MPRA
Munich Personal RePEc Archive

Airport and Access Mode Choice in Germany: A Generalized Neural Logit Model Approach

Gelhausen, Marc Christopher

German Aerospace Center (DLR), Air Transport and Airport Research

August 2006

Online at https://mpra.ub.uni-muenchen.de/4236/
MPRA Paper No. 4236, posted 24 Jul 2007 UTC

AIRPORT AND ACCESS MODE CHOICE IN GERMANY:
A GENERALIZED NEURAL LOGIT MODEL APPROACH

Marc Ch. Gelhausen
German Aerospace Center (DLR)
Air Transport and Airport Research

1. INTRODUCTION

The purpose of this paper is to present a novel approach in discrete choice modelling based on artificial neural networks and is an excerpt of the doctoral thesis of the author. Research is mainly focussed on the distribution assumptions of the random component of the utility function to model correlations among alternatives in the choice set due to unobserved alternative attributes. Only a few works deal with the subject of nonlinear utility functions partly because of the difficulties arising in determining a priori the form of nonlinearity of the utility function. Box-Cox and Box-Tukey transformations (see i.e. Maier et al. 1990, pp. 126ff.) enable to model some limited forms of nonlinear utility functions.

Some research has been done in combining discrete choice models with artificial neural networks to model a nonparametric nonlinear utility function. Bentz and Merunka (Bentz et al. 2000) and Gelhausen (Gelhausen 2003) describe two different ways to build a logit-model as an artificial neural network. This approach shows significant better empirical results than a standard logit-model with a linear utility function (Bentz et al. 2000; Hruschka et al. 2002; Probst 2002). An implementation of a nested logit-model with an arbitrary nesting structure is possible (Wilken et al. 2005), but because of the complexity of the resulting network severe estimation and performance problems occur.

The Generalized Neural Logit-Model enables to model a nonparametric nonlinear utility function and arbitrary correlations among alternatives in the choice set due to unobserved attributes. Correlations among alternatives are modelled similar to the dogit-model. An efficient implementation of the Generalized Neural Logit-Model is possible as it is close to standard artificial neural networks.

The outline of this paper is as follows:

Chapter 2 explains the concept of alternative groups in discrete choice models as already introduced in Gelhausen et al. (2006). This concept facilitates complexity reduction and the development of a model, which is applicable to alternatives outside the estimation data set.

Chapter 3 describes the theory of the Generalized Neural Logit-Model and its implementation as artificial neural network.

The Generalized Neural Logit-Model is evaluated empirically by means of an application case in chapter 4. The chosen application case is airport and access mode choice in Germany and the benchmark is a nested logit approach (Gelhausen et al. 2006).

The paper ends with a summary and conclusion.

2. GROUPING OF ALTERNATIVES IN DISCRETE CHOICE MODELS

The fundamental hypothesis of discrete choice models is the assumption of individual utility maximisation. Alternatives are evaluated by means of an utility function and the one with the highest utility is supposed to be chosen. From an external point of view the utility of an alternative for a specific individual is a random variable, so that the utility U_i for alternative i is composed of a deterministic component V_i and a random component ε_i (Maier et al. 1990, p. 100):

(2.01) $$U_i = V_i + \varepsilon_i$$

The random component of the utility function is introduced for various reasons, i.e. a lack of observability of the relevant attributes of the alternatives or their incomplete measurability (Maier et al. 1990, pp. 98f.).

From an external point of view, only evidence in terms of the probability of an alternative being the one with the highest utility can be given, because of the random component in the utility function. Specific discrete choice models differ in terms of their assumptions of the random component. The most prominent member of this class of models is the logit-model with independently and identically distributed random components. The choice probability of an alternative i is computed as (Train 2003, p. 40):

(2.02) $$P(a_i = a_{opt}) = \frac{e^{\mu V_i}}{\sum_j e^{\mu V_j}}$$

As a consequence of the independently and identically distributed random components of the utility functions the ratio of two choice probabilities is only dependent on the utility of those two alternatives (Ben-Akiva et al. 1985, p. 108):

(2.03) $$\frac{P(a_i = a_{opt})}{P(a_j = a_{opt})} = \frac{\frac{e^{\mu V_i}}{\sum_k e^{\mu V_k}}}{\frac{e^{\mu V_j}}{\sum_k e^{\mu V_k}}} = \frac{e^{\mu V_i}}{e^{\mu V_j}}$$

This property of the logit-model is called “Independence from Irrelevant Alternatives” (IIA) and it is both a weakness and a strength of the model. Due to the distribution assumptions of the random component of the utility function it is not possible to model correlations among alternatives owing to unobserved factors. A major advantage of the IIA-property is the possibility to estimate the model parameters, excluding alternative-specific variables, on a subset of the alternatives (McFadden 1974, p. 113; McFadden 1978, pp. 87ff.; Ortuzar et al. 2001, pp. 227f.; Train 2003, pp. 52f.) and the possibility of an evaluation of new alternatives without the need to re-estimate alternative-unspecific model parameters (Domencich et al. 1975, pp. 69f.). The problem of estimating alternative-specific variables from a subset of alternatives will be discussed below.

The nested logit-model relaxes the IIA-restriction to some extent without losing the closed-form expression of the choice probabilities. For this purpose the random component in (2.01) is split up into a part ε_i^a, which varies over all alternatives i and a part ε_k^c, which is identical for all alternatives of a nest k (Maier et al. 1990, pp. 154f.):

$$U_i = V_i + \varepsilon_i^a + \varepsilon_k^c \tag{2.04}$$

It is possible to model correlations due to unobserved factors among subsets of the alternatives, so that the choice set is partitioned into clusters with highly correlated alternatives. (2.05) is an example of a covariance matrix for four alternatives partitioned into two clusters with the first two belonging to cluster one and the last two assigned to cluster two.

$$\Omega = \begin{bmatrix} \sigma_{11}^2(\mu_1^c) & \sigma_{12}^2(\varepsilon_1^c) & 0 & 0 \\ \sigma_{21}^2(\varepsilon_1^c) & \sigma_{22}^2(\mu_1^c) & 0 & 0 \\ 0 & 0 & \sigma_{33}^2(\mu_2^c) & \sigma_{34}^2(\varepsilon_2^c) \\ 0 & 0 & \sigma_{43}^2(\varepsilon_2^c) & \sigma_{44}^2(\mu_2^c) \end{bmatrix} \tag{2.05}$$

Each cluster k is characterized by an individual scale parameter μ_k^c and an identical non-negative covariance for all alternatives i within a cluster k. Alternatives of different clusters are assumed not to be correlated.

For technical reasons the choice probabilities $P(a_i = a_{opt})$ are decomposed into an unconditional choice probability $P(c_k = c_{opt})$ that cluster k is chosen, and a conditional choice probability $P(a_i = a_{opt} \mid a_i \in c_k)$, that alternative i from cluster k is chosen (Maier et al. 1990, p. 156):

$$P(a_i = a_{opt}) = P(a_i = a_{opt} \mid a_i \in c_k) * P(c_k = c_{opt}) \tag{2.06}$$

The conditional choice probabilities comply with the logit-model and the choice set is restricted to the alternatives of the appropriate nest. The choice probability of a nest k is determined by its maximum utility V_k^c (Maier et al. 1990, p. 157):

$$V_k^c = \frac{1}{\mu} \ln \sum_{i \in k} e^{\mu V_i} \tag{2.07}$$

The choice probability of an alternative i in nest k can be written as (Maier et al. 1990, p. 158):

$$(2.08) \qquad P(a_i = a_{opt}) = \frac{e^{\mu V_i}}{\sum_{j \in k} e^{\mu V_j}} * \frac{e^{\mu_k^c V_k^c}}{\sum_{l} e^{\mu_l^c V_l^c}}$$

The hierarchical structure of (2.08) does not imply a sequential decision process. An extension to more than two levels is possible (see i.e. Ben-Akiva et al. 1985, pp. 291ff.).

In the nested logit-model the IIA-property holds only for two alternatives of the same cluster:

$$(2.09) \qquad \left. P(a_1 = a_{opt} \mid a_1 \in c_1) * P(c_1 = c_{opt}) \middle/ P(a_2 = a_{opt} \mid a_2 \in c_1) * P(c_1 = c_{opt}) \right.$$

$$= \left. \frac{e^{\mu V_1}}{\sum_{j \in c_1} e^{\mu V_j}} * \frac{e^{\mu_1^c V_1^c}}{\sum_{l} e^{\mu_l^c V_l^c}} \middle/ \frac{e^{\mu V_2}}{\sum_{j \in c_1} e^{\mu V_j}} * \frac{e^{\mu_1^c V_1^c}}{\sum_{l} e^{\mu_l^c V_l^c}} \right.$$

$$= \frac{e^{\mu V_1}}{e^{\mu V_2}}$$

The ratio of the choice probabilities for two alternatives of different clusters depends on the characteristics of all alternatives of those two clusters:

$$(2.10) \qquad \left. P(a_1 = a_{opt} \mid a_1 \in c_1) * P(c_1 = c_{opt}) \middle/ P(a_2 = a_{opt} \mid a_2 \in c_2) * P(c_2 = c_{opt}) \right.$$

$$= \left. \frac{e^{\mu V_1}}{\sum_{j \in c_1} e^{\mu V_j}} * \frac{e^{\mu_1^c V_1^c}}{\sum_{l} e^{\mu_l^c V_l^c}} \middle/ \frac{e^{\mu V_2}}{\sum_{j \in c_2} e^{\mu V_j}} * \frac{e^{\mu_2^c V_2^c}}{\sum_{l} e^{\mu_l^c V_l^c}} \right.$$

$$= \left. \frac{e^{\mu V_1}}{\sum_{j \in c_1} e^{\mu V_j}} \middle/ \frac{e^{\mu V_2}}{\sum_{j \in c_2} e^{\mu V_j}} \right. * \frac{e^{\mu_1^c V_1^c}}{e^{\mu_2^c V_2^c}}$$

As the nested logit-model lacks the IIA-property for some pairs of alternatives, model estimation on a subset of the choice set equal to the logit-model is not possible.

