing the alternative. The attributes describing the nest only vary over nests; they do not vary over alternatives within the nests. The attributes describing the alternatives vary over the alternatives within a nest. Setting $U_{ni} = W_{nk} + Y_{ni} + \varepsilon_{ni}$, in which W_{nk} only depends on attributes describing nest k, and Y_{ni} depends on attributes describing alternative j, allows to write the conditional and marginal probabilities to be expressed as 10 and 11 respectively.

$$\mathbb{P}_{ni\,B_k} = \frac{e^{Y_{ni}/\lambda_k}}{\sum_{j \subset B_k} e^{Y_{nj}/\lambda_k}} \tag{10}$$

$$\mathbb{P}_{nB_k} = \frac{e^{W_{nk} + \lambda_k I_{nk}}}{\sum_{l=1}^{K} e^{W_{nl} + \lambda_l I_{nl}}} \tag{11}$$

in which $I_{nk} = \ln \sum_{j \in B_k} e^{Y_{nj}/\lambda_k}$. For the derivation of equations 10 and 11 the interested reader is referred to Train (2009). Observe from equation 11 that the attributes varying over nests but not over alternatives within each nest are included. The quantity λI_{nk} is considered to be the expected utility individual n receives from the choice among alternatives within nest B_k. Quantity I_{nk} is referred to as the inclusive utility of nest B_k and links the upper and lower model. The upper model refers to the choice of nest; the lower model to the choice of alternative within the nest.

Nested logistic regression models maintain the property that the interval of equation 3.1 is of closed form. This convenient property allows for estimation of the model's parameters by standard maximum likelihood techniques. However, maximization could still be a demanding task due to the rugged landscape of the log-likelihood function; convergence to a global optimum is not guaranteed. To point estimation in the right direction, the analyst could commence with estimating a standard multinomial logistic regression model. The obtained parameters could then be used as starting values for the estimation of nested models. This procedure ensures appropriate starting values of the nested model and potentially eases the convergence to a global minimum. Note however, that convergences cannot be guaranteed; optimization techniques are only pointed to a possible appropriate direction.

Nested models still impose some restrictions. Note that the assumption $B = \bigcup_k B_k$ and $B_k \cap B_{k'} = \emptyset \, \forall k \neq k'$ constrain each alternative to only be part of one nest, which could potentially be an inappropriate modeling assumption. Alternatives sharing a nest are put together since it is assumed that they have similar unobserved characteristics. Of course, the possibility of an alternative sharing these characteristics with multiple nests exist; it would be convenient if alternatives could belong to multiple nests. Cross-nested logistic regression models allow for this relaxation and are discussed in the subsequent section.

3.4 Cross-Nested Multinomial Logistic Regression

Cross-nested models are roughly similar to nested models with one important difference. The assumption of alternatives belonging to one nest is relaxed; alternatives can belong to multiple nests. To allow for this property, allocation parameters $\alpha_{ik} \geq 0$ are added to the model, indicating the degree to which alternative i belongs to nest k. Intuitively, $\alpha_{ik} = 0$ states that alternative i does not belong to nest k. For interpretability reasons the allocation parameters are usually scaled to $\sum_k \alpha_{ik} = 1 \forall i$. This is not a restrictive assumption however. In the remainder of this section it is assumed that this normalization has occurred. The parameter λ_k still serves the same function as in nested models and portrays the degree of independence across alternatives within nest k. Then, the probability that individual n chooses alternative i in a cross-nested structure is then defined as

$$\mathbb{P}_{ni} = \frac{\sum_{k} (\alpha_{ik} e^{V_{ni}})^{1/\lambda_k} \left(\sum_{j \in B_k} (\alpha_{jk} e^{V_{nj}})^{1/\lambda_k}\right)^{\lambda_k - 1}}{\sum_{l=1}^{K} \left(\sum_{j \in B_l} (\alpha_{jl} e^{V_{nj}})^{1/\lambda_l}\right)^{\lambda_l}} \tag{12}$$

Observe that the probability specification of a cross-nested model shares many characteristics with the probability specification of a nested structure given in equation 8. The difference lies in the numerator of equation 12 including a summation over all nests containing alternative i. The attentive reader could have observed that a cross-nested model collapses to a nested model if all alternatives are only present in one nest: $\alpha_{ik} = 1$ for $i \in B_k$.

