Diplomarbeit

Titel der Diplomarbeit

„Demand Estimation with a Model of Discrete Choice“

Verfasser

Fabian Schroeder

Angestrebter akademischer Grad

Magister der Sozial- und Wirtschaftswissenschaften
(Mag. rer. soc. oec.)

Wien, im Februar 2009

Studienkennzahl lt. Studienblatt:	A-140
Studienrichtung lt. Studienblatt:	Diplomstudium Volkswirtschaft
Betreuer:	Ao. Univ. Prof. Mag. Dr. Burcin Yurtoglu

Misura ciò che è misurabile, e rendi misurabile ciò che non lo è.

Galileo Galilei

Demand Estimation with a Model of Discrete Choice

Fabian Schroeder

February 19, 2009

Contents

Abstract

For the purpose of estimating demand functions, economists have developed a broad range of models, which are as idiosyncratic as are the markets they describe. The objective of my thesis is to explore possibilities to estimate residential demand for internet access in Austria. Two models will be presented, each situated at the very opposite ends of the spectrum. They differ with respect to some fundamental assumptions concerning the nature of the market - such as the homogeneity or heterogeneity of the goods -, the level of aggregation requiring different sources of data, and a number of other specifications. The first model is based on a simple macro-level setting and will assume homogenous goods. This framework, however, is of limited use for most research questions, and thus motivates the application of the much richer and more powerful second model. It considers a differentiated product market and is based on a micro-level model of discrete choice. Even within the family of discrete choice models, there remains a wide range of possible frameworks. The identification of the most suitable given the unique characteristics of the market and data availability is the purpose of the final chapters.

1 Introduction

It is the art of the econometrician to create an adequate model of a market. The specific characteristics have to be taken into account. Consider the market for internet access, for example. Due to the extraordinary rate of diffusion, which is likely without a parallel in the history of technological change and the dramatic development of applications for internet access, a model that considers observations over time can hardly be applied. Internet access in the year 2000 was an entirely different product than in the year 2008. This fact prohibits the use of panel data. Other important idiosyncrasies on which model selection was based includes, among others: the binary nature of the decision to acquire internet access, the limited number of supplied goods, and the fact that most internet connections are not transportable.

Without claiming to have found the most suitable model, nor to possess an exhaustive number of possible models I will pick two very different settings and undertake a comparative analysis of each. The first model represents the starting point of the work. Among others, it will be based on the very limiting maintained assumption of homogenous goods. The relaxation of this assumption comes at a high cost. We will see how the assumption of differentiated products leads to a much more demanding framework. The initial chapters will highlight the maintained assumptions that lead to the models and compare them with regard to explanatory power and suitability.

2 Model Selection

The most critical question is whether the goods are considered homogenous or heterogeneous. In the case of internet access this question is not as straightforward as it may look. Both positions are to some extent justified. From a technical perspective, internet access consists of information packages sent from the user to internet hubs. Even though they can be delivered by means of different technologies the very nature of the good is the same. However, if internet goods are effectively perfect substitutes, one will fail to find an explanation for observed differences in price. They are a good indicator for whether two goods belong to the same market or not. From the consumer's perspective internet access is as heterogeneous as are its applications. These include: the obtaining of information, communication, audio and video entertainment and online gaming. On the lower end of the spectrum of internet users stands the person whose main use for the internet is reading the newspaper and obtaining information. He will very likely consider all types of internet access equal, thus price will play a role. He even has close substitutes such as real newspapers and other means of communication. On the other extreme of the spectrum there is the online gamer, who is hungry for bandwidth and download volume. There exists no close substitute for his use of the internet. The mentioned applications require completely different speeds and volumes and therefore a slight difference in the speed of the internet connection can completely alter the possibilities of its use. Speed and volume are however not the only means of active product differentiation. Mobility as in the case of internet via HSDPA technology, the provision of email addresses, web space, or the setting the overbooking factor are other important means.

In any case, the decision whether to assume homogenous or heterogeneous goods for modelling purposes will most likely be based on the intention of estimating demand. This underlying purpose will shape the basic outlines of the model in use. A wise researcher will chose a model that is sufficiently complex to capture all the required entities involved in answering his question but also keep it as simple as possible. Consider the following example: For the purpose of defining a market - a necessary step in any market survey - the researcher will be interested in the extent to which consumers react to a price change of one good by switching to a similar good. In this case it makes no sense to assume homogenous goods. For purposes, for which the elasticities between the goods in the market are irrelevant, the assumption of homogenous goods will simplify the model dramatically.

Model number one stands in the neo-classical tradition. It will attempt to uncover price and income demand elasticities for broadband internet access from market data under the assumption of a perfectly competitive equilibrium. The most critical maintained assumption is that broadband internet goods are perfect substitutes and represent a market of its own. As a consequence there exists only one market price which is the one that we observe. Market demand will have the general form:

$$\begin{aligned} Q &= f(P, ...) \qquad (1) \\ P &= g(Q, ..) \end{aligned}$$

Model number two will assume heterogeneous goods and will include all internet connections, not only broadband. A differentiated n product markets consequently requires a system of n demand functions

$$\begin{aligned} P_1 &= f(Q_1, ..., Q_n, ...) \qquad (2) \\ &\vdots \\ P_n &= f(Q_1, ..., Q_n, ...) \\ Q_1 &= g(P_1, ..., P_n, ...) \\ &\vdots \\ Q_n &= g(P_1, ..., P_n, ...) \end{aligned}$$

and a system of n supply functions - usually derived from first order profit maximising conditions. I will now briefly present one possible version of framework (1) in order to illustrate its limitations. This will serve as a motivation for the introduction of model number two.

2.1 Model One

Model One will attempt to uncover price and income elasticities from cross-country data. To reduce unobserved cross-country heterogeneity I have only included data of the 30 OECD member countries, which constitute a fairly homogenous group. The only proxy I used to capture country differences was GDP in PPP dollars. This approach may seem a little bit strange, after all I am interested in elasticities for Austria. The most promising setting for my purpose would have probably been one using regional Austrian data - comparing the prices and quantities of internet access by province or even township. Unfortunately, this data did not show sufficient variation in price to have any explanatory power. This is probably due to the size of Austria and the fact that most suppliers offer one price for the whole country. Also it is not clear whether the assumption of homogeneity of people within one province of Austria is less problematic. This issue is clearly a result of the macro-level approach to demand estimation. It will again be raised in the next chapter where we will see how the second model deals with it.