If it is feasible to form groups of at least approximately similar clusters and to assign an identical covariance matrix for all clusters of the same group, an estimation of alternative-unspecific model-parameters equal to the logit-model on a subset of alternatives is possible. Each group of clusters must be represented by at least one member in this subset to enable the estimation of all cluster-specific scale parameters. (2.11) shows a covariance-matrix for six alternatives belonging to three groups, with two alternatives per group. Figure 2.1 shows the dependence between a group and a cluster for this example.

$$\Omega = \begin{bmatrix} A & 0 & 0 & 0 & 0 & 0 \\ 0 & B & 0 & 0 & 0 & 0 \\ 0 & 0 & B & 0 & 0 & 0 \\ 0 & 0 & 0 & C & 0 & 0 \\ 0 & 0 & 0 & 0 & A & 0 \\ 0 & 0 & 0 & 0 & 0 & C \end{bmatrix} \qquad (2.11)$$

The letters A, B and C represent the covariance structure of a cluster. Same letters indicate an equal covariance structure for different clusters. Figure 2.01 illustrates the assignment of clusters to groups.

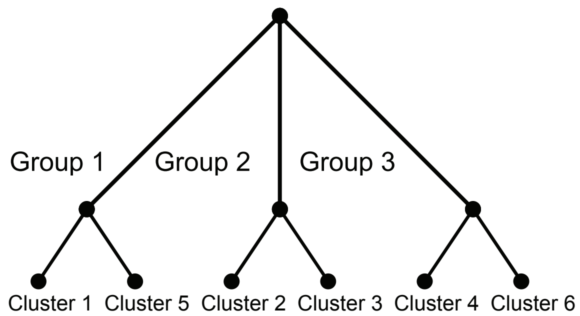


Figure 2.01: Dependence between Clusters and Groups

If identical alternative-specific model-parameters, especially alternative-specific constants, can be assumed reasonably well for different clusters of the same group, an estimation of all model-parameters is feasible on a subset of all alternatives as described above.

Applying the concept of grouping in the logit-model is possible, however, serves only to estimating alternative-specific variables, as there are no different scale parameters due to independently and identically distributed random components in the utility function.

The main advantage of this approach does not only lie in the reduction of computational costs for very large choice sets, as many econometric software packages limit the maximum number of clusters and alternatives for nested logit estimations, but also in a better way of developing a more generally applicable choice model beyond the alternatives of the estimation data set, i.e. in the context of scenario analysis.

A less popular member of discrete choice models is the dogit-model. Correlations among alternatives in the choice set are modelled by means of a functional combination of the utility functions of each alternative with an alternative-specific parameter θ_i (Gaudry et al.1979, p. 105):

$$P(a_i = a_{opt}) = \frac{e^{\mu V_i} + \theta_i \sum_j e^{\mu V_j}}{\left(1 + \sum_j \theta_j\right) \sum_j e^{\mu V_j}} \tag{2.12}$$

Dogit- and logit-model are equal for all θ_l being zero so that the IIA-property holds for arbitrary pairs of alternatives. The vector of parameters θ describes to what extend the IIA-property does not hold.

In some empirical cases the dogit-model is superior in terms of model fit to a logit approach (see i.e. McCarthy 1997). However, the IIA-property does not hold in a systematic way in a genuine dogit-model with a nonzero vector θ so that the aforementioned concept of alternative groups is not applicable.

3. THE GENERALIZED NEURAL LOGIT-MODEL

3.1 Theory of the Generalized Neural Logit-Model

In the Generalized Neural Logit-Model the distribution assumptions of the random component of the utility function are the same as those of the logit-model. Correlations among alternatives due to unobserved attributes are modelled by means of a combination of the utility functions of each alternative. This approach shows some similarities to the dogit-model, however, it offers more flexibility in terms of modelling correlations among alternatives.

An essential part of the Generalized Neural Logit-Model is a linear combination of the utility functions of each alternative:

(3.01)
$$V_i^{LK} = \sum_{j \in A_P^{LK}} \gamma_{ij} * V_j \quad i, j \in A_p^{LK}$$

where

γ_{ij}: Coefficient of the linear combination of the alternatives i and j

Alternatives in a subset A_p^{LK} of the total choice set A^l are correlated. The correlation structure among alternatives is modelled by means of a hierarchy of utility functions. This approach shows some similarities to the nested logit-model. Due to the linear dependence between utility functions of different levels a two-stage hierarchy is sufficient. The choice probabilities are computed the same way as in the logit-model with (3.01) being the utility function:

(3.02)
$$P(a_i = a_{opt}) = \frac{e^{\mu V_i^{LK}}}{\sum_j e^{\mu V_j^{LK}}}$$

The Generalized Neural Logit-Model belongs to the class of General Extreme Value-Models, so that utility maximizing behaviour is modelled. The derivation of the model is identical to the logit-model with (3.01) being the utility function (Train 2003, p. 97f):

The function G is defined as:

(3.03)
$$G = \sum_j Y_j$$

where

(3.04)
$$Y_j = e^{\mu V_j^{LK}}$$

Inserting G and its first derivation G_i in

$$(3.05) \qquad P(a_i = a_{opt}) = \frac{Y_i G_i}{G}$$

produces

$$(3.06) \qquad P(a_i = a_{opt}) = \frac{Y_i}{\sum_j Y_j}$$

(3.06) equals (3.02).

The definition of the subsets A_p^{LK} depends on the correlations among the alternatives to be modelled. Four cases are distinguished:

- No correlations (logit-model)
- Correlations among all alternatives in the choice set
- Correlations among alternatives in disjoint clusters
- Limited correlations among all alternatives in the choice set

No correlations (logit-model)

Each subset A_p^{LK} equals exact one alternative and all coefficients γ_{ij} are set to a value of one. The IIA-property holds for arbitrary combinations of alternatives.

Correlations among all alternatives in the choice set

There is only one subset A_p^{LK}, which equals the total choice set. The coefficients can take any value. The IIA-property does not hold for any combination of alternatives.

Correlations among alternatives in disjoint clusters

In this case alternatives are grouped in disjoint clusters A_p^{LK} similar to the nested logit approach to model arbitrary correlations among alternatives in each subset. The IIA-property does only hold on the cluster level. The aforementioned concept of alternative groups can be applied. Clusters of the same group have an identical matrix of coefficients of linear combination instead of an identical covariance matrix:

$$(3.07) \qquad \Omega^g = \begin{bmatrix} \gamma_{11}^g & \cdots & \gamma_{1J}^g \\ \vdots & \ddots & \vdots \\ \gamma_{I1}^g & \cdots & \gamma_{IJ}^g \end{bmatrix}$$

where

γ_{ij}^{g} : Coefficient of linear combination of the alternatives i and j of a cluster of group g

Limited correlations among all alternatives in the choice set

In this case all alternatives may be correlated, so that the IIA-property does not hold for any pair of alternatives, but the coefficients of linear combination underlie a systematic structure, which enables model estimation on a subset of the complete choice set. Structural groups of alternatives and clusters can be identified according to the logit- and nested logit-model. Their definition is problem-dependent. In this study correlations among alternatives of the same cluster and between alternatives of different cluster groups are considered. Therefore this approach is a medium between case two and case three. The alternative subsets A_p^{LK} are composed of the cluster of the considered alternative and all clusters of different groups.

A group-dependent coefficient of linear combination $\gamma_{im}^{kl\,g}$ is assigned to every alternative i of the cluster k and alternative m of the cluster l. This coefficient is identical for two pairs of alternatives (a, b) and (c, d), if (a, c) and (b, d) belong to different clusters of the same group.

The number of coefficients $\gamma_{im}^{kl\,g}$ equals the square of the number of alternatives on the lowest level of the cluster structure. Because of the identity of certain coefficients every cluster of group g has an identical matrix of structural coefficients of linear combination:

(3.08)
$$\Omega^{g} = \begin{bmatrix} o_{11}^{g1} & \cdots & o_{1J}^{g1} & \cdots & o_{11}^{gG} & \cdots & o_{1M}^{gG} \\ \vdots & \ddots & \vdots & \cdots & \vdots & \ddots & \vdots \\ o_{I1}^{g1} & \cdots & o_{IJ}^{g1} & \cdots & o_{I1}^{gG} & \cdots & o_{IM}^{gG} \end{bmatrix}$$

where

o_{im}^{gG} : Coefficient of linear combination of alternative i of a cluster of group g and alternative m of a cluster of group G

The assignment of elements of matrix (3.08) to the coefficients $\gamma_{im}^{kl\,g}$ results from the grouping of the clusters:

(3.09)
$$o_{im}^{gG} = \gamma_{im}^{kl\,g} \Rightarrow k \in \text{group } g \wedge l \in \text{group } G$$

The coefficients o_{im}^{gG} and $\gamma_{im}^{kl\,g}$ receive the value zero in the case of an assignment of two different clusters to the same group, as this kind of correlations among alternatives are not considered in this study. This fact is pointed out above in the definition of the subsets A_p^{LK}.

Matrix (3.07) is a special case of (3.08).

In the cases one, two and three formula (3.01) contains only non-equivalent alternatives. Two alternatives are equivalent, if they are on the same position in the cluster structure and their clusters belong to the same group. They have the same index m in (3.09).

In case four the above mentioned does not have to hold. By a normalization of the coefficients the value of (3.01) is according to the other three cases only dependent on the quality of non-equivalent linear combined alternatives and independent from the number of equivalent alternatives. For this purpose the coefficients o_{im}^{gG} and $\gamma_{im}^{kl\,g}$ have to be divided by the number equivalent alternatives in (3.01) for model estimation and model application:

(3.10) $$o_{im}^{gG^{Norm}} = \frac{o_{im}^{gG}}{N_m^G}$$

where

N_m^G: Number of equivalent alternatives m of different cluster of the same group G in (3.01)

Model estimation on a subset of the total choice set by means of the IIA-property equal to the logit- and nested logit-model is not possible, because the IIA-property does hold neither on cluster- nor on alternative-level. However, identical coefficients being independent from the number of summands in (3.01) are assigned to certain alternatives because of the grouping of clusters and the normalization of coefficients. According to a random sample model estimation is possible in the case of every cluster group being represented at least with one member.