Just as for nested models, the probability of individual n choosing alternative i can be decomposed to a marginal probability depicting the probability of choosing nest k (equation 13), and a conditional probability depicting the probability of choosing alternative i given

$$\mathbb{P}_{nk} = \frac{\left(\sum_{j \in B_k} (\alpha_{jk} e^{V_{nj}})^{1/\lambda_k}\right)^{\lambda_k}}{\sum_{l=1}^{K} \left(\sum_{j \in B_l} (\alpha_{jl} e^{V_{nj}})^{1/\lambda_l}\right)^{\lambda_l}} \tag{13}$$

$$\mathbb{P}_{ni|B_k} = \frac{(\alpha_{ik} e^{V_{ni}})^{1/\lambda_k}}{\sum_{j \in B_k} (\alpha_{jk} e^{V_{nj}})^{1/\lambda_k}} \tag{14}$$

Note that the inclusive utility has dropped out of the equations due to the relaxation that alternatives can belong to multiple nests. For the mathematical derivation of these equations, the reader is referred to Train (2009). Concerning optimization of the model's parameter, convergence cannot be guaranteed. Just as for the nested model, the rugged landscape of the log-likelihood function could produce several local optima. Again, parameters of the cross-nested model could be initialized by first estimating either a standard or nested model and using these estimates as initial parameter values.

3.5 Application to Vehicle Leasing Data

This section addresses how the models discussed in this chapter can be applied to the vehicle leasing data. The section is mainly concerned with defining appropriate nests for the nested and cross-nested models. No nests are required for a standard multinomial logistic regression model; fitting such a model on the vehicle leasing data is straightforward. However, as discussed in previous sections, the outcome of the standard model is very useful. The estimates of the parameters are used as starting values for both the nested and cross-nested models. Besides pointing the maximization procedure in the right direction, this procedure significantly reduces computation time.

All three models use the same explanatory variables, which can be classified as numerical and categorical variables. Recall from section 2 that the numerical variables consist of the variables *catalogue price, commercial discount amount, standard discount percentage, total accessories amount, ufwt amount, mileage per month* and *contract duration*. The remaining variables consist of the categorical variables *country, client segment, fuel type, vehicle type, vehicle segment, body style, lease type, switch quarter* and *make*. Regarding categorical variables, rather than estimating one β per state of the variable, one β per state of the variable per alternative is estimated. Namely, for state Diesel of the variable *fuel type*, one β per alternative is estimated ($\beta_{Diesel\text{-}Audi}$, $\beta_{Diesel\text{-}BMW}$, etc.). Of course, as discussed in section 3.1, one of the βs per categorical variable is held fixed; all other βs are estimated with respect to the fixed β. In addition, a categorical variable with n unique states, produces $n - 1$ (times the number of unique alternatives) different βs to be estimated, since the n^{th} state is a perfect linear combination of the previous $n - 1$ states. Additionally, one β per alternative for each numerical variable is estimated ($\beta_{catalogue\ price\text{-}Audi}$, $\beta_{catalogue\ price\text{-}BMW}$, etc.).

The multinomial logistic regression uses all these explanatory variables to analyze and predict. As discussed in section 3.2, the IIA assumption does not allow for correlation across alternatives. One can imagine however, that for instance adding a Fiat 500 to the choice set should not change the decision maker's choice when he is looking for an SUV type of vehicle. To allow for correlations across alternatives, the alternatives are divided into nests. This division proves relatively straightforward due to the vehicle segmentation provided by the leasing company. Section 2 states that the leasing company assigns each model of vehicle to one particular vehicle segment. Note however, that the make of vehicle can belong to multiple vehicle segments, since a make of vehicle consists of multiple models. The exact distribution of the target variable *new make* over the variable *new vehicle segment* is shown in Table 4. Note that the variable new vehicle segment is not used for prediction. It is solely used to divide new make into nests.

Table 2. Overview of the distribution of the target variable *new make* over the variable *new vehicle segment*. This segmentation is provided by the leasing company. The vehicle segment to which an alternative is assigned to most often is stated in bold. The rows indicate the alternatives, whilst the columns indicate the states of the variable *vehicle segment*.