I will use a simple log-linear model of demand and supply with the following structure:

$$\ln Q_i = a + b(\ln P_i) + c(\ln Y_i) + e_i \tag{3}$$

$$\ln P_i = \alpha + \beta \ln(D_i) + \epsilon_i \tag{4}$$

Q_i denotes a measure of the quantity of internet access and P_i a measure of the price. Y_i is a proxy for real income and e_i and ϵ_i are well behaved stochastic error terms. In order to fix the endogeneity problem I have added D_i - population density of country i - as an instrument for supply. The intuition for this is the following: A high population density will reduce the per capita infrastructure costs for bandwidth connections and therefore stand as a proxy for the cost of supply. A more detailed explanation of the variables can be found in table (2.1). The results of the 2-stage least squares

Variable	**Description**	**Mean**	**St.Dev.**	**Data Source**
Q: bb_pen	Broadband penetration (subscribers per 100 inhabitants)	21.58	8.86	OECD, 2006
P1: pri_sub	Average broadband monthly subscription price	52.32	13.74	OECD, 2006
P2: pri_mb	Average broadband monthly price per advertised Mbit/s			OECD, 2006
Y: gdp_ppp	GDP per capita price power parity	32222	11425	World Bank, 2006
D: pop_den	Population density (inhabitants per km^2	133.91	124.62	OECD, Dec 2007

Table 1: Data Sample

estimation can be found in table (2.1). The model delivered fairly robust results. Slight modifications in the model specification did not alter the outcome completely. I also tried two different proxies for the price of broadband internet, delivering the same results. The explanatory power of this simple model was excellent. The adjusted R^2 was in the order 0.6 in most models. The coefficients had the expected sign prompted by economic theory and were highly significant. Any undertaken expansions of the model with some

ln_bb_pen	**ln_bb_price**	**ln_gdp_ppp**	**Constant**
Coefficiant	-0.992347	1.360549	-7.159
t-Value	-2.31	7.86	-4.44
P> $\|t\|$	0.029	0.000	0.000
Adjusted $R^2 = 0.6977$	F-statistic = 0.0000	Observations: 30	

Table 2: Results of the 2-Stage Least Squares Estimation

additional variables such as country dummies or broadband speed did not prove successful. They did either not add any explanatory power or the sign lacked plausibility.

The result can be interpreted as follows: The average OECD broadband price elasticity is approximately -1. This implies that a 10 % increase in price would decrease demand by 10%. The average OECD income elasticity is approximately 1.4. This indicates that a 10% increase in the average income would increase demand by 14%. Broadband access at home still seems not to be affordable for everybody.

Depending on the purpose of research this model can be too limited with respect to the assumption that goods as well as consumers are homogenous. As a result of the homogeneity of goods assumption, which is a nontestable maintained assumption, one has to calculate the averages of broadband prices within a country. Although broadband internet access is offered in a variety of different packages, including different speed, and volume, they were considered all the same and averages were calculated. By deleting variation within a country a lot of useful information was lost. Additionally our model does also not allow us to shed light on substitution patterns between broadband and non-broadband internet, or to differentiate between mobile and non mobile broadband. Another problematic assumption of this model was that price was the only differentiating feature between OECD countries. This means that consumer's preferences were assumed to be generally equal. Therefore, I will introduce a much richer model of the market for Austrian internet goods in the following chapters that can account for both differentiated goods as well as differences in consumers tastes.

The only real difficulty in Model One is how to solve the endogeneity problem. Demand parameters can be consistently estimated in the presence of unobserved demand factors via the use of traditional instrumental variables methods. In framework (2) another problem arises. The sheer number of parameters to be estimated can become burdensome. Consider a model of demand that only includes price elasticities.

Definition 1 (Cross/Own-Price Elasticity)

$$\mathcal{E}_{i,j} = \frac{\partial Q_i}{\partial P_j} \frac{P_j}{Q_i} \tag{5}$$

A set of n goods renders n^2 price elasticities to be estimated. Therefore it will be necessary to introduce some kind of a structure that includes the use of a richer data set than just price and quantity information. This can be achieved in many ways. The chosen approach will impose such structure basing demand on a micro level model of consumer choice. The underlying behavioral model - the Random Utility Model which will be introduced in the next chapter - will explain the utility a consumer derives from the acquisition of a product with a set of product characteristics and a set of socio-economic data of the consumer standing as proxies for his individual taste. These utilities will render the probability of a consumer choosing a certain internet connection given his personal attributes. Disaggregate individual demand can then be aggregated to deliver market level demand. This was first done by Goldberg (1995) (3). In this pioneering work she developed the following structure: q_j^e, the demand for good j is equal to the sum over all households or individual of the probabilities that alternative j was chosen.

$$Q_c = \sum_{i \in \mathcal{I}} \mathbf{P}(\text{agent } i \text{ buys product } c) = \sum_{i \in \mathcal{I}} \Psi(p_1, ..., p_n, \boldsymbol{x}_1, ..., \boldsymbol{x}_n, \boldsymbol{\omega}_i, \theta) \tag{6}$$

$\Psi(\boldsymbol{p}, \boldsymbol{x}_1, ..., \boldsymbol{x}_n, \boldsymbol{\omega}_i, \theta)$ are the (Nested) Logit choice probabilities that depend on the price p and non price attributes x of all goods, as well as a set of socio-economic variables standing as proxies for household i's taste, and a distribution θ.

In the following third chapter I will introduce random utility models and derive the very popular Logit and Nested Logit model. This introduction will go as far as is necessary for the basic understanding of the model I will use for estimation. I would like to recommend the book *Discrete Choice Methods* by Kenneth Train (4) to the reader interested in a more thorough discussion of discrete choice models.

3 Discrete Choice Models

3.1 Introduction

Discrete Choice Models have been developed to model the decision of an agent consisting of a choice made among a finite set of alternatives. Many decisions is real life are of this kind. Classic examples include the decision whether participate in the labour market or not, consumer choices in markets such as the transportation market, the market for insurances, and - not surprisingly - the market for internet goods. One cannot purchase half a broadband connection and it rarely makes sense to purchase more than one. The discreteness can be considered a bias towards countable numbers and makes estimation more difficult.

The regular linear regression models tends to fail to capture the specific nature of discrete decisions. Linear models of the form $\mathbf{P}(Eventjoccurs) = F(Parameters)$ - Linear Probability Models - have been developed. They, however, have a number of shortcomings. See Greene(2003)6, for example. The setting $\mathbf{P}(Y = 1|\boldsymbol{x}) = F(\boldsymbol{x}, \boldsymbol{\beta})$ and $\mathbf{P}(Y = 0|\boldsymbol{x}) = F(\boldsymbol{x}, \beta)$ with $F(\boldsymbol{x}, \boldsymbol{\beta}) = \boldsymbol{x}'\boldsymbol{\beta}$ will lead to heteroscedasticity. The variance of ϵ is a function of $\boldsymbol{\beta}$.

In order to model the decision making process of individuals, we will have to make specific assumptions. We will distinguish here among assumptions about

- **The decision-maker**: ($\mathcal{I}$) These assumptions define who is the decision-maker, and what are his/her characteristics. In many situation it is not clear whether the decison-maker is a person, a household or the board of a company. This assumption is often more critical than it may seem. Our example of internet access is a perfect example. Whereas regular internet connections are usually shared by the members of a household, mobile internet is more often acquired by an individual. The following questions need to be answered: Who is the decision making person in the household? What socio-economic data can be considered to characterize them both?