3.2 The Generalized Neural Logit-Model as Artificial Neural Network

The Generalized Neural Logit-Modell as artificial neural network is composed of different modules, which have to be configured and put together problem-specific:

- Utility functions
- Linear combinations
- Logit-function

Utility function

For the problem considered a three-layer multilayer perceptron is sufficient for universal function approximation (Hornik et al. 1989, pp. 359ff.; Fausett 1994, p. 329).

The activation function of the input and output neurons constitutes the identical function. Hidden neurons have a tangens hyperbolicus function as activation function. The linear part of the utility function is described by means of the direct connections from the input to the output neurons marked in blue. This represents a linear perceptron in itself. The nonlinear part of the utility function is modelled by means of the connections between the neurons marked in black. The input neurons correspond to alternative attributes and the output neurons match the utility of an alternative. Figure 3.01 displays a nonlinear utility function as artificial neural network in an abstract way. The box in the upper part of the figure shows the kind of activation function for the appropriate layer of the artificial neural network.

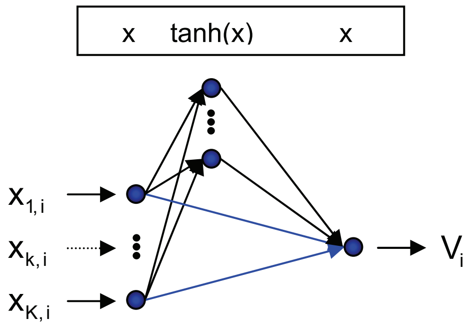


Figure 3.01: Nonlinear Utility Function as Multilayer Perceptron

Linear Combinations

Linear combinations are modelled by means of a two-layer linear perceptron. The identical function is the activation function for the input and output neurons. The Input neurons correspond to utility values of an alternative and the output neurons match the alternative-specific linear combinations of those utility values. The connection weights correspond to the coefficients of linear combination. Connections marked in red are constrained to a value of one.

Figure 3.02 displays the aforementioned four cases of correlation among the alternatives of the choice set as artificial neural network.

A possible grouping structure of clusters is indicated by an appropriate highlighting of the connections in black and blue. A possible grouping structure of clusters is easily identifiable because of the structure of the subsequent figures as only one cluster per group is displayed. An additional indexing of utility functions relating to the clusters is omitted for reasons of clearness.

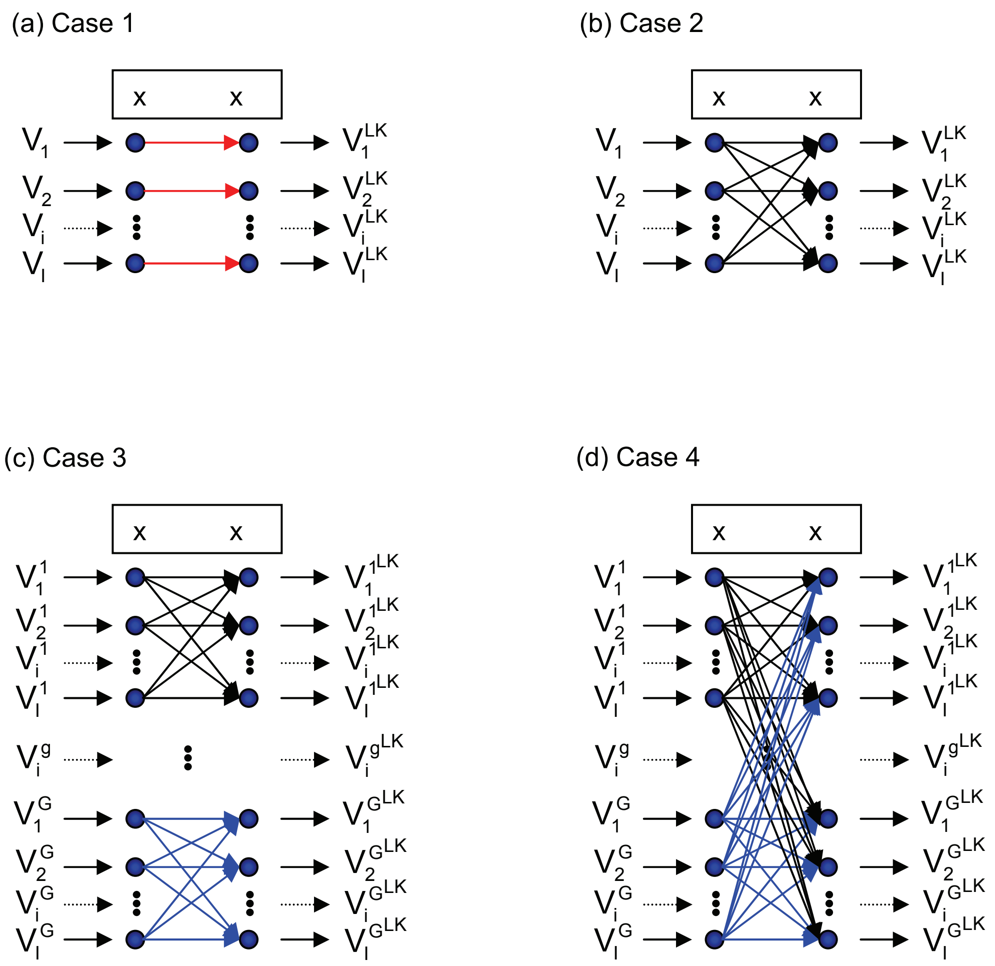


Figure 3.02: Linear Combination as Artificial Neural Network

Logit-Function

The logit-function is modelled by means of a three-layer multilayer perceptron. The activation function of the input neurons is $f(x)=e^x$ and the function $f(x)=1/x$ is assigned to the hidden neurons as activation function. The output neurons posses the identical function as activation function with (3.11) being the net input function. This is a multiplicative combination instead of the usual summation of the inputs into a neuron.

(3.11) $$net_j = \prod_i w_{ij} * o_i$$

where

w_{ij}: Connection weight between neuron of layer i and neuron of layer j
o_i: Output of neuron of layer i

This type of neuron is called "combiner neuron" (NeuroDimension 2005, p. 276f). The input neurons represent the linear combinations of the utility values and the output neurons conform to the choice probabilities of the alternatives in the choice set. Figure 3.03 shows the logit-function as artificial neural network. The connections marked in red are constrained to a value of one.

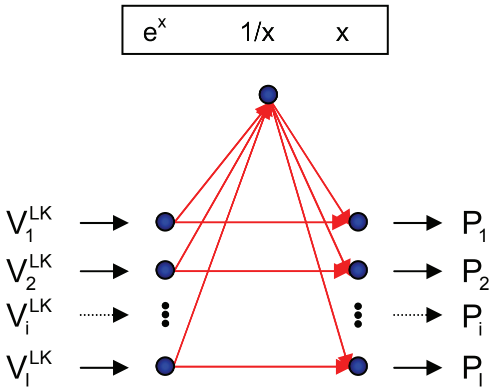


Figure 3.03: Logit-Function as Artificial Neural Network

Generalized Neural Logit-Model

Figure 3.04 to 3.07 display the Generalized Neural Logit-Model for all four aforementioned cases. For reasons of clearness of the implementation alternatives and clusters are grouped although this is not necessary for the cases one and two. The number of utility functions and output neurons equals the number of alternatives on the lowest level of the cluster structure. Identical utility functions in terms of connection structure and weights can be achieved by weight sharing (Bishop 2003, p. 349, LeCun et al. 1989, pp. 542ff., Rumelhart et al. 1986, p. 349) in the case of all alternatives being evaluated by means of the same utility function. In the case of an endogenity of exogenous factors due to unobserved alternative attributes (Bhat 2003, pp. 16f.) a dependence of the utility function in terms of the considered alternative is possible as alternative attributes are evaluated differently dependent on the relevant alternative.