	A	**B**	**C**	**D**	**E**	**F**	**LCV**	**MPV**	**Pickup**	**S**	**SUV**
Audi	0	167	1413	**2923**	1196	3	0	0	0	42	1308
BMW	0	0	789	**2202**	1194	10	0	525	0	2	1835
Citroen	40	316	520	73	0	0	**1333**	916	0	0	10
Ford	3	747	**1646**	926	0	0	1290	886	66	6	390
Mercedes-Benz	0	0	574	**2706**	1046	3	427	204	0	5	1158
Nissan	0	71	108	0	0	0	56	12	22	0	**1173**
Opel	2	359	**1714**	1121	0	0	322	251	0	0	224
Other	192	1139	1564	595	88	54	422	420	111	4	**2490**
Peugeot	37	616	**1839**	693	0	0	1190	499	0	1	1142
Renault	7	2287	1166	425	0	0	**2298**	864	0	0	448
Skoda	1	72	**1294**	752	0	0	0	1	0	0	134
Volkswagen	28	312	2345	**3360**	24	0	1221	1476	35	0	942
Volvo	0	0	287	496	340	0	0	0	0	0	**906**
Total	310	6086	15259	16272	3888	70	8559	6054	234	60	12160

From Table 4**,** it becomes clear that all makes belong to multiple vehicle segments. For the nested multinomial logistic regression model, alternatives are restricted to be part on only one nest. To determine the nesting structure of this model, and to determine which nest each alternative belongs to, each alternative is assigned to the vehicle segment in which the alternative occurs most. This assignment is shown in Table 5. Observe from this table that only four unique vehicle segments are assigned to be a nest: C, D, LCV and SUV. Each nest has at least two alternatives that belong to it. A schematic representation of the nesting structure of the model is visualized in Figure 1**.**

Table 3 Assignment of nests to each alternative of the choice set. Each nest corresponds to the nest in which the alternative occurs most.

Alternative	Nest	Alternative	Nest
Audi	D	Other	SUV
BMW	D	Peugeot	C
Citroen	LCV	Renault	LCV
Ford	C	Skoda	C
Mercedes-Benz	D	Volkswagen	D
Nissan	SUV	Volvo	SUV
Opel	C		

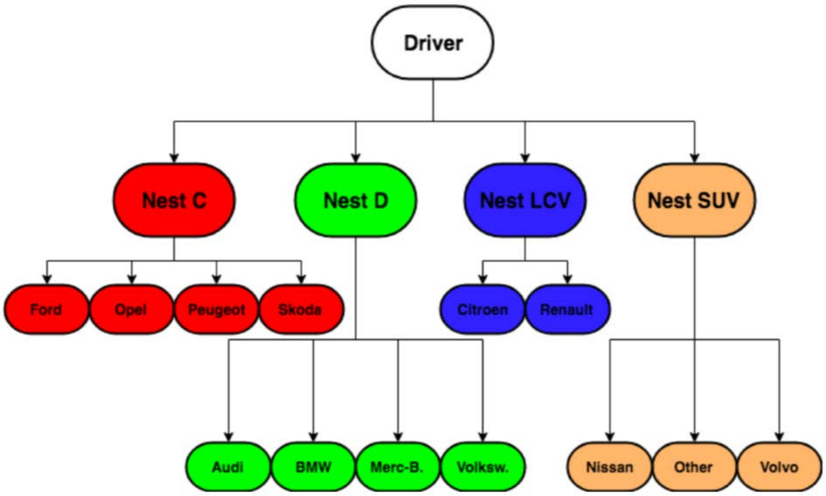

Figure 1. Schematic overview of the nesting structure used for the nested multinomial logistic regression model. Each nest is associated with its own colour. In addition, each alternative is only part of one nest.

Regarding cross-nested multinomial logistic regression, the restriction of each alternative belonging to one nest is dropped. Each alternative is allowed to be contained in multiple nests. The model estimates the allocation parameters α, indicating the degree to which an alternative belongs to a particular nest. To determine the nests each alternative belongs to, and the allocation parameters α corresponding to these nests, Table 4 is used. First, each number of this table is divided by the total occurrences of the alternative in the data. This procedure results in a table denoting the fraction of the number of times an alternative is assigned to a particular vehicle segment over the total number of occurrences of the alternative in the data. The idea is that these fractions function as starting values for the allocation parameters α in the cross-nested multinomial logistic regression model. All fractions less than 0.10 are considered to equal zero. For interpretability reasons, the allocation parameters are usually scaled to $\sum_k \alpha_{ik} = 1 \forall i$, with i denoting the alternative and k the nest. Therefore, all fractions less than 0.10, but greater than zero are added to different α. Since these α solely function as starting values for the cross-nested model, this should not impose a problem. Table 6 depicts the starting values of the allocation parameters α stemming from the above-mentioned procedures. Figure 2 depicts a schematic overview of the cross-nested structure. Note that for interpretability reasons, only three nests are depicted.