- **The alternatives**: ($\mathcal{C}$) These assumptions determine the possible options of the decision-maker. *To fit within a discrete choice framework, the set of alternatives, called the choice set, needs to exhibit three characteristics. First, the alternatives must be mutually exclusive from the* *decision makers perspective. Choosing one alternative necessarily implies not choosing any of the other alternatives. The decision maker chooses only one alternative from the choice set. Second, the choice set must be exhaustive, in that all possible alternatives are included. The decision maker necessarily chooses one of the alternatives. Third, the number of alternatives must be finite. The researcher can count the alternatives and eventually be finished counting.*4

The essential assumption of a mutually exclusive choice set - basically meaning that every agent can only choose one alternative out of this set - is not very restrictive since the choice set can be expanded to include multiple good alternatives. Think of an agent that decides to use broadband internet at home and additionally has a mobile internet connection when travelling. All that needs to be done is to expand the choice set with this alternative. Then 'Broadband only', 'Mobile only' and 'both BB and Mobile' are all separate alternatives. A similar method can be applied in order to satisfy the second assumption. By simply adding the alternative 'none of the other alternatives',which basically means that the agent has no internet connection, one can assure that the choice set is exhaustive. If there was no outside option the agents of our model would be forced to choose an alternative. Firstly this is not what we observe in reality and secondly this would render implausible consumer behaviour e.g. the choice is dependent only on the difference in prices of the goods and not the absolute prices. The third assumption actually is restrictive. A situation where an agent has infinitely many alternatives to choose from cannot be captured by a discrete choice model. This condition is the defining characteristic of discrete choice models and distinguishes them from a regular regression.

In our case the set of alternatives will include: $\mathcal{C}$ = {Cable, ADSL, Mobile, DialUp and No Internet}.

- **The attributes**: These assumptions identify the attributes of each potential alternative that the decision-maker is taking into account to make his/her decision. The question what attributes matter to an individual is a crucial issue. It involves extensive inquiry by the analyst possibly involving questioning consumers. In the case of internet goods one can imagine the following characteristics to be of importance: Speed, Price, Download limits, less obvious but possibly also important: the corporate image of the company, customer service, ... In practice the availability of data and/or the possibility to quantify a certain characteristic will have a big impact on which attributes the analyst includes in his model.

However, not only the attributes of the alternative matter in the decision process. The underlying consideration for a decision differ strongly between agents. This variation in consumer taste can rarely be observed. Therefore the analyst uses socio-economic parameter serving as proxies in order to reflect the heterogeneity of the agents. One of the main challenges is to capture as much information as possible in order to minimize the unobserved heterogeneity and reduce it to pure noise.

- **The decision rules**: They describe the process used by the decision-maker to reach his/her choice.

  Discrete choice models are based on an underlying behavioural model which attempts to explain the variation in people's behaviour. The complexity of human behaviour,however, suggests that a choice model should explicitly capture some level of uncertainty. The neoclassical economic theory fails to do so. The exact source of uncertainty is an open question. Some models assume that the decision rules are intrinsically stochastic, and even a complete knowledge of the problem would not overcome the uncertainty. Others consider that the decision rules are deterministic, and motivate the uncertainty from the impossibility of the analyst to observe and capture all dimensions of the problem, due to its high complexity. This question will further be addressed in the chapter on nested Logit models.

  While the source of the variability of human behaviour is open to debate we will proceed with the assumption that agents are able to compare any two alternatives and choose the alternative from which they derive the most utility. This behaviour is referred as utility maximising behaviour and is an essential assumption of Random Utility (Maximisation) Models (RUM).

3.2 Random Utility Models

Consider a situation where an agent is supposed to choose from a set of alternatives. Since he is assumed to have perfect discriminatory capability he will choose the alternative from which he derives the most utility. Therefore the probability that agent i chooses alternative a from a set of choices $\mathcal{C}$ is given by

$$\mathbf{P}_{i,a} = \mathbf{P}(U_{i,a} = \max_{c \in \mathcal{C}} U_{i,c}) \tag{7}$$

The agent chooses option a if and only if $U_a \geq U_c, \forall c \in \mathcal{C}$. From the analyst's perspective the utility that the agent derives from this option is unobservable. What he can observe, however, is a set of attributes of the alternatives $\boldsymbol{y}$ and a set of characteristics of the agent $\boldsymbol{z}_i$ that have explanatory power with regard to the agent's utility. Let $\boldsymbol{x}_{i,a} := (\boldsymbol{y}, \boldsymbol{z}_i)$. Due to the existence of unobserved entities it is therefore necessary to distinguish between the real utility $U_{i,a}$ of agent i and that part of utility that can be explained by observable attributes $V_{i,a}$, which I will refer to as the deterministic part of utility. Therefore $U_{i,a} \neq V_{i,a}$ This difference stems from four different sources of uncertainty: unobserved alternative attributes, unobserved individual attributes, measurement errors and proxy, or instrumental, variables.5. In order to reflect this uncertainty utility is modelled as a random variable.

$$U_{i,a} = V_{i,a} + \epsilon_{i,a}$$

The utility of agent i chosing option a is given by a determined part $V_{i,a}$ and a stochastic part $\epsilon_{i,a}$. They are assumed to be independent and additive (maintained assumption). The deterministic part is a function of both characteristics of the agent as well as attributes of the alternative.

$$V_{i,a} = F(\boldsymbol{y}_{i,a}, \boldsymbol{z}_i)$$

Since it should be clear by now that we are dealing with probabilities on the level of an agent i will drop the subscript i for notational simplicity! The probability that decision maker i chooses option a is

$$\begin{aligned}\mathbf{P}_a &= \mathbf{P}(U_a > U_c, \forall c \neq a) = \\ &\quad \mathbf{P}(V_a + \epsilon_a > V_c + \epsilon_c, \forall c \neq a) = \\ &\quad \mathbf{P}(\epsilon_a - \epsilon_c > V_c - V_a, \forall c \neq a)\end{aligned}$$

Agent i will choose alternative a if the differences in the unobserved part of utility makes up for the difference in observed utility. A look at this equation reveals two interesting characteristics of RUMs. First, only differences in utility matter. If a constant is added to every utility, the outcome of the decision process should not be altered. Secondly, there is no natural scale for utility. If all utilities are multiplied by a constant, again probabilities should not change. Random utility models therefore require normalization, an issue that will be brought up in chapter (3.5).
Denoting the joint vector of random terms as $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, ..., \epsilon_n)$ (one for every element of the choice set) with a density function $f(\boldsymbol{\epsilon})$ and by using the indicator function $\mathbf{1}[.]$ that takes the value 1, whenever the expression in the brackets is true and 0 if not, the probability can be rewritten as:

$$\mathbf{P}_a = \int_{\boldsymbol{\epsilon}} \mathbf{1}[\epsilon_a - \epsilon_c > V_c - V_a, \forall c \neq a] f(\boldsymbol{\epsilon}) d\boldsymbol{\epsilon}$$