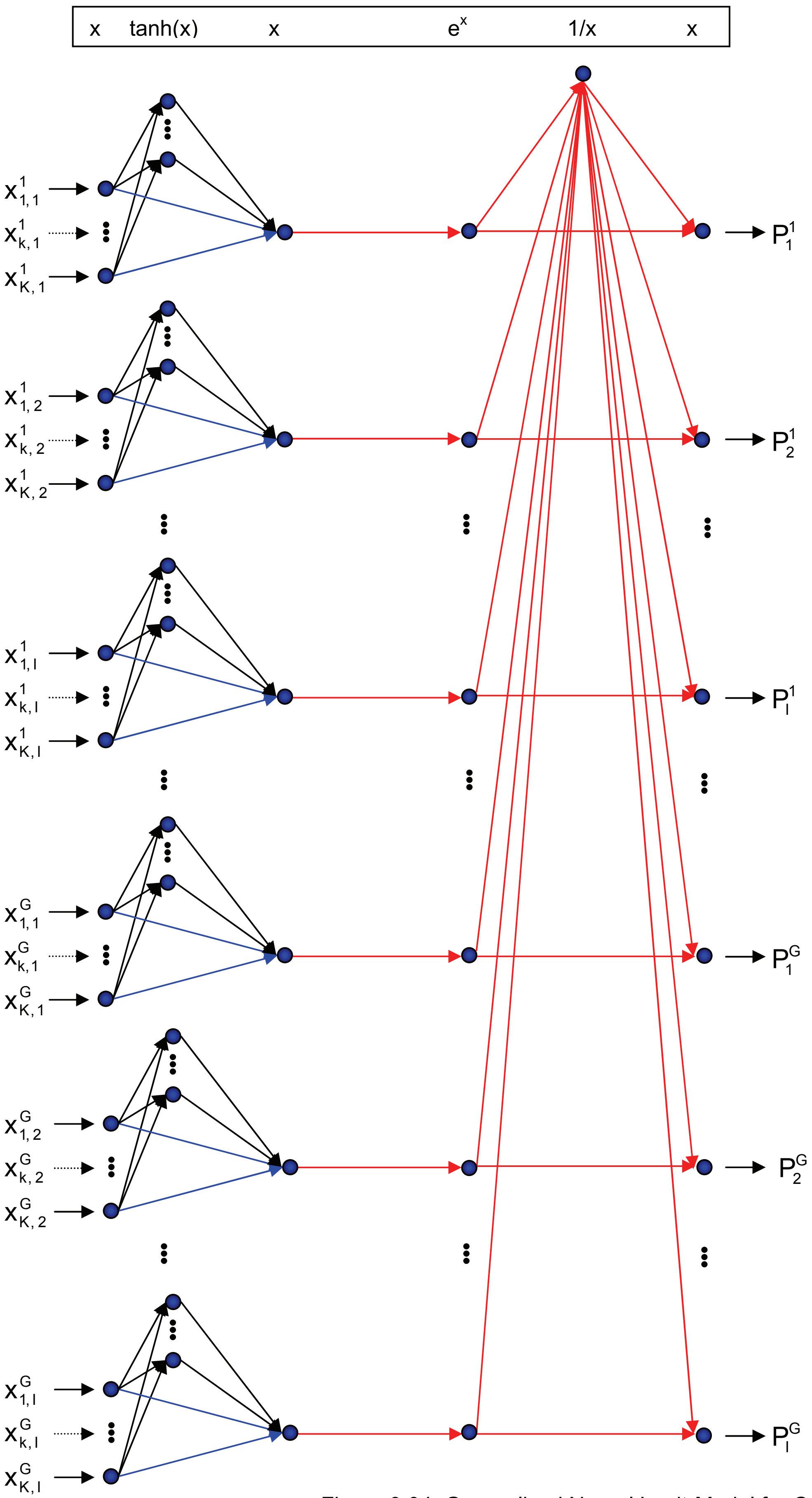


Figure 3.04: Generalized Neural Logit-Model for Case 1

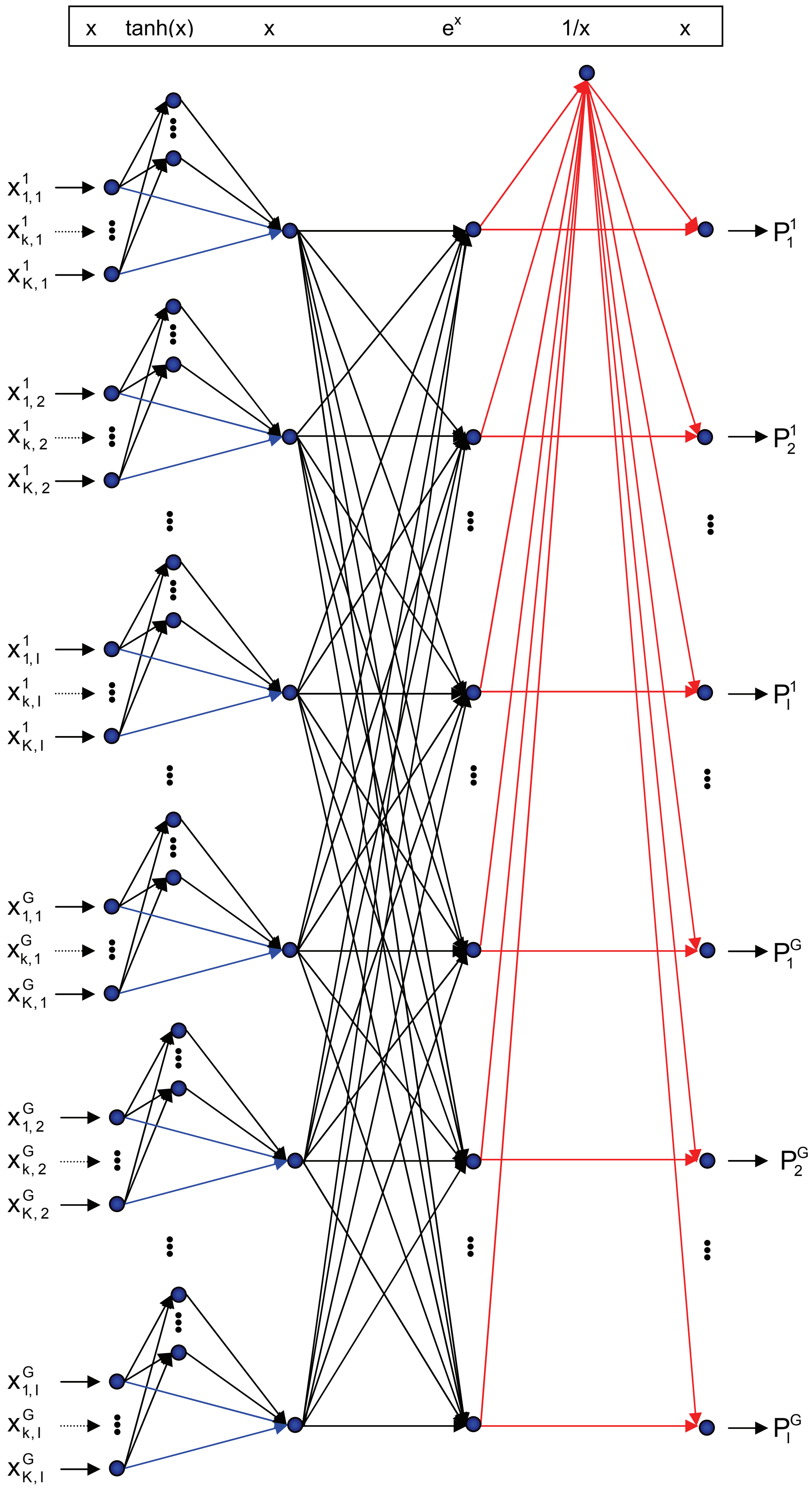


Figure 3.05: Generalized Neural Logit-Model for Case 2

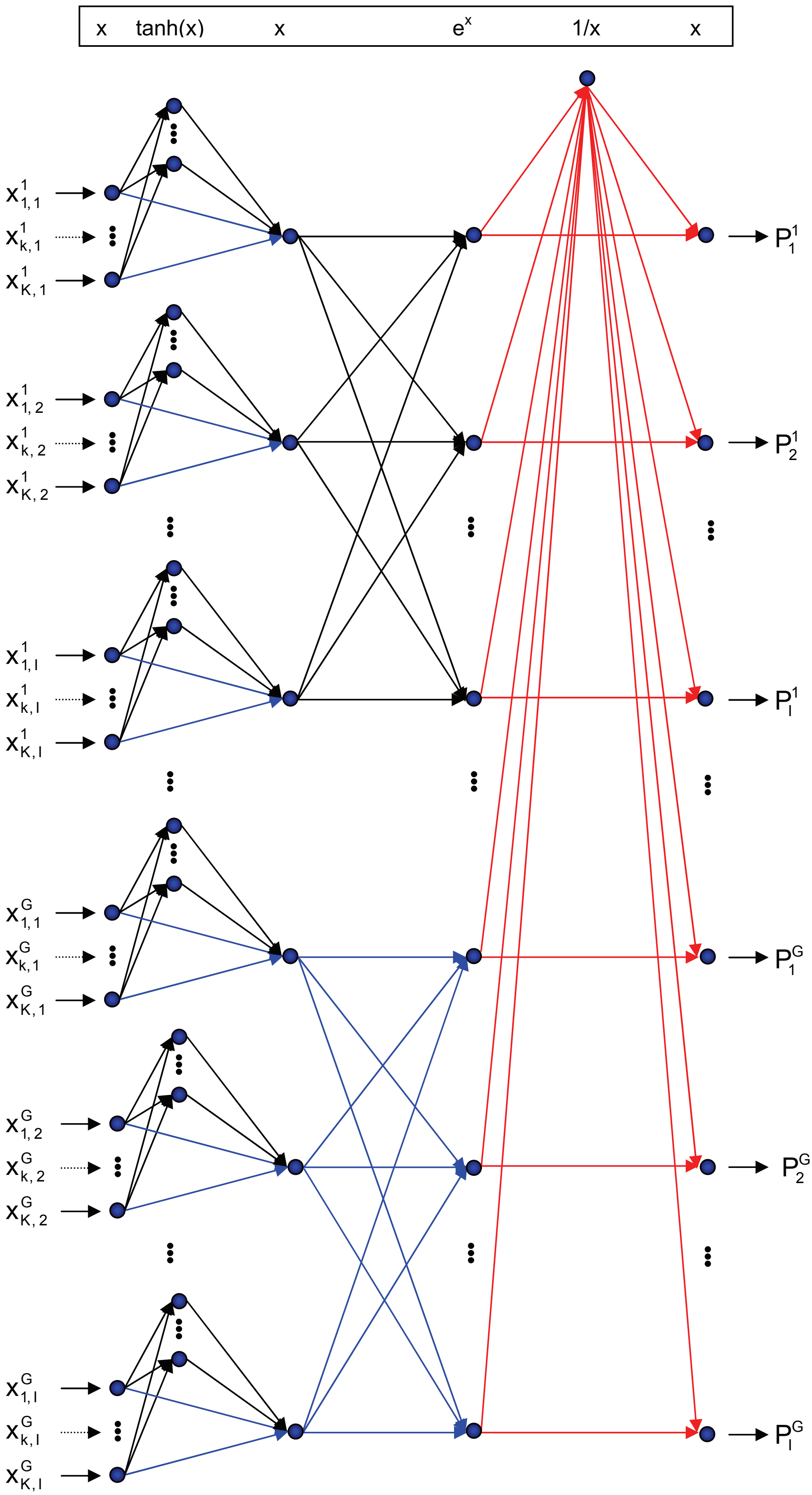


Figure 3.06: Generalized Neural Logit-Model for Case 3

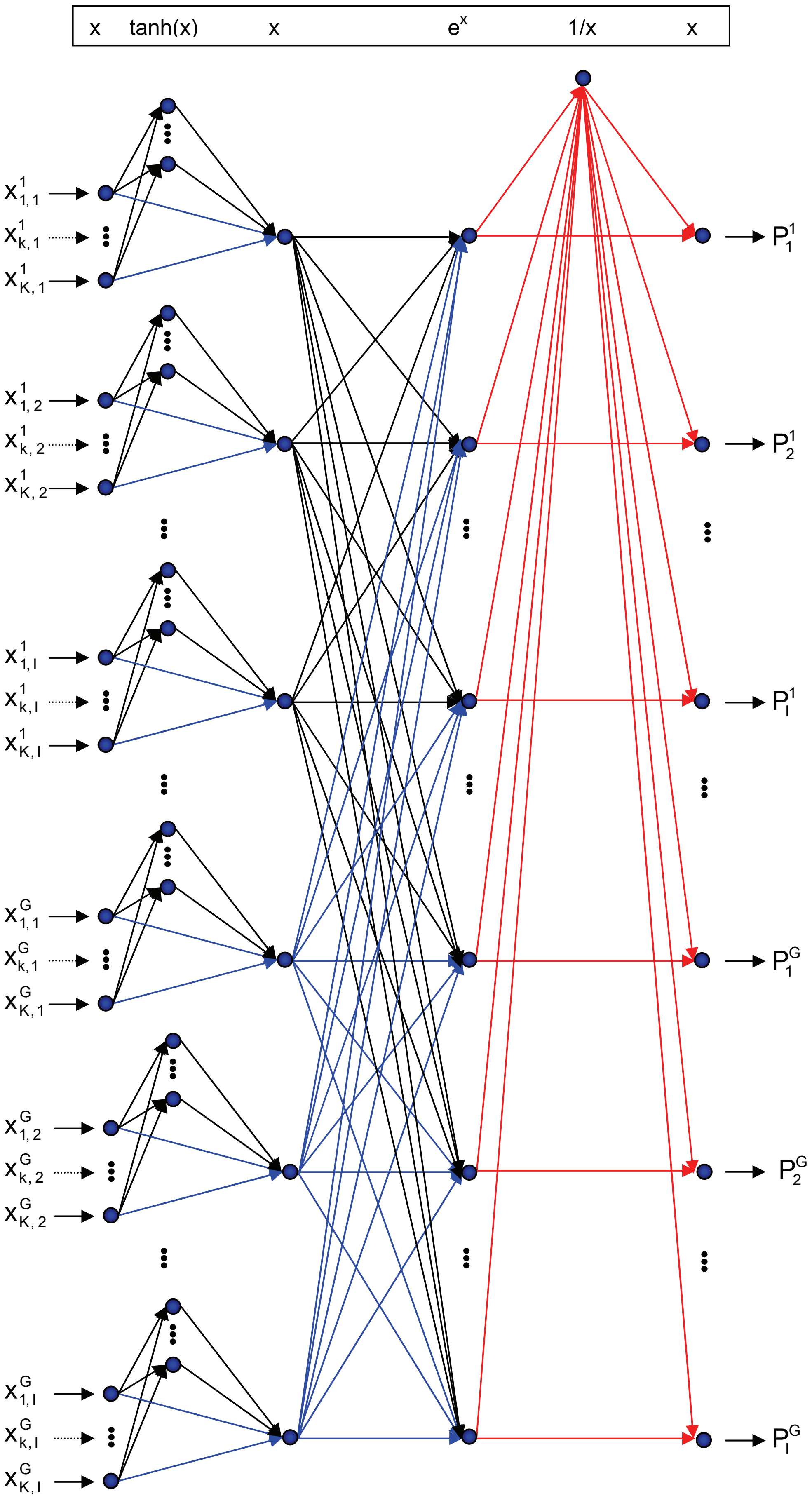


Figure 3.07: Generalized Neural Logit-Model for Case 4

4. APPLICATION CASE: AIRPORT AND ACCESS MODE CHOICE

4.1 Introduction

The chosen application case for an empirical evaluation of the Generalized Neural Logit-Model is airport and access mode choice of air travellers. A nested logit approach with a linear utility function serves as benchmark in terms of model fit. The available database, airport categories, model definition and model estimation of the nested logit-model is discussed in great detail in Gelhausen and Wilken (2006, pp. 10ff.). Only some fundamental facts are introduced briefly below. A full discussion of these issues would be beyond the scope of this chapter. Purpose of this chapter is to discuss both approaches concerning model quality and some conclusions relating to choice behaviour in airport and access mode choice.

Figure 4.02 shows the full alternative set of the database (Gelhausen et al. 2006, p. 11). Only the access mode “car” includes parking at the airport for the duration of the journey. For “kiss and ride” the number of trips is double compared to all other access modes as the car is parked at the trip origin. The “taxi” alternative includes taxis and private bus services operating on demand only. The access mode “bus” contains scheduled public-transit buses. “Urban transit” and “train” are distinguished in terms of the tariff paid. If the tariff of the Deutsche Bahn applies, it is a train, otherwise it is an urban railway (Gelhausen et al. 2006, p. 11).

	Car	Kiss and Ride	Rental Car	Taxi	Bus	Urban Transit	Train
Berlin	x	x	x	x	x	x	
Bremen	x	x	x	x		x	
Dortmund	x	x	x	x	x		
Dresden	x	x	x	x	x	x	
Düsseldorf	x	x	x	x	x	x	x
Erfurt	x	x	x	x	x		
Frankfurt a. M.	x	x	x	x	x	x	x
Frankfurt Hahn	x	x	x	x	x		
Friedrichshafen	x	x	x	x	x	x	
Hamburg	x	x	x	x	x		
Hannover	x	x	x	x	x	x	
Karlsruhe-Baden	x	x	x	x	x		
Köln/Bonn	x	x	x	x	x		
Leipzig/Halle	x	x	x	x	x		x
Lübeck	x	x	x	x	x		
München	x	x	x	x	x	x	
Münster/Osnabrück	x	x	x	x	x		
Niederrhein	x	x	x	x	x		
Nürnberg	x	x	x	x	x	x	
Paderborn/Lippstadt	x	x	x	x	x		
Saarbrücken	x	x	x	x	x		
Stuttgart	x	x	x	x	x	x	

Figure 4.01: Airports and Available Access Modes

According to the length and purpose of a journey different market segments are defined (Gelhausen et al. 2006, pp. 10f.):

- Journeys to domestic destinations, subdivided into private (BRD P) and business (BRD B) trip purpose
- Journeys to European destinations for business trip purpose (EUR B)
- Journeys to European destinations for private short stay reasons up to four days (EUR S)
- Journeys to European destinations for holiday reasons for five days or longer (EUR H)
- Journeys to intercontinental destinations, subdivided into private (INT P) and business (INT B) trip purpose