Table 4. Distribution of the alternatives over the different nests. Each value depicts the starting value of the allocation parameter α associated with the alternative and the nest. Note that the sum of each row is equal to 1, satisfying the normalization requirement. The rows indicate the alternative, whilst the columns indicate the nests.

	B	C	D	E	LCV	MPV	SUV
Audi		0.23	0.41	0.17			0.19
BMW		0.20	0.34	0.18			0.28
Citroen		0.29			0.42	0.29	
Ford	0.19	0.28	0.16		0.22	0.15	
Mercedes-Benz			0.64	0.17			0.19
Nissan							1.00
Opel		0.72	0.28				
Other	0.41	0.22					0.37
Peugeot		0.49	0.12		0.20		0.19
Renault	0.41	0.16			0.31	0.12	
Skoda		0.67	0.33				
Volkswagen		0.38	0.34		0.13	0.15	
Volvo		0.14	0.24	0.17			0.45

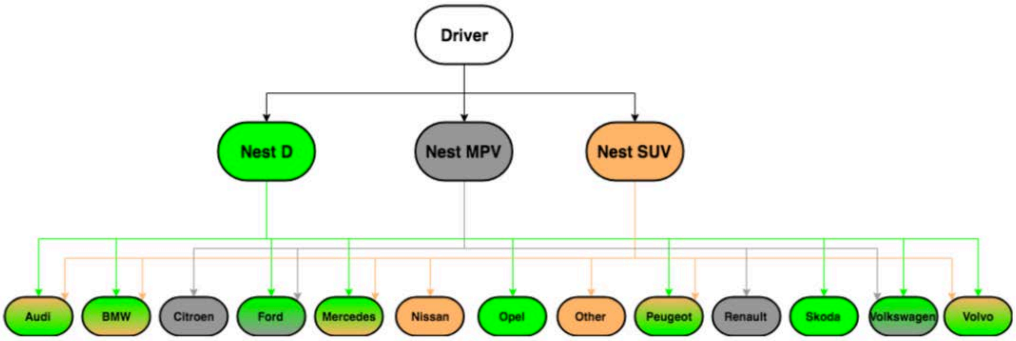

Figure 2. Schematic overview of the nesting structure used for the cross-nested multinomial logistic regression model. Each nest is associated with its own colour. Contrary to the nested model, alternatives are allowed to belong to multiple nests. The colour(s) of the alternatives indicate the nest(s) they belong to. Note that for interpretability reasons, only three nests are shown.

4. Results

This section describes the obtained results by all three discussed logit models. Recall that all models utilize the same data, and that estimates of the multinomial logit model are used as starting values for both the nested and cross-nested models. Observe from Table 7 the results regarding all discrete choice models. The leftmost column states performance measures, whereas the remaining columns indicate the values associated with these measures for the standard, nested, and cross-nested multinomial logistic regression models respectively. The statistic $\mathcal{L}(0)$ corresponds to the null log likelihood. The null log likelihood is the log likelihood of the sample for a logistic regression model such that the deterministic part of the utility function is zero for all alternatives (Bierlaire, 2015). In addition, $\mathcal{L}(c)$ is defined as the log likelihood of the sample where the deterministic part of utility of each alternative contains only the alternative specific constants. The statistic $\mathcal{L}(\hat{\beta})$ denotes the final log likelihood of the estimated model, and $-2[\mathcal{L}(0) - \mathcal{L}(\hat{\beta})]$ denotes the likelihood ratio test. The likelihood ratio test compares the goodness of fit of two models, the null and final model. Lastly, $\boldsymbol{\rho}^2$ and $\hat{\boldsymbol{\rho}}^2$ are defined as the likelihood ratio index and the adjusted likelihood ratio index respectively. The latter is a slight adjustment of the former sine it considers the number of estimated parameters K.

Table 5. Performance of al logit models on training data. Each model was estimated using 62,056 observations. Values equal for all three models are omitted in the two rightmost columns.