The choice probability is the n-1 dimensional integral over the distribution of the difference of the error terms. Up to this point the model was kept as general as possible. In order to make a Random Utility Model operational it will be necessary to make a maintained assumption concerning both the distribution of the random terms and the functional form of deterministic utility. The attribute 'maintained' stresses the fact that this assumption cannot be tested and will remain an assumption. It is not possible to clarify whether the distribution fits the data or not.
Whereas deterministic utility is almost always modelled as a linear function $V_{i,a} = \boldsymbol{x}'\boldsymbol{\beta}$, the decision about the distribution of the random term is not as straigthforward and will be the critical assumption by which we differentiate between different discrete choice models.
Any continuous probability distribution will do. The most widely used 6 are the logarithmic and the normal distribution delivering the Logit and the Probit model respectively. The assumption of jointly normal distributed error terms seems very natural. However it comes with the disadvantage that estimation requires the solution of multidimensional integrals since the normal distribution does not have a closed form. This causes computational difficulties which deters many researchers from the Probit model. The most common discrete choice model is,therefore, the

3.3 The Logit Model

The Logit model is obtained by assuming that all random terms are identically and independently distributed Extreme Value of type one.

$$\epsilon_c \sim EV1 \qquad \forall c \in \mathcal{C}$$

Note that sofar the ϵ_cs were indexed in recognition that the stochastic component can be different across alternatives. They may be correlated between pairs of alternatives and have different distributions. However, by introducing assumption (3.3) the error terms will be not cross-correlated and identically distributed. As a consequence we will be able to drop the subscript of the ϵs.
The Generalized Extreme Value Distribution of Type 1 (1) - also referred to Gumbel [1] distribution - has a density function given by

$$f(\epsilon) = e^{-\epsilon} e^{-e^{\epsilon}}$$

[1]Emil Julius Gumbel (1891 - 1966)

and cumulative density given by

$$F(\epsilon) = e^{-e^{\epsilon}}$$

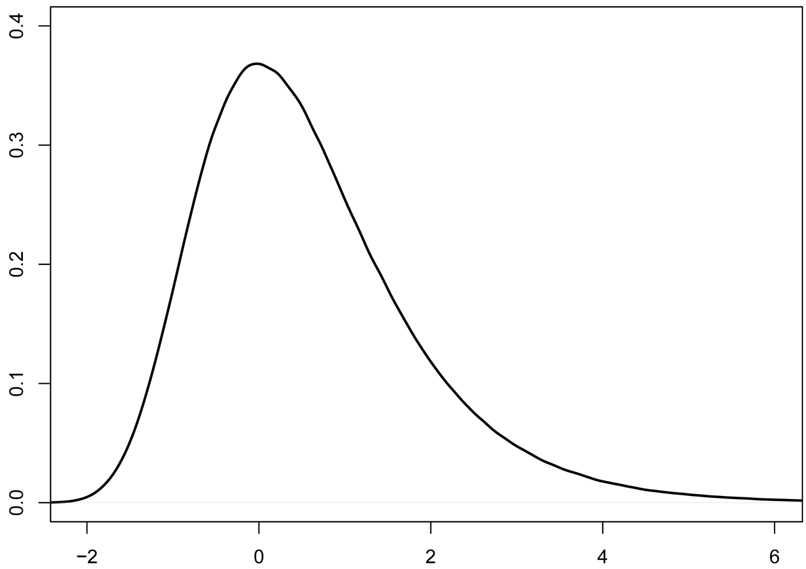


Figure 1: The Gumbel Distribution

The variance of the Extreme Value type 1 distribution is $\pi^2/6$. This has important consequences for the normalization of the model, an issue we will come back to later. The difference between two draws of GEV type 1 variables has a logistic distribution with the distribution function and cumulative distribution respectively

$$\epsilon_c - \epsilon_d =: \epsilon^*_{cd} \sim Log$$

$$F(\epsilon^*_{i,cd}) = \frac{e^{\epsilon^*_{i,cd}}}{1 + e^{\epsilon^*_{i,cd}}}$$

The Multinomial Logistic distribution is very close to the normal distribution with slightly fatter tails. The critical part of the assumption is that the unobserved factors are uncorrelated over alternatives, as well as having the same variance for all alternatives. While this assumption is very restrictive it delivers a very convenient form for the choice probability. Logit is by far the most widely used discrete choice model due to its mathematical convenience. However, the assumption of independence can be inappropriate in some situations.A correlation between error terms basically means that there is some similarity between two different alternatives which was not sufficiently captured by the chosen explanatory variables. In many cases this is caused by a lack of data. Put in other words: The objection for the analyst is to model utility sufficiently well, as to reduce the random terms to white noise. If there is correlation between the error terms, the chosen explanatory variables for utility do not have sufficient explanatory power. This problem can be fixed by either trying to find a more suitable model to explain utility or by relaxing the assumption of iid error terms, which will lead to the Generalized Extreme Value Model, introduced in chapter 3.7.

3.4 Derivation of the Logit Model Probabilities following McFadden (1974)

Starting point is the general behavioural model developed in (3.2) where agent i chooses alternative a only when its utility is greater than of other alternative.

$$\mathbf{P}(\epsilon_a + V_a > \epsilon_c + V_c, \forall c \neq a) =$$
$$\mathbf{P}(\epsilon_c < \epsilon_a + V_a - V_c, \forall c \neq a)$$

The assumption of iid random terms will lead to a cumulative distribution which is simply the product of cumulative distribution functions.

$$\mathbf{P}_a | \epsilon_a = \prod_{c \neq a} e^{-e^{-(\epsilon_c + V_c - V_a)}}$$

The probability that option a is chosen is therefore the product of probabilities that ϵ_a is greater than $\epsilon_c - V_a + V_c \quad \forall c$. Therefore the probability of alternative a being chosen is the expected utility of that alternative, which is obtained by calculating the integral over all values of ϵ_a and weighing them by their densities.

$$\mathbf{P}_a = \int_{-\infty}^{\infty} \left(\prod_{c \neq a} e^{-e^{-(\epsilon_b + V_b - V_a)}} \right) e^{-\epsilon_a} e^{-e^{\epsilon_a}} d\epsilon_a$$

Solving the integral, which is done in appendix (.1), will deliver the following expression

$$\mathbf{P}_a = \frac{e^{V_a}}{\sum_{c \in \mathcal{C}} e^{V_c}}$$

Assuming that V is linear $V_a = \boldsymbol{x}'\boldsymbol{\beta}$, we can finally write the closed form expression for the probabilities of the Logit Model.

$$\mathbf{P}_a = \frac{e^{\boldsymbol{x}'_a\boldsymbol{\beta}}}{\sum_{c\in\mathcal{C}} e^{\boldsymbol{x}'_c\boldsymbol{\beta}}} \tag{8}$$

Throughout this work we will refer to equation (8) as the Logit model. In the literature, although the nomenclature is not consistent, often the differentiation is made between the Multinomial Logit model (MNL) and the Conditional Logit model (CL). This work will refer to the MNL as a special case of the CL model where all attributes are individual specific. This convention is consistent with the software of choice, STATA, where the commands are `clogit` and `mlogit`, respectively.
Note that by regarding expression (8) one can see that the Logit model satisfies

$$\lim_{\boldsymbol{x}'\boldsymbol{\beta}\to+\infty} \mathbf{P}(Y = a|\boldsymbol{x}) = 1$$

$$\lim_{\boldsymbol{x}'\boldsymbol{\beta}\to-\infty} \mathbf{P}(Y = a|\boldsymbol{x}) = 0$$

and is therefore more adequate in estimating probabilities than linear probability models which had the problem that 'probabilities' were not confined to $[0, 1]$. Y stands for the stochastic decision process.