Figure 4.02 displays the employed alternative attributes and their definitions (Gelhausen et al. 2006, p. 12).

Variable (Abbreviation)	**Definition**
Access Cost (COST)	Cost in € per Person incl. Parking Fees, Double Trip Length
Access Time (TIME)	Time in Minutes, Double Trip Lengtht
Waiting Time (WAIT)	Inverse of the Daily Frequency
Inverse of the Population Density (INVPD)	Inverse of Residents per km^2
Inverse of the Competition on a Direct Flight Connection(COMP)	Inverse of the Number of Alliances and Independent Airlines
Quality of Terminal Access (AAS)	binary (good/bad)
Existence of a Direct Flight Connection (DIRECT)	binary (good/bad)
Frequency of a Direct Flight Connection (DFREQ)	Number Flights per week
Existence of a Low-Cost Connection (LC)	binary (yes/no)
Frequency of a Low-Cost Connection(LCFREQ)	Number Low-Cost Flights per week
Existence of a Charter Flight Connection (CC)	binary (yes/no)
Frequency of a Charter Flight Connection (CCFREQ)	Number Charter Flights per week

Figure 4.02: Definition of Alternative Attributes

Airports are categorized from a demand-oriented point of view by means of Kohonen’s Self-Organizing Maps to form groups of clusters consisting of one airport category and all access modes (Gelhausen et al. 2006, pp. 14ff.). Figure 4.03 shows the chosen attributes for distinguishing airport categories (Gelhausen et al. 2006, p. 12).

Attributes (Abbreviation)	Definition
Number of Domestic Low-Cost Flights (LCBRD)	Flights per Week
Number of Domestic Charter Flights (CCBRD)	Flights per Week
Number of Domestic Full Service Flights (LBRD)	Flights per Week
Number of European Low-Cost Flights (LCEUR)	Flights per Week
Number of European Charter Flights (CCEUR)	Flights per Week
Number of European Full Service Flights (LEUR)	Flights per Week
Number of Intercontinental Low-Cost Flights (LCINT)	Flights per Week
Number of Intercontinental Charter Flights (CCINT)	Flights per Week
Number of Intercontinental Full Service Flights (LINT)	Flights per Week
Number of Domestic Destinations(NUMBRD)	Number of Destinations
Number of European Destinations (NUMEUR)	Number of Destinations
Number of Intercontinental Destinations (NUMINT)	Number of Destinations

Figure 4.03: Attributes for Airport Categorization

Figure 4.04 displays the airports of the German Air Traveller Survey (Berster et al. 2005), which was used as database for model estimation, and the appropriate category for each of those airports.

Category	Airport (IATA-Code)
AP 1	Frankfurt a. M. (FRA)
AP 1	München (MUC)
AP 2	Düsseldorf (DUS)
AP 2	Hamburg (HAM)
AP 2	Köln/Bonn (CGN)
AP 2	Stuttgart (STR)
AP 3	Bremen (BRE)
AP 3	Dortmund (DTM)
AP 3	Dresden (DRS)
AP 3	Erfurt (ERF)
AP 3	Frankfurt Hahn (HHN)
AP 3	Friedrichshafen (FDH)
AP 3	Hannover (HAJ)
AP 3	Karlsruhe/Baden (FKB)
AP 3	Leipzig/Halle (LEJ)
AP 3	Lübeck (LBC)
AP 3	Münster/Osnabrück (FMO)
AP 3	Niederrhein (NRN)
AP 3	Nürnberg (NUE)
AP 3	Paderborn/Lippstadt (PAD)
AP 3	Saarbrücken (SCN)

Figure 4.04: Assignment of Airports to Categories

Figure 4.05 and 4.06 illustrates the properties of the identified three airport categories both in percentages and in absolute values (Gelhausen et al. 2006, p. 17).