Summary Statistics	MNL	nested MNL	cross-nested MNL
$\mathcal{L}(0)$	-159,170.497		
$\mathcal{L}(c)$	-152,117.028		
$\mathcal{L}(\hat{\beta})$	-106,158.670	-106,121.880	-105,995.517
$-2[\mathcal{L}(0) - \mathcal{L}(\hat{\beta})]$	106,023.655	106,097.235	106,349.960
ρ^2	0.333	0.333	0.334
$\bar{\rho}^2$	0.329	0.329	0.330

Observe from Table 7 that the cross-nested model achieves best results on all performance measures. Of course, it is to be expected that both the nested and cross-nested models outperform the standard multinomial logit model, since estimates of this model are taken as starting values of the nested and cross-nested models. The likelihood ratio test indicates that all three models are significant improvements relative to the null-model. Comparing this statistic for all three models, the more restrictive assumptions are relaxed, the better fit the model is on training data. Only for the cross-nested model the improved final log likelihood with respect to the other two models, results in a better (adjusted) likelihood ratio index. Observe that the improvement of final log likelihood value of the nested model is not significant enough to obtain a better likelihood ratio index.

Since performance of all three models is roughly similar, it need be checked if dropping the IIA assumption is relevant. To do so, note the estimates of the nest parameters stated in Table 8. Recall from section 3.3 that the nested model collapses to a standard multinomial logistic regression model if all nest parameters are equal to one, $\lambda_k = 1\ \forall k$. Note that both nest D and SUV are significant irrelevant of the maintained significance level. Nest C is deemed appropriate depending on the maintained significance level. Albeit slight, all estimated nest parameters are greater than one, hence validating relaxing of IIA.

Table 6. Relevance of nest parameters of the nested multinomial logit model.

Nest	Estimate	Std. Error	*t*-stat	*p*-value
C	1.24	0.124	1.92	0.05
D	1.66	0.126	5.18	0.00
LCV	1.00	FIXED		
SUV	1.73	0.249	2.93	0.00

Further relaxation of assumptions leads to the cross-nested model, of which the nest and allocation parameters are displayed in Table 9 and Table 10 respectively. Note that some allocation parameters are fixed. These parameters were estimated with infinite standard error at first. The cross-nested model was estimated again fixing these allocation parameters at the estimated value produced on the first run. Interestingly, all nest parameters are significant and all estimates of these parameters differ significantly from one, hence relaxing assumptions is again validated. Observe that nest MPV has the highest estimated parameter value. The higher the nest parameter estimate, the more correlated alternatives of this nest are within the nest, rather than outside the nest.

Table 7. Relevance of the nest parameters of the cross-nested logit model. Note that no statistics are displayed for nest LCV. This parameter was estimated with infinite standard error at first. On the second run this parameter was therefore held fixed at the estimated value of the first run.

Nest	Estimate	Std. Error	*t*-stat	*p*-value
B	1.75	0.319	5.490	0.00
C	1.54	0.034	45.28	0.00
D	2.08	0.066	31.46	0.00
E	1.24	0.148	8.420	0.00
LCV	1.10	FIXED		

MPV	2.89	0.160	18.02	0.00
SUV	1.94	0.177	10.98	0.00

Even though all nest parameters of the cross-nested model are deemed significant and relevant, Table 10 states some allocation parameters indicating that inclusion of the corresponding alternative in the concerned nest is not strongly supported by the given data. In other words, some allocation parameters are deemed insignificant. The third column of this table states the starting value of the respective allocation parameters. Recall that these estimates are solely based on the nesting structure provided by the vehicle leasing company. Note that most estimated values do not differ much from the provided starting value of the parameter. Interestingly enough, estimates of allocation parameters that differ much from the corresponding starting value are typically significant. Lastly, note that for each of the alternatives at least one of the allocation parameters is significant, indicating inclusion in one of the nests is indicated by the data.

Table 8. Statistics on the allocation parameters α. In α_{ik}, *i* denotes the alternative and k the corresponding nest. All values for which statistics are not displayed were fixed at run-time. These parameters were estimated with infinite standard error at first. On the second run they were fixed at the produced output value of the first run.