3.5 Normalization

First we will see that only differences in utility matter - a consequence of our underlying behavioural model - and that this has several important implications for the identification and specification of discrete choice models. Consider for example a set of explanatory variables that include

3.5.1 Socio-Economic Variables

We would like to use both alternative specific attributes as well as individual specific attributes to properly explain the decision process. For illustration purposes, remember that we have distinguished between the two. Let $\boldsymbol{x} = (\boldsymbol{y}, \boldsymbol{z})$, where $\boldsymbol{y}$ stands for attributes of the alternatives and $\boldsymbol{z}$ stand for the characteristics of the individual. Integrating this in (8) delivers

$$P_a = \frac{e^{\boldsymbol{y}'_a\boldsymbol{\beta}_y + \boldsymbol{z}'_a\boldsymbol{\beta}_z}}{\sum_j e^{\boldsymbol{y}'_a\boldsymbol{\beta}_y + \boldsymbol{z}'_a\boldsymbol{\beta}_z}}$$

This illustrates the fact that attributes that are constants over alternatives do not affect probabilities. Stated in a more intuitive way: only differences in utility matter when chosing from a choice set, not absolute values. If we add a constant k to every determined part of utility $V_c = \boldsymbol{x}'\boldsymbol{\beta} + k_c, \forall c \in \mathcal{C}$ then the differences between the utilities will not change and the decision of the agent is unaffected. This has important consequences for our model. If we want to incorporate individual specific characteristics which are by nature constant over all alternatives, we will have to modify our model. One method is to create a set of dummy variables for the choices an multiply them with the constant. Hereby we allow the coefficient to vary across choices instead of the characteristics. Another way to incorporate individual specific characteristics is the following: Suppose that the income of agent i (y) is one of the socio-economic characteristics in $\boldsymbol{z}$, with which we attempt to explain consumption choices for internet goods. The price p_c of alternative c is one the parameters in $\boldsymbol{y}$. Then using price relative to income p_c/y_i instead of incorporating them separately will do the trick. It incorporates both effects: utility of alternative a (V_a) will decrease if the price increases and increase when it becomes relatively cheaper.
The fact that only relative utilities matter also has identification issues.

3.5.2 Normalization of the Constant

When we specify the observed part of utility as $V_c = \boldsymbol{x}'\boldsymbol{\beta} + k_c, \forall c \in \mathcal{C}$ with k_c being a constant specific to alternative c. k_c can be interpreted as representing the average impact of all unincluded variables of our model on utility. These constants have the important task to ensure that the mean of the ϵ_c is 0. However, since only the differences between the utilities matter, it is possible to add any additional constant δ to k_c, yielding $\tilde{k}_c := k_c + \delta$ without changing the probabilities. Any model with the same differences is equivalent. As far as estimation is concerned, it is impossible to estimate these constants since there is an infinite number of δ with deliver the same utilities. However, this problem can easily be avoided by performing a harmless normalization. Simply set $k_a = 0$ for any a. The constant can then be interpreted as the average impact of unincluded variables relative to good a.

3.5.3 The Scale of Utility is Irrelevant

Since utility has no natural scale, we can also multiply utility U_a with a constant λ without affecting the outcome of the model. The order of alternatives with respect to the utility they yield will be unaffected by a homogenous transformation. Therefore the equations $U_a = V_a + \epsilon_a$ and $U_a = \lambda V_a + \lambda\epsilon_a$ are equivalent in that respect. However, the scale of utility is not irrelevant when it comes to interpreting the result of the estimation. An innocent normalization is therefore necessary and is usually obtained by simply normalizing the variance of the error terms. This is equivalent to normalizing the scale of utility because, when multiplying utility with λ alters the variance of the error term by a factor λ^2. For the Logit model where error terms are independent and identically distributed the standard normalization is achieved by setting the variance to $\frac{\pi^2}{6}$, which is chosen for convenience.

3.6 Strengths and Weaknesses of the Logit Model

The capability to capture taste variations between agents is one of its greatest features. Households can attach varying importance to different attributes of the alternatives. Whereas households with a high demand for multimedia applications consider bandwidth the most important criteria, low income households put more weight on the price of the internet connection. The Logit model can incorporate systematic variations in tastes, but will fail when taste varies with non-observed attributes or even randomly.

The greatest strength of the Logit model, however, is that it delivers simple closed form probabilities. This basically means that no further calculations are necessary to obtain the desired probabilities. However the simplicity of its closed form (8) is also a consequence of an often limiting assumption - namely the assumption of uncorrelated error terms. Only under this assumption do Logit probabilities have such a convenient form. This assumptions leads to a property called the

3.6.1 Independence of Irrelevant Alternatives

A critical property of the Mulitnomial Logit model is the Independence of Irrelevant Alternatives (IIA). It can be stated as follows: The ratio of the probabilities of any two alternatives is independent of the choice set. That is,

Definition 1 (IIA) *for any choice sets $\mathcal{S}$ and $\mathcal{T}$ such that $\mathcal{S} \subseteq \mathcal{T} \subseteq \mathcal{C}$, for any alternative a and b in $\mathcal{S}$, we have*

$$\frac{\mathbf{P}_{\mathcal{S}(a)}}{\mathbf{P}_{\mathcal{S}(b)}} = \frac{\mathbf{P}_{\mathcal{T}(a)}}{\mathbf{P}_{\mathcal{T}(b)}} \tag{9}$$

It can easily be seen that the Logit model fulfills this criteria. For any two alternatives a, b from our choice set $\mathcal{C}$ the relative probabilities are:

$$\frac{\mathbf{P}_a}{\mathbf{P}_b} = \frac{e^{V_a} / \sum_{c \in \mathcal{C}} e^{V_c}}{e^{V_b} / \sum_{c \in \mathcal{C}} e^{V_c}} =$$
$$= \frac{e^{V_a}}{e^{V_b}} = e^{V_a - V_b}$$

Since the relative probabilities of any two alternatives a and b only depend on their own characteristics and are completely independent of the existence of other alternatives, the Logit model is said to exhibit the feature of Independence of Irrelevant Alternatives (IIA).