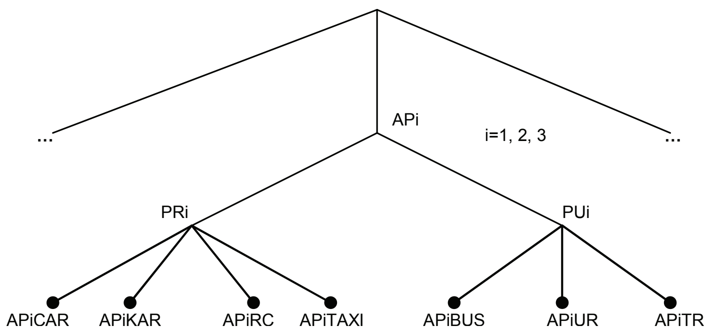
	LCBRD	CCBRD	LBRD	LCEUR	CCEUR	LEUR	LCINT	CCINT	LINT	NUMBRD	NUMEUR	NUMINT
AP 1	3.18	0.43	20.39	0.87	5.83	55.81	0.00	1.24	12.25	8.31	60.27	31.42
AP 2	8.97	0.58	28.27	11.65	11.76	37.24	0.02	0.71	0.79	16.23	74.62	9.16
AP 3	1.29	0.86	39.22	32.57	15.57	10.05	0.02	0.42	0.00	19.94	78.90	1.16

Figure 4.05: Properties of Airport Categories (in %)

	LCBRD	CCBRD	LBRD	LCEUR	CCEUR	LEUR	LCINT	CCINT	LINT	NUMBRD	NUMEUR	NUMINT
AP 1	106	16	756	32	225	2138	0	49	517	19	144	83
AP 2	104	7	348	129	153	487	0	11	11	17	80	12
AP 3	3	1	80	47	25	39	0	0	0	6	22	1

Figure 4.06: Properties of Airport Categories (absolute)

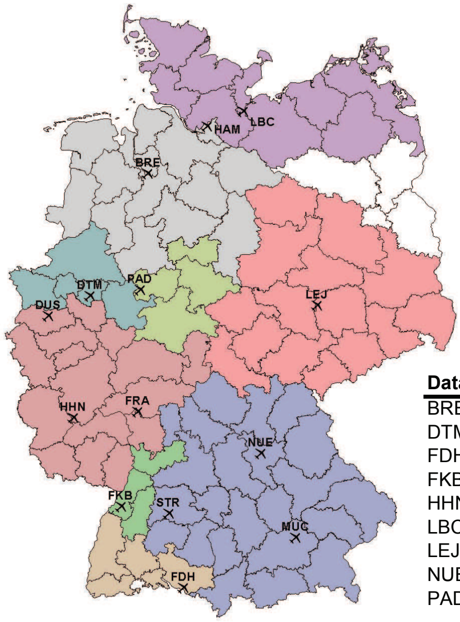
Figure 4.07 illustrates the nesting structure of each cluster group consisting of one airport category and all access modes (Gelhausen et al. 2006, p. 20).



Alternative	Abbreviation
AP 1/Car	AP1CAR
AP 1/Kiss and Ride	AP1KAR
AP 1/Rental Car	AP1RC
AP 1/Taxi	AP1TAXI
AP 1/Bus	AP1BUS
AP 1/Urban Railway	AP1UR
AP 1/Train	AP1TR
AP 2/Car	AP2CAR
AP 2/Kiss and Ride	AP2KAR
AP 2/Rental Car	AP2RC
AP 2/Taxi	AP2TAXI
AP 2/Bus	AP2BUS
AP 2/Urban Railway	AP2UR
AP 2/Train	AP2TR
AP 3/Car	AP3CAR
AP 3/Kiss and Ride	AP3KAR
AP 3/Rental Car	AP3RC
AP 3/Taxi	AP3TAXI
AP 3/Bus	AP3BUS
AP 3/Urban Railway	AP3UR
AP 3/Train	AP3TR

Figure 4.07: Nesting Structure

For model estimation the estimation data set is partitioned into several disjoint data subsets. Each data subset contains only a subset of the full set of airport-access mode alternatives, namely one airport of each category and its access modes. Each data subset includes observations of individuals, who have chosen one of the alternatives of the reduced alternative set. By a suitable definition of data subsets, it is possible to estimate a model with the full set of seven access modes for all three airport categories. For this purpose, the inclusion of the airports Frankfurt/Main, Düsseldorf and Leipzig/Halle is necessary, as these are the only airports of their category with an access via train in 2003. The individual data subsets are merged into a single new estimation data set. The number of alternatives is reduced from 122 to 21. By weighting each observation the estimation data set is statistically representative. Figure 4.08 shows the definition of the data subsets. The nearest airport of each category is assigned to each data set marked in different colours. Every subset is named according to its airport of the third category (Gelhausen et al. 2006, p. 18).



Data Subset	Airport (IATA-Code)
BRE	FRA, HAM, BRE
DTM	FRA, DUS, DTM
FDH	MUC, STR, FDH
FKB	FRA, STR, FKB
HHN	FRA, DUS, HHN
LBC	FRA, HAM, LBC
LEJ	FRA, HAM, LEJ
NUE	MUC, STR, NUE
PAD	FRA, DUS, PAD

Fig. 4.08: Data Subsets and Assignment of Airports

After selecting the airports and access modes for a specific application case, they are assigned to categories with the appropriate model parameters. The model can be applied to any number of airports. An application of the estimated model to other airports and airport/access mode combinations than those of the estimation data set is possible as a result of the grouping of clusters. Figure 4.09 summarizes the general process of the model estimation and its application (Gelhausen et al. 2006, pp. 18f.).

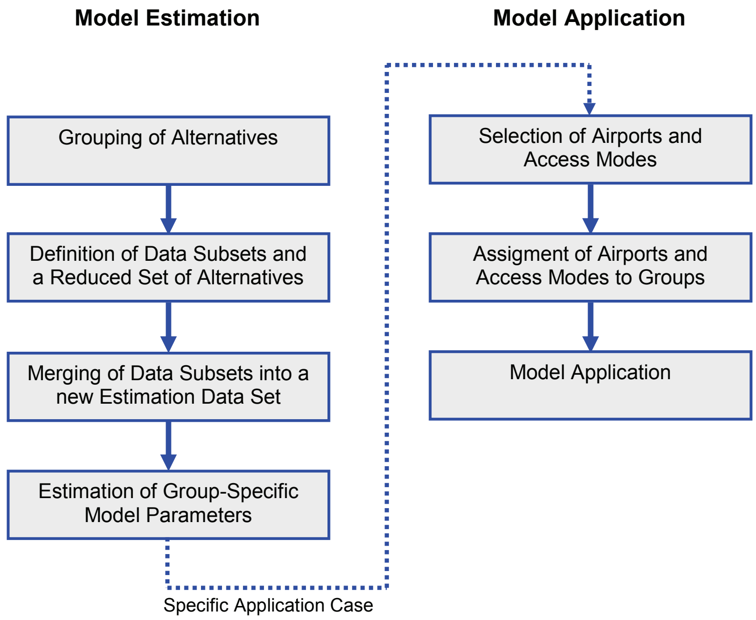


Fig. 4.09: Estimation and Application of Airport and Access Mode Choice Model

The deterministic part of the utility function of the nested logit-model is linear (Gelhausen et al. 2006, p. 19):

$$V_i = alt_i + \sum_k b_k * x_{k,i} \tag{4.01}$$

where

alt_i: Alternative-specific constant of alternative i
b_k: Coefficient of attribute k
$x_{k,i}$: Value of attribute k for alternative i

Figure 4.10 displays the estimated coefficients of the alternative attributes, scale parameters, goodness-of-fit measures and the likelihood-ratio test statistics for all seven market segments (Gelhausen et al. 2006, p. 28). Scale parameters are normalized on the lowest level of the nesting structure to a value of one. For the alternative-specific constants, p- and t-values and the standard deviation of the estimated coefficients see Gelhausen and Wilken (Gelhausen et al. 2006, pp. 21ff.).

Variable	BRD P	BRD B	EUR S	EUR H	EUR B	INT P	INT B
COST	-0.0263035	-0.0204609	-0.0199987	-0.0173617	-0.0216885	-0.0138527	-0.00936472
TIME	-0.0081889	-0.0152572	-0.0061063	-0.00857067	-0.00795957	-0.00541014	-0.00535887
WAIT	-28.8061	-18.935	-8.33078	-4.40982	-9.94709	-18.7546	-35.7591
INVPD	-187.86	-21.8829	-215.876	-235.641	x	-25.6109	-32.2589
COMP	-0.158635	x	-1.22176	-1.13258	-0.182127	x	x
AAS	0.920627	1.12781	0.20336	0.46823	0.504623	0.840462	0.382595
DIRECT	2.29637	3.64119	3.63327	3.31697	1.43564	1.85847	0.439344
DFREQ	0.00682913	0.00601159	0.0104684	0.0153856	0.0177437	x	x
LC	x	x	0.0863075	0.563633	0.275153	x	x
LCFREQ	x	x	0.0631856	x	0.0761092	x	x
PR1	1.07092	1.02375	0.764486	0.61189	0.808397	1.13266	1.03073
PU1	0.745385	0.978059	0.593257	0.3847	0.386155	0.983045	0.32899
PR2	0.492518	1.00829	0.767123	0.570138	0.783306	1.06067	1.3532
PU2	0.390636	0.992109	0.543582	0.437515	0.708662	0.927296	0.832438
PR3	0.817955	1.00988	0.821821	0.610065	0.937914	0.813943	0.91783
PU3	0.428619	0.999286	0.395656	0.551239	0.805435	0.137029	0.718249
AP1	1.81029	1.01119	1.80601	1.65075	1.61072	1.10489	2.10553
AP2	2.10174	1.00887	1.76862	1.92646	1.67197	1.19742	1.16102
AP3	2.35248	1.01164	1.74828	1.99236	1.77295	1.23031	1.73837
pseudo-R^2(null) in %	57.41	54.10	52.40	52.29	48.58	48.89	47.46
pseudo-R^2(const) in %	43.82	40.47	41.94	38.22	35.96	32.86	28.30
LR (MNL)	82414	8740	43774	349740	311756	599974	131576
α=0.5%	25.19	23.59	23.59	23.59	23.59	23.59	23.59

Figure 4.10: Overview Estimation Results per Market Segment

4.2 Generalized Neural Logit-Model

The Generalized Neural Logit-Model is estimated on the same data set as the aforementioned nested logit approach and the grouping structure is identical. The subdivision of access modes in private and public modes of travel is omitted as a two-stage hierarchy is sufficient. Case 4 is chosen in terms of correlation among alternatives as this approach is more flexible than a nested logit model yet it enables the development of a model, which is applicable to alternatives outside the estimation data set. The selection of explanatory variables is based on the nested logit model because of the possibility of statistical significance tests and simple plausibility checks. To consider endogenity of exogenous factors no weight sharing is applied. Figure 4.11 exemplifies the structure of the Generalized Neural Logit-Model for the case of private journeys to domestic destinations (BRD P).