Allocation Parameter	**Estimate**	**Start. Value**	**Std. Error**	***t*-stat**	***p*-value**
α_{Audi-C}	0.367	0.23	0.222	1.65	0.10
α_{Audi-D}	0.378	0.41	0.243	1.56	0.12
α_{Audi-E}	0.111	0.17	FIXED		
$\alpha_{Audi-SUV}$	0.145	0.19	FIXED		
α_{BMW-C}	0.103	0.20	0.105	0.98	0.33
α_{BMW-D}	0.665	0.34	0.055	12.18	0.00
α_{BMW-E}	0.091	0.18	0.095	0.96	0.34
$\alpha_{BMW-SUV}$	0.141	0.28	0.042	3.35	0.00
$\alpha_{Citroen-C}$	0.147	0.29	0.109	1.35	0.18
$\alpha_{Citroen-LCV}$	0.212	0.42	FIXED		
$\alpha_{Citroen-MPV}$	0.641	0.29	0.165	3.88	0.00
α_{Ford-B}	0.096	0.19	0.071	1.34	0.18
α_{Ford-C}	0.153	0.28	0.100	1.54	0.12
α_{Ford-D}	0.485	0.16	0.122	3.97	0.00
$\alpha_{Ford-LCV}$	0.190	0.22	0.061	3.13	0.00
$\alpha_{Ford-MPV}$	0.076	0.15	0.019	4.08	0.00
$\alpha_{Mercedes\text{-}Benz-D}$	0.685	0.64	0.051	13.33	0.00
$\alpha_{Mercedes\text{-}Benz-E}$	0.102	0.17	0.068	1.50	0.13
$\alpha_{Mercedes\text{-}Benz-SUV}$	0.213	0.19	0.042	5.01	0.00
$\alpha_{Nissan-SUV}$	1.000	1.00	FIXED		
α_{Opel-C}	0.365	0.72	FIXED		
α_{Opel-D}	0.635	0.28	FIXED		
$\alpha_{Other-B}$	0.231	0.41	0.104	2.23	0.03
$\alpha_{Other-C}$	0.111	0.22	0.038	2.89	0.00
$\alpha_{Other-SUV}$	0.658	0.37	FIXED		
$\alpha_{Peugeot-C}$	0.379	0.49	FIXED		
$\alpha_{Peugeot-D}$	0.351	0.12	0.074	4.74	0.00
$\alpha_{Peugeot-LCV}$	0.101	0.20	FIXED		
$\alpha_{Peugeot-SUV}$	0.168	0.19	0.030	5.62	0.00
$\alpha_{Renault-B}$	0.336	0.41	0.113	2.98	0.00
$\alpha_{Renault-C}$	0.184	0.16	0.097	1.91	0.06
$\alpha_{Renault-LCV}$	0.378	0.31	0.196	1.92	0.05
$\alpha_{Renault-MPV}$	0.102	0.12	0.023	4.43	0.00

$\alpha_{Skoda-C}$	0.338	0.67	0.078	4.32	0.00
$\alpha_{Skoda-D}$	0.662	0.33	0.078	8.44	0.00
$\alpha_{Volkswagen-C}$	0.225	0.38	0.148	1.52	0.13
$\alpha_{Volkswagen-D}$	0.211	0.34	0.068	3.11	0.00
$\alpha_{Volkswagen-LCV}$	0.206	0.13	0.060	3.41	0.00
$\alpha_{Volkswagen-MPV}$	0.357	0.15	FIXED		
$\alpha_{Volvo-C}$	0.071	0.14	FIXED		
$\alpha_{Volvo-D}$	0.131	0.24	0.077	1.70	0.09
$\alpha_{Volvo-E}$	0.086	0.17	0.088	0.98	0.33
$\alpha_{Volvo-SUV}$	0.712	0.45	0.093	7.64	0.00

Since this research aims to construct predictive models rather than descriptive models, performance on test data is important. Table 11 states the achieved accuracy of each model on test data. Recall that the test data consist of the most recent 10% of switches. In addition to the achieved accuracy of the aforementioned models, accuracy of a benchmark model is stated as well. This benchmark model assumes loyalty of drivers. It assumes that a driver chooses the same make as he or she was driving prior to his or her current vehicle, hence this model solely copies the variable make to the choice variable new make. Observe that, in the test data, almost 45% of drivers remain loyal to their previously driven make. Interestingly enough, the nested model generalizes worst on unseen data, whilst best performance of the cross-nested model on training data directly relates to best performance on unseen data. Note that an accuracy value of 1% translates to roughly 69 correctly predicted switches.