3.6.2 Substitution Patterns

The IIA assumption will imply restrictive (and sometimes unrealistic) substitution patterns between the alternatives in our choice set. Consider a change in an attribute of one alternative. This will necessarily alter the probability of both this good as well as the probability of the other goods. Suppose for the sake of illustration a decrease in the price of cable internet. This will increase the probability of cable internet being chosen and since the probabilities of all goods necessarily have to sum up to one, decrease the probability of the other goods to be chosen. Consumers will substitute between goods in the choice set, in our case some consumers will move from other forms of internet to cable. The Logit model implies a certain pattern of substitution across alternatives - proportional substitution. The cross elasticity in a Logit model has the form[2]:

$$\mathcal{E}_{a,b} = -\beta_x x_b P_b \tag{10}$$

Where $\beta_x = \frac{\partial V_a}{\partial x_b}$ the change in deterministic utility of good a when an attribute x (the price) of good b changes. This cross elasticity is the same for all goods. Therefore an increase in the probability of cable internet of 20% will lead to a decrease of 20% for all alternatives. Furthermore, the relative probabilities of all other goods remain the same. This substitution pattern seems unrealistic in the case of internet products. Intuition would prompt us to believe that a price reduction of cable internet, which is a broadband product, will have a stronger affect on the probability of another broadband product (e.g. ADSL) than on mobile internet. Therefore, proportionate shifting, a manifestation of the IIA property, may deliver biased forecasts. This is clearly the most problematic weakness of the Logit model which we will deal with in the following chapter.

[2]see Derivation of Cross Elasticities for the Logit Model in the Appendix

3.6.3 Testing for IIA

Until now we have argued that proportional substitution patterns do not "seem" realistic. Sound economic research requires that we test for this attribute. The definition of IIA also provides a way to test for whether a data set satisfies the IIA condition or not. The model can be reestimated using only a subset of alternatives $\mathcal{D} \subset \mathcal{C}$. Since the relative probabilities of any two alternatives are supposed to be independent of the existence of other options, the estimates obtained from a subset should deliver the same probabilities as from the original set. This method is referred to as the Hausman test and its statistic can be written

$$(\boldsymbol{\beta}_s - \boldsymbol{\beta}_f)'[V_s - V_f]^{-1}(\boldsymbol{\beta}_s - \boldsymbol{\beta}_f) \sim \chi^2$$

where the subscript s indicates that the estimators where based on the restricted set of alternatives and f indicates that the full choice set was used. V is the covariance matrix of parameters.

3.6.4 Estimation

This model is solved by Maximum Likelihood. The coefficients $\boldsymbol{\beta}$ are obtained by maximising the log likelihood function, which means identifying those coefficients which makes the set observations most likely. This process involves multidimensional numerical methods such as variations of the Newton Method. McFadden demonstrated that the log-likelihood function with these choice probabilities is globally concave in parameters beta, and therefore the solution exists and is unique.2

$$LL(\boldsymbol{\beta}) = \sum_i \sum_{c \in \mathcal{C}} \mathbf{1}_i \ln \mathbf{P}_c$$

Where $\mathbf{1}$ is an indicator function taking the value one if agent i chooses alternative c and zero if not.

3.7 Generalized Extreme Value Models

Using the market for internet access as an example, I have argued that proportional substitution seems unrealistic. It does not seem compatible with observed decision behaviour. To overcome these restrictive substitution patterns between alternatives, that are a result of the IIA assumption, we will extend our model and relax this assumption. This led to the development of Generalized Extreme Value Models. They constitute a large class of models with the unifying attribute, that the unobserved portions of utility for all alternatives are jointly distributed generalized extreme value. This distribution is a generalization of the univariate extreme value distribution which has the critical attribute, that it allows for correlations over alternatives.

Definition 1 (Generalized Extreme Value Models (GEV)) *Consider a function $G = G(x_1, x_2, ..., x_m)$ with $G_c := \frac{\partial G}{\partial x_c}$. If the function G satisfies the following conditions:*

- *G is nonnegative.* $G(x_1, ..., x_m) \geq 0 \quad \forall x_c$
- *G is homogeneous-of-degree-one.* $G(\rho x_1, ..., \rho x_m) = \rho G(x_1, ..., x_m) \qquad \rho \in \mathbf{R}^+$
- $G \to \infty$ *when* $x_c \to \infty \qquad \forall c.$
- *for* $(i_1, ...i_k) \subseteq (1, ..., m)$, $\frac{\partial^k G}{\partial x_{i_1}, ..., \partial x_{i_k}}$ *is nonnegative if k is odd and non-positive if k is even.*

Then

$$P_c = \frac{e^{V_c} G_c}{G} \tag{11}$$

defines a probabilistic choice model from alternatives $c = 1, ..., m$, which is consistent with utility maximisation.

This definition of the class of models is given by McFadden (1978). It constitutes a very formal approach to GEV, basically defining a class of models over a class of functions with the stated attributes. The amount of intuition behind these attributes is very little, however, at the same time it accurately defines the GEV family. McFadden has proven that any function satisfying the above conditions, delivers a probabilistic choice model.
The Logit model is a member of this class. It arises when all correlations between alternatives are zero. In that case the GEV distribution is simply the product of independent extreme value distributions. The Logit choice probabilities are derived from (11) in Appendix (.3).
In the following two sections the Nested Logit Model will be introduced. The first approach will be an intuitive one, explaining the basic idea behind the model. Then the choice probabilities of the Nested Logit model will be derived from the basic form of GEV models (11).

3.7.1 The Nested Logit Model

The Nested Logit model is the most common member of the GEV family. This is due to the simplicity of its functional form, which still allows to incorporate a broad range of substitution patterns. It has found applications in various fields such as transportation research, logistics, marketing and will also be the models of our choice. The idea of the Nested Logit model lies in the grouping of "similar" alternatives into nests and thus structuring the alternatives. Similarity can - in this case - be characterized by correlation of the random terms - the sum of all unobserved attributes. If the random terms of two alternatives in the choice set correlate significantly then the econometrician has to conclude that there are similarities between these alternatives he was not able to capture with his chosen explanatory variables. The IIA condition will not hold. If this is the case one can try to group these alternatives to a subset with the hope that - considering they are similar in nature - the random terms of the probabilities in this subset are pure white noise.

The NLM is appropriate whenever the set of alternatives $\mathcal{C}$ can be divided into disjoint subsets - called nests $\mathcal{N} = \{N_1, ..., N_k\}$ with the properties $N_i \cap N_j = \emptyset, \forall i \neq j$ and $\bigcup_j N_j = \mathcal{C}$ and the IIA assumption holds within nests but does not have to hold across nests.

$$Cov(\epsilon_a, \epsilon_b) \neq 0, \forall a, b \in N_j$$

$$Cov(\epsilon_a, \epsilon_b) = 0, \forall a \in N_i \text{ and } b \in N_j, \qquad i \neq j$$

To turn back to our example, suppose there are five different internet connections available. ADSL, Cable, Narrowband, MobileInternet, and NoInternet. The observed substitution behaviour has the feature that whenever MobileInternet or Narrowband are taken out of the choice set, the probabilities of the two broadband connections (ADSL and Cable) increase by the same percentage. They however change by different percentages than the other alternatives. Then ADSL and Cable fulfill the IIA criteria - the relative probability of the two alternatives is irrelevant of the choice set.

$$\frac{\mathbf{P}_{\mathcal{S}(ADSL)}}{\mathbf{P}_{\mathcal{S}(Cable)}} = \frac{\mathbf{P}_{\mathcal{C}(ADSL)}}{\mathbf{P}_{\mathcal{C}(Cable)}}, \forall \mathcal{S} \subseteq \mathcal{C}$$

We can therefore put these two alternatives in a nest.