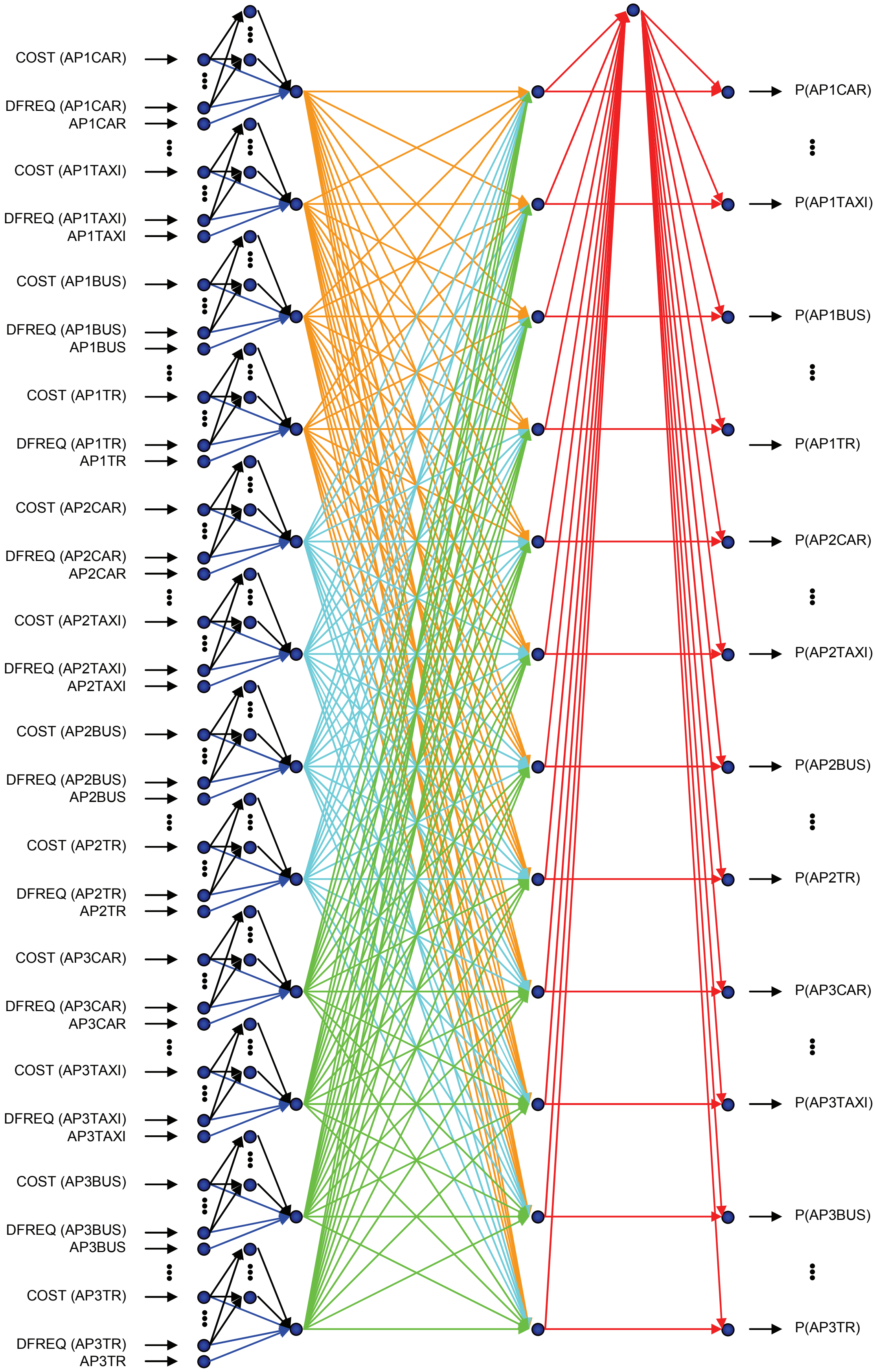


Figure 4.11: Structure of the Generalized Neural Logit-Model for the Market Segment BRD P

The method of network structure specification follows the idea of Miller, Todd and Hedge (Miller et al. 1989). The estimation data set is split up into a training set and a cross-validation set with a share of 85% and 15% respectively. Different network topologies are generated by a genetic search being trained on the training set and evaluated on the cross-validation set in terms of their ability to generalise. After the best network topology in terms of a minimal cross-validation error has been found the network is trained on the entire estimation data set without an early stopping of training, so that a maximum of information is available for the final estimation of the connection weights. This ensures a maximum of statistical efficiency (Anders 1997, p. 116). The ability of an artificial neural network does not decline in the case of an appropriate network structure (Anders 1997, pp. 117f.). The method of least squares is employed for the estimation of connection weights with conjugate gradients being the numerical optimisation method. Input variables are scaled on the interval [-1; 1]. Figure 4.12 summarises the training parameters.

Estimation Method	Least Squares
Optimisation Method	Conjugate Gradients
Scaling	yes, [-1; 1]
Genetic Search	
Share of Cross-Validation	15%
Population Size	10
Selection Rule	Roulette
Cross-over	Multi-Point
P(Cross-over)	0,9
P(Mutation)	0,01
Coding	Direct Encoding

Figure 4.12: Training Parameters

Because of computational costs the population size is chosen small. An optimal network topology is found within ten generations.

Model quality in terms of model fit is assessed by means of the pseudo-R^2. Benchmark is a model without any variables (R2null) and a market share model (R2const). Figure 4.13 illustrates the pseudo-R^2 by market segment for the Generalized Neural Logit-Model (GNL) and the nested logit approach (NL).

	R2(null)	**in %**		**R2(const)**	**in %**	
Market Segment	**NL**	**GNL**	**Diff. to NL**	**NL**	**GNL**	**Diff. to NL**
BRD P	57,41	61,35	**3,94**	43,82	49,74	**5,92**
BRD B	54,10	58,13	**4,03**	40,47	47,16	**6,69**
EUR S	52,40	58,09	**5,69**	41,94	49,99	**8,05**
EUR H	52,29	56,51	**4,22**	38,22	45,10	**6,88**
EUR B	48,58	51,96	**3,38**	35,96	41,79	**5,83**
INT P	48,89	55,10	**6,21**	32,86	42,01	**9,15**
INT B	47,46	56,01	**8,55**	28,30	41,26	**12,96**

Figure 4.13: Comparison of Model Fit

Especially for the market segments of intercontinental journeys for both private and business purpose the Generalized Neural Logit-Model shows a clear increase in model fit compared to the nested logit approach. The increase of R2(const) is about 45% for the market segment INT B compared to the nested logit model. The pseudo-R^2 is more evenly distributed over the market segments and lies between 41% and 49% in the case of R2(const). This corresponds to an R^2 of linear regression of 82% and 92% (Domencich et al. 1975, p. 124).

Figure 4.14 contrasts relative alternative frequencies and computed choice probabilities for the market segment EUR S.

Alternative	Relative Frequency	Logit-Model	abs. Diff.	NL-Model	abs. Diff.	GNL-Model	abs. Diff.
AP1CAR	4,7068	5,7485	1,04	5,4012	0,69	5,2468	0,54
AP1KR	9,1049	9,9923	0,89	10,0309	0,93	9,4534	0,35
AP1RC	0,1929	0,1543	0,04	0,1543	0,04	0,2043	0,01
AP1TAXI	2,7006	2,8549	0,15	2,7392	0,04	2,7109	0,01
AP1BUS	0,5401	0,4630	0,08	0,5015	0,04	0,4558	0,08
AP1UR	5,5170	6,2500	0,73	6,4429	0,93	6,2622	0,75
AP1TR	1,3503	1,7361	0,39	1,7747	0,42	2,3327	0,98
AP2CAR	9,9537	9,4907	0,46	9,3364	0,62	9,9918	0,04
AP2KR	16,4738	14,1204	2,35	14,1975	2,28	15,1639	1,31
AP2RC	0,3472	0,2315	0,12	0,1929	0,15	0,2491	0,10
AP2TAXI	6,5201	6,5972	0,08	6,5972	0,08	6,6302	0,11
AP2BUS	2,1219	2,7778	0,66	2,6620	0,54	2,3235	0,20
AP2UR	4,2438	5,0154	0,77	5,3627	1,12	4,2061	0,04
AP2TR	4,4753	3,5880	0,89	3,7809	0,69	3,8147	0,66
AP3CAR	16,3580	16,7052	0,35	16,2423	0,12	17,5289	1,17
AP3KR	11,0725	10,0309	1,04	9,7994	1,27	9,6736	1,40
AP3RC	0,1157	0,0772	0,04	0,0772	0,04	0,1048	0,01
AP3TAXI	2,4306	2,7778	0,35	2,7392	0,31	2,1798	0,25
AP3BUS	1,2731	0,9259	0,35	1,3889	0,12	0,9809	0,29
AP3UR	0,4244	0,4244	0,00	0,5401	0,12	0,4718	0,05
AP3TR	0,0772	0,0386	0,04	0,0386	0,04	0,0149	0,06
E(abs. Diff.))			0,51		0,50		0,40
σ(abs. Diff.)			0,54		0,55		0,45

Figure 4.14: Relative Alternative Frequencies and Computed Choice Probabilities for EUR S

Airport and access mode choice behaviour is governed by a complex nonlinear utility function and correlations among alternatives beyond the capabilities of a nested logit approach with a linear utility function as the clear increase in model fit demonstrates.

Figure 4.15 and 4.16 illustrate the dependency between two selected alternative attributes and the choice probability of one alternative. There is a clear nonlinear relationship between access time, access cost and the choice probability of the appropriate alternative in domestic air travel for business purpose. These travellers are very access time-sensitive. This relationship is of more linear form with a greater importance of access cost in the market segment of intercontinental air travel for private reasons.