Table 9. Performance of all logit models on test data. Each model predicts the most likely next make of vehicle given previous contract attributes. The test data contain 6,896 matched contracts.

Model	**Accuracy**
MNL	48.26%
nested MNL	47.96%
cross-nested MNL	48.39%
benchmark	44.90%

Next, the relevance and influence of the explanatory variables is discussed. Note that insignificant parameters are also displayed. These parameters serve two purposes. Firstly, insignificance of parameters could serve an explanatory purpose regarding inclusion in models. Secondly, exclusion of these parameters causes both the final log-likelihood of models and the achieved accuracy on test data to worsen. Therefore, all insignificant estimates were maintained and used for prediction. Since nearly 45% of drivers stays loyal to their make, it is expected that parameters regarding make are of great importance. Table 12 portrays the estimates for these parameters. Per state of make only the two parameters with the highest estimated value are shown. Note that for nearly all these βs, the one indicating make loyalty has the highest value. Only the parameters $\beta_{make\ BMW\ BMW}$ and $\beta_{make\ Renault\ Renault}$ are not the estimate with the highest value for that particular state of the variable. In addition, note that all these estimates are significant. This table portrays the estimates of the multinomial logistic regression model. Similar phenomena are observed for both the nested and cross-nested models.

Table 10 Importance of previously driven make of vehicle for the multinomial logit model. The two βs with the highest value are shown per previous make. The parameter $\beta_{make\ Audi\ Audi}$ is omitted, since this parameter is fixed at zero. Note that insignificant βs are not considered. All alternatives are stated in italics.

Parameter	**Estimate**	**Std. Error**	***t*-stat**	***p*-value**
$\beta_{make\ Audi\ Peugeot}$	-0.85	0.248	-3.44	0.00
$\beta_{make\ BMW\ Opel}$	1.02	0.226	4.49	0.00
$\beta_{make\ BMW\ BMW}$	0.92	0.104	8.81	0.00
$\beta_{make\ Citroen\ Citroen}$	3.32	0.434	7.65	0.00

$\beta_{make\ Citroen\ Peugeot}$	2.30	0.261	8.80	0.00
$\beta_{make\ Ford\ Ford}$	2.70	0.165	16.39	0.00
$\beta_{make\ Ford\ Citroen}$	2.12	0.430	4.93	0.00
$\beta_{make\ Merc\text{-}B\ Merc\text{-}B}$	1.54	0.122	12.58	0.00
$\beta_{make\ Merc\text{-}B\ Opel}$	0.56	0.252	2.22	0.03
$\beta_{make\ Nissan\ Nissan}$	3.35	0.363	9.23	0.00
$\beta_{make\ Nissan\ Opel}$	1.50	0.328	4.56	0.00
$\beta_{make\ Opel\ Opel}$	2.88	0.225	12.79	0.00
$\beta_{make\ Opel\ Peugeot}$	1.52	0.249	6.12	0.00
$\beta_{make\ Other\ Other}$	1.85	0.134	13.82	0.00
$\beta_{make\ Other\ Citroen}$	1.51	0.430	3.52	0.00
$\beta_{make\ Peugeot\ Peugeot}$	2.53	0.247	10.26	0.00
$\beta_{make\ Peugeot\ Citroen}$	1.76	0.429	4.11	0.00
$\beta_{make\ Renault\ Citroen}$	2.54	0.428	5.92	0.00
$\beta_{make\ Renault\ Renault}$	2.07	0.186	11.11	0.00
$\beta_{make\ Skoda\ Skoda}$	2.63	0.190	13.83	0.00
$\beta_{make\ Skoda\ Nissan}$	1.57	0.385	4.08	0.00
$\beta_{make\ Volkswagen\ Volkswagen}$	1.41	0.107	13.16	0.00
$\beta_{make\ Volkswagen\ Peugeot}$	0.57	0.240	2.39	0.02