In the following we will decompose a (two-level) NLM into two Logit models, because resulting expression is easily interpretable and allows to

build up intuition for NLMs. Once we have divided the alternatives of the choice set $\mathcal{C}$ into nests - we will dissect the decision process essentially in two separate decisions - the resulting NLM can be regarded as the product of two Logit models. The choice set $\mathcal{N}$ of the 'first' decision contains the created nests $\mathcal{N} = \{N_1, ..., N_J\}$. The choice set of the 'second' decision are the alternatives $c_{j_1}, ..., c_{j_n}$ within one nest N_j. Note that the two parts of the decision process - even if they are going to be illustrated in a tree structure - do not happen sequentially. They are essentially one decision.
If V_c is split into two parts $V_c = W_{N_j} + Y_c$ (additively separable, linear-in-parameters) where W is the subset of attributes that is constant for all alternatives within a nest and Y varies over alternatives in a nest, then for this specification the probabilities can be written as the product of two standard Logit probabilities.

$$\mathbf{P}_a = \mathbf{P}_{a|N_j}\mathbf{P}_{N_j} \text{ with } a \in N_j \tag{12}$$

Where $\mathbf{P}_{a|N_j}$ is the conditional probability that alternative a is chosen, given that an agent has chosen the nest N_j and $\mathbf{P}_{N_j}$ is the probability that an alternative in nest N_j is chosen. Suppose our choice set consisted of four alternatives $\mathcal{C} = \{$ADSL, Cable, Mobile, Narrowband, NoInternet$\}$. We could group them into two different subsets $N_{Broadband}$ and $N_{NoBroadband}$. Then the choice probability for cable is equal to the probability that broadband is chosen times the probability that cable is chosen given $N_{Broadband}$ is chosen.

$$\mathbf{P}_{N_j} = \frac{e^{W_{N_j} + \lambda_{N_j} I_{N_j}}}{\sum_{N_l \in \mathcal{N}} e^{W_{N_l} + \lambda_{N_l} I_{N_l}}} \tag{13}$$

$$\mathbf{P}_{a|N_j} = \frac{e^{Y_a/\lambda_{N_j}}}{\sum_{c \in N_j} e^{Y_c/\lambda_{N_j}}} \tag{14}$$

$$I_{N_j} = \ln \sum_{c \in N_j} e^{Y_c/\lambda_{N_j}} \tag{15}$$

Equation (15) defines the inclusive value of the corresponding nest. Its function is to link the 'upper' model with the 'lower' model. Since the agent does not directly derive any utility from choosing a nest, the inclusive value takes over this function. I_{N_j} can be interpretated as the expected utility that an agent receives from choosing between the alternatives of nest $N_j \in \mathcal{N}$.
(13), (14), and (15) put together, yield the choice probability of the Nested Logit model.

$$\mathbf{P}_a = \frac{e^{V_a/\lambda_{N_j}} (\sum_{c \in \mathcal{N}_j} e^{V_c/\lambda_{N_j}})^{\lambda_{N_j} - 1}}{\sum_{N_l \in \mathcal{N}} (\sum_{c \in N_l} e^{V_c/\lambda_{N_l}})^{\lambda_{N_l} - 1}} \tag{16}$$

The values of λ_{N_j} can vary over nests reflecting different levels of independence between nests and therefore different levels of correlation between unobserved attributes. One can easily see that a value of $\lambda_{N_j} = 1$ for all nests N_j renders the standard Multinomial Logit model (8). This highlightes that the NLM is actually a generalization of the Logit model. This can also be seen regarding the distribution function. The NLM is obtained under the assumption that the vector of unobserved utility $\boldsymbol{\epsilon} = (\epsilon_1, ..., \epsilon_m)$, has cumulative distribution

$$F(\boldsymbol{\epsilon}) = \exp\Big(- \sum_{N_j} \Big(\sum_{c \in N_j} e^{-\epsilon_c/\lambda_j} \Big)^{\lambda_j} \Big)$$

Again, a value of $\lambda_j = 1, \forall N_j$ renders the distribution function where ϵs are idependent and with a univariate extreme value distribution, rendering the Logit model. Testing whether $\lambda_j = 1$, $\forall N_j \in \mathcal{N}$ is equivalent to performing the Hausman Test for each nest. λ_j can be interpretated as a measure of the degree of independence in unobserved utility among alternatives in nest j. A value $\lambda_j = 1$ incicates that alternatives within nest j are completely independent, and therefore uncorrelated. Consequently, if the alternatives in all nests are completely independent, then the IIA condition holds and the Nested Logit model reduces to the standard Logit model.

$$\mathbf{P}_a = \exp\left(- \sum_{N_j \in \mathcal{N}} \sum_{c \in N_J} e^{-\epsilon_a} \right) = \prod_{c \neq a} e^{-e^{-\epsilon_a}}$$

For a derivation of the choice probabilities from $G = \sum_{N_j \in \mathcal{N}} \left(\sum_{c \in N_j} (e^{V_c})^{1/\lambda_j} \right)^{\lambda_j}$ and showing that this model in a member of the GEV family please see Appendix (.4).

There are many different specifications of Nested Logit models. The difference "*lies in the explicit scaling of the deterministic utility component ...*"(7) The NLM version presented in this work is also referred to as the 2-Level Utility Maximization Nested Logit Model (RU2 UMNL). For the same identification problems mentioned earlier, one of the scale parameters needs to be normalized, which is normally achieved by seeting any scale parameter to 1. A normalization of a scale parameter of the 'upper' model will lead to the RU2 UMNL, whereas a normalization on the 'lower' level will induce a RU1 UMNL. No normalization at all will induce the non-normalized Nested Logit model (NNNL). However, only the RU2 UMNL is consistent with utility maximisation behaviour without any further restrictions7. This can be formally defined as follows

Definition 1 (Consistency with Random Utility Theory) *Consistency with utility maximization theory implies, that each alternative's choice probability* $\mathbf{P}_a$ *must not change when a constant term* δ *is added to each alternative's deterministic utility component* V_a.

$$\tilde{\mathbf{P}}_c(\tilde{V}_1, ..., \tilde{V}_J) = \mathbf{P}_c(V_1, ..., V_J) \quad \forall c \in \mathcal{C} \tag{17}$$

where

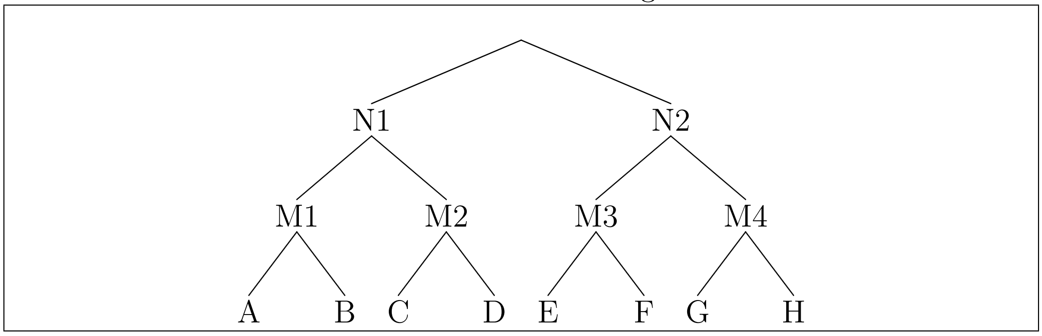
$$\tilde{V}_c := V_c + \delta \quad \forall c \in \mathcal{C} \tag{18}$$

The proof that the RU2 UMNL satisfies this condition will be given in Appendix (.5).