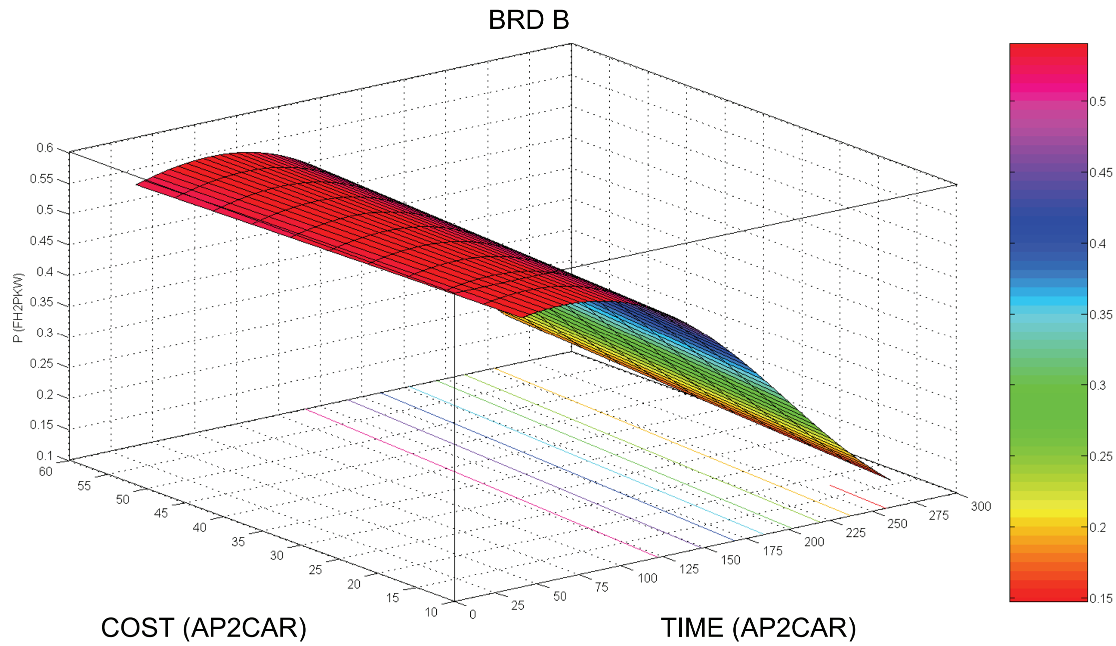


Figure 4.15: Analysis of the Utility Function for BRD B

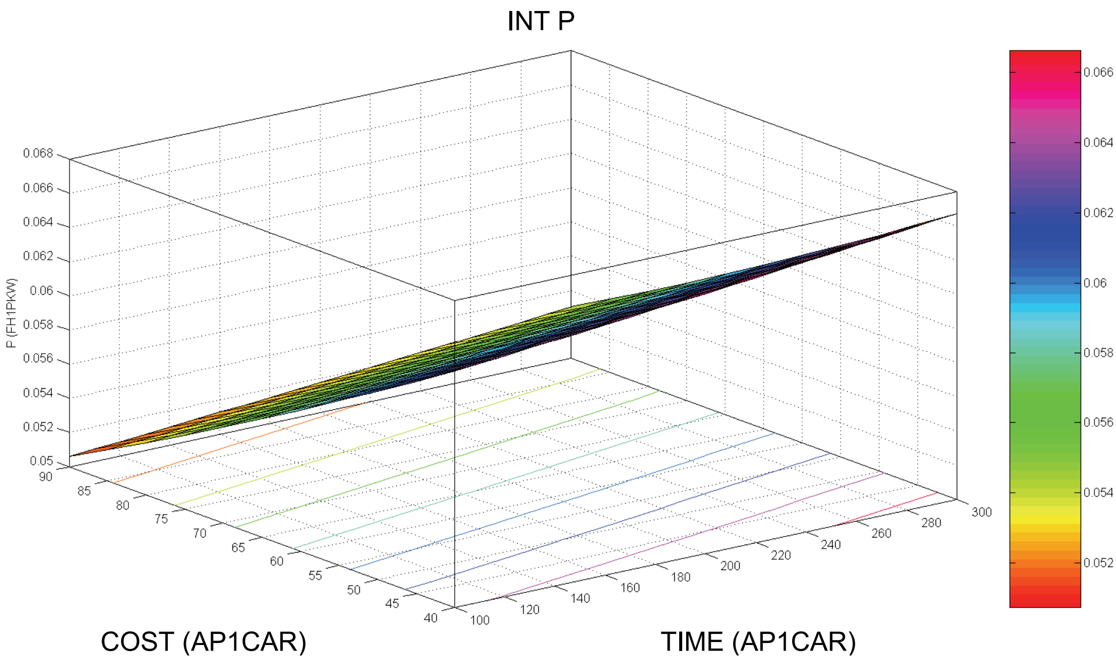


Figure 4.16: Analysis of the Utility Function for INT P

5. SUMMARY AND CONCLUSIONS

This paper presents a novel approach in discrete choice modelling called "Generalized Neural Logit-Model". This approach is based upon the General Extreme Value-framework and is implemented as artificial neural network. Its main advantages lie in a nonparametric nonlinear utility function and the capability to model arbitrary correlations among alternatives in the choice set.

The first part of this paper deals with the concept of alternatives and cluster groups. It enables the development of discrete choice models applicable to alternatives outside the estimation data set.

The next chapter introduces the Generalized Neural Logit-Model. The first part is about the theoretical framework followed by an implementation as artificial neural network.

The Generalized Neural Logit-Model is evaluated empirically by means of an application case. The chosen application case is airport and access mode choice in Germany and a nested logit approach based on the concept of cluster groups serves as a benchmark. To form cluster groups airports are categorized from a demand-oriented point of view by means of Kohonen's Self-Organizing Maps.

The Generalized Neural Logit-Model is superior to the nested logit approach in terms of model fit as the considered problem is governed by a complex nonlinear utility function and correlations among alternatives beyond the nested logit approach with a linear utility function. The pseudo-R^2 based on a market share model as a benchmark is in the range of 41% to 49% and lies up to 45% above the nested logit approach depending on the market segment. This corresponds to an R^2 of linear regression of 82% to 92%, so that a model of very good quality can be obtained for all market segments.

REFERENCES

Anders, U. (1997) *Statistische neuronale Netze*, München.

Bhat, Ch. R. (2003) Econometric Choice Formulations: Alternative Model Structures, Estimation Techniques and Emerging Directions, *paper presented at the 10th International Conference on Travel Behaviour Research*.

Ben-Akiva, M., Lerman, S. (1985) *Discrete Choice Analysis: Theory and Applications to Travel Demand*, Cambridge.

Bentz, Y., Merunka, D. (2000) Neural Networks and the Multinomial Logit for Brand Choice Modelling: A Hybrid Approach, **Journal of Forecasting, vol. 19** 177-200

Berster, P., Focke, H., Pabst, H., Wilken, D. (2005) *Gemeinsame Auswertung der ADV-Fluggastbefragung 2003*, Köln (unveröffentlicht).

Bishop, C. M. (2003) *Neural Networks for Pattern Recognition*, Oxford.

Domencich, T. A., McFadden, D. (1975). *Urban Travel Demand - A Behavioral Analysis*, New York.

Fausett, L. (1994) *Fundamentals of Neural Networks - Architectures, Algorithms and Applications*, Englewood Cliffs

Gaudry, M. J. I., Dagenais, M. G. (1979) The Dogit Model, **Transportation Research, vol. 13 B** 105-111.

Gelhausen, M. (2003) *Evaluation zufallsnutzenbasierter Verfahren und künstlicher neuronaler Netze zur Prognose von Flughafen- und Feedermoduswahl*, Aachen

Gelhausen, M., Wilken, D. (2006) Airport and Access Mode Choice - A Generalized Nested Logit Model Approach, *paper presented at the 10th Air Transport Research Society Conference 2006 in Nagoya,* Japan.

Hornik, K., Stinchcombe, M., White, H. (1989) Multilayer Feedforward Networks are Universal Approximators, **Neural Networks, vol. 3** 359-366.

Hruschka, H., Fettes, W., Probst, M., Mies, Ch. (2002) A Flexible Brand Choice Model Based on Neural Net Methodology - A Comparison to the Linear Utility Multinomial Logit Model and its Latent Class Extension, **OR Spectrum, vol. 24** 127 – 143.

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., Jackel, L. D. (1989) Backpropagation Applied to Handwritten Zip Code Recognition, **Neural Computation, vol. 1** 541-551.

Maier, G., Weiss, P. (1990) *Modelle diskreter Entscheidungen - Theorie und Anwendung in den Sozial- und Wirtschaftswissenschaften*, Wien.

McCarthy, P. S. (1997) The Role of Captivity in Aggregate Share Models of Intercity Passenger Travel, **Journal of Transport Economics and Policy, 31** (3) 293-308.

McFadden, D. (1974) Conditional Logit Analysis of Qualitative Choice Behavior, in: Zarembka, P.: *Frontiers in Econometrics* 105-142, New York.

McFadden, D. (1978) Modelling the Choice of Residental Location, in: Karlqvist, A., Lundqvist, L., Snickars, F., Weibull, J.: *Spatial Interaction Theory and Planning Models* 75-96, Amsterdam.

Miller, G. F., Todd, P. M., Hedge, S. U. (1989) Designing Neural Networks Using Genetic Algorithms, in: Schaffer, J. D.: **Proceedings** *of the 3rd International Conference on Genetic Algorithms* 379 – 384.

Ortuzar, J. de D., Willumsen, L. G. (2001) *Modelling Transport*, London.

Probst, M. (2002) *Neuronale Netze zur Bestimmung nichtlinearer Nutzenfunktionen in Markenwahlmodellen,* Regensburg.

Rumelhardt, D. E., Hinton, G. E., Williams, R. J. (1986) Learning Internal Representations by Error Propagation, in: Rumelhart, D. E., McClelland, J. L.: *Parallel Distributed Processing: Exploration in the Microstructure of Cognition* vol. 1: Foundations 318 – 362, Cambridge.

Train, K. E. (2003) *Discrete Choice Methods with Simulation*, Cambridge.

Wilken, D., Berster, P., Gelhausen, M. (2005). Airport Choice in Germany – New Empirical Evidence of the German Air Traveller Survey 2003, *paper presented at the 9th Air Transport Research Society Conference 2005 in Rio de Janeiro,* Brazil.