Regarding the remaining explanatory variables, observe that estimates of the parameters included in all three models portray similar characteristics. Most parameters considered important in one model, prove important for the other two models as well. Recall that the leasing company segments drivers by country and client segment. The estimates regarding these parameters indicate that segmentation by country is often more relevant than by client segment. Noteworthy, the estimates regarding the state Norway of parameter country are all significant for prediction. This is the only branch of the company for which true data is provided. The categorical variable *fuel type* seems to play an important predictive role, even though that the distribution of the states of this parameter is dis-balanced. In addition, the parameter vehicle segment proves its relevance depending on the state of the variable. It seems that the states chosen to serve as nests prove slightly more predictive power than the excluded states. Observe that nearly all states of the variable *body style* do not add much explanatory power to the model. Only the states *Car Van* and *Delivery Van* add some strength to the model. Regarding the numerical variables included, the variable *catalogue price* proves its explanatory power for all alternatives except Volvo. In addition, the parameters *mileage month* and *standard discount percentage* influence most drivers when choosing a new make of vehicle.

5. Discussion

Logit models have historically proven successful for analyzing and predicting mode choice. This paper presented a mode choice case study by which performance of several logit models, each relaxing correlation assumptions more, was compared. It was expected that the models in which correlation assumptions were relaxed would perform best. The presented results did indeed indicate better performance of such models. More precisely, these models fit training data better and generalize better on unseen data, obtaining higher accuracy on test data.

The standard multinomial logit model was estimated first. Due to a guarantee of convergence, such models prove extremely convenient when relaxing assumptions on the distribution of error terms. When defining a nesting structure the estimated parameters of the standard model were taken as starting values for the parameters of the nested models. The presented results indicate that the leasing company defined nesting structure proves accurate, with almost all estimated nest parameters being significant. Whereas performance of the nested logistic regression model was slightly better than that of the standard model on training data, it generalized worse on unseen data. Performance of the cross-nested model was better on both training and test data. Considering that the increase in log-likelihood of the nested model relative to the standard model was negligible, both an increase in log-likelihood and accuracy on test data was observed for the cross-nested model. Even though the nesting structure used for the cross-nested model was considered accurate, the training data did not support the entire allocation of makes across nests. Relaxing assumptions on error term distribution does indeed improve performance of models, albeit slight. From estimation results it can be concluded that correlations across alternatives indeed exist, hence assigning each make one or multiple nests is justified.

In conclusion, relaxing assumptions on error term distribution allows for better capturing of correlations across alternatives. For the leasing company to use the models discussed in this thesis, the company should keep track of the drivers of their vehicles. Doing so will allow for accurate matching of contracts, whilst concurrently enhancing predictive power by addition of explanatory variables.

References

Ben-Akiva, M. and Bierlaire, M. (2003). Discrete choice models with applications to depar- ture time and route choice. In Handbook of transportation science, pages 7–37. Springer.

Bhat, C. R. and Sardesai, R. (2006). The impact of stop-making and travel time reliability on commute mode choice. Transportation Research Part B: Methodological, 40(9):709– 730.

Bierlaire, M. (2015). Bisonbiogeme: estimating a first model. Technical report.

Chu, Y.-L. (2009). Work departure time analysis using dogit ordered generalized extreme value model. Transportation Research Record: Journal of the Transportation Research Board, (2132):42–49.

De Jong, G., Daly, A., Pieters, M., Vellay, C., Bradley, M., and Hofman, F. (2003). A model for time of day and mode choice using error components logit. Transportation Research Part E: Logistics and Transportation Review, 39(3):245–268.

Ding, C., Mishra, S., Lin, Y., and Xie, B. (2014). Cross-nested joint model of travel mode and departure time choice for urban commuting trips: Case study in maryland– washington, dc region. Journal of Urban Planning and Development, 141(4):04014036.

Hess, S., Fowler, M., Adler, T., and Bahreinian, A. (2012). A joint model for vehicle type and fuel type choice: evidence from a cross-nested logit study. Transportation, 39(3):593– 625.

Koppelman, F. S. and Bhat, C. (2006). A self instructing course in mode choice modeling: multinomial and nested logit models.

McFadden, D. (1978). Modeling the choice of residential location. Transportation Research Record, (673).

McFadden, D. (2001). Economic choices. American economic review, 91(3):351–378.

McFadden, D. et al. (1973). Conditional logit analysis of qualitative choice behavior.

Train, K. E. (2009). Discrete choice methods with simulation. Cambridge university press.

Vovsha, P. (1997). Application of cross-nested logit model to mode choice in tel aviv, israel, metropolitan area. Transportation Research Record: Journal of the Transportation Research Board, (1607):6–15.