3.7.2 3-Level Nested Logit Models

If the correlation structure of the ϵs is more complicated than to be readibly captured with a two level nest structure the NLM can be extended to encorporate any level of nests. This allows more complex substitution patterns. A NLM is said to have two levels if it consists of one level of marginal probabilities ('upper level') and one level of conditional probabilities ('lower level'). In some situations two level NLMs seem to be inapropriate. A three level NLM is obtained when every nest $N_j \in \mathcal{N}$ is divided into subnests $M_k \subseteq N_j$ with $\bigcup_k M_k = N_j$ and $M_k \cap M_l = \emptyset$, $\forall k \neq l$ from which the agent then chooses an alternative. It can therefore be illustrated as a sequence of three decisions.

Table 3: 3 Level Nested Logit Model



Since we will apply a three level Nested Logit specification to our example of internet goods, the three level Nested Logit model will quickly be presented. The idea behind the structure of the model is equivalent to the two-level NLM enriched by one level of decisionmaking. In a three level specification the probability that alternative a is chosen is given by

$$\mathbf{P}_a = \mathbf{P}_{a|M_k|N_j}\mathbf{P}_{M_k|N_j}\mathbf{P}_{N_j} \tag{19}$$

where $\mathbf{P}_{a|M_k|N_j}$ is the conditional probability that alternative a is chosen, given subnest M_k was chosen given nest N_j was chosen. Since I am going to use three level Nested Logit models, I will just list the required probabilities and inclusive values. They are simply an extension of the two-level case.

- $$\mathbf{P}_{a|M_k|N_j} = \frac{\exp(V_{a,M_k,N_j}/\lambda_{M_k,N_j})}{\sum_{c\in M_k}\exp(V_{c,M_k,N_j}/\lambda_{M_k,N_j})}$$

- $$\mathbf{P}_{M_k|N_j} = \frac{\exp(V_{M_k,N_j}/\lambda_{N_j} + \lambda_{M_k,N_j}I_{M_k,N_j}/\lambda_{N_j})}{\sum_{M_l\in N_j}\exp(V_{M_l,N_j}/\lambda_{M_l,N_j} + \lambda_{M_l,N_j}I_{M_l,N_j}/\lambda_{N_j})}$$

- $$I_{M_k,N_j} = \ln\sum_{c\in M_k}\exp(V_{c,M_k,N_j}/\lambda_{M_k,N_j})$$

- $$\mathbf{P}_{N_j} = \frac{\exp(V_{N_j} + \lambda_{N_j}I_{N_j})}{\sum_{N_k\in N}\exp(V_{N_k} + \lambda_{N_k}I_{N_k})}$$

- $$I_{N_j} = \ln\sum_{M_k\in N_j}\exp(V_{M_k,N_j}/\lambda_{N_j} + \lambda_{M_k,N_j}I_{M_k,N_j}/\lambda_{N_j})$$

4 Data Description and Sampling Methods

This chapter will briefly introduce the data used in my estimations and address a number of sampling issues.

4.1 The Austrian Market for Residential Internet Access

On the Austrian market for internet access there are four different types of connection technologies:

- Cable is delivered over upgraded cable TV networks.
- DSL (Digital Subscriber Line) is a family of technologies that provide digital data transmission with high frequencies over the wires of a local telephone network.
- Mobile broadband describe various types of wireless high-speed internet access via UMTS/HSDPA.
- Narrowband (NB) stands for different types of ISDN and DialUp connections also using telephone lines.

When the survey was conducted more than one million out of 1.36 million broadband connections were held by residential customers. Firms were not included in the sample since they arguably are not part of the same market. Internet connections for companies fall under completely different pricing tariffs. By the end of 2006, 52% of all households in Austria had an internet connection with numbers growing fast and steadily. Narrowband connections still accounted for 19% of the market, however broadband connections - DSL, Cable, and Mobile - were on the rise. Mobile internet is the newest technology. The UMTS standard was introduced in 2003, the HSDPA standard only became available in 2006. The number of mobile internet users was about 220,000 - a small but fast growing share. However, internet connection availability is not spread evenly. Whereas more than 90% of households in Austria have DSL coverage, cable and mobile broadband tends to only be available in more densely populated regions. The cable network coverage is about 50%. HSDPA net coverage is usually given in cities with more than 5000 inhabitants. Narrowband internet - running over telephone lines - is available in 100% of households given that the survey was conducted using the telephone. I will therefore have to distinguish geographically according to the availability of internet access options when performing the estimations. Different choice sets will lead to different decisions. Since there was no detailed information regarding the availability of mobile internet for the sake of estimation we will assume that mobile internet is at least available in regions where cable internet is available. We will therefore distinguish between regions with the full choice set (Area 1) and regions where only DSL and Narrowband are available (Area 2). More detailed statistics on market shares can be found in chapter 6 in table (4.3).

4.2 Data Source

The liberalizing of the telecom sector in Europe began with a with a Green paper by the European Commission in 1987. Consequently, in the following years more and more of the predominantly state-run companies were privatized. However, the process of liberalization was accompanied by new regulatory duties and instruments. The EU directives were enacted in 2002 followed by the implementation into Austrian national law in 2003. The European Commission has enacted that national regulation authorities closely monitor these markets and has created a three-stage market analysis process, including the points (1) market definition, which is supposed to provide a framework for a (2) market analysis, which should determine whether effective competition prevails and should identify the companies with significant market share, and finally, if necessary, (3) impose regulatory instruments. This was laid down in the Directive 2002/21/EC of the European Parliament and of the Council of 7 March 2002 on a common regulatory framework for electronic communications networks and services.

The data stems from a survey commissioned by the RTR (the Austrian National Regulatory Authority) conducted in November 2006. This survey was conducted for the biannual market analysis required by the European Commission. Professor B. Yurtoglu, Department of Economics, University of Vienna, participated in the creation of this study by acting as an advisor to the Regulatory Authority. This work resulted in the paper "Demand Estimation and market definietion for broadband internet services" by Cardona, Schwarz, Yurtoglu and Zulehner. (8). Ida and Kuroda (9) published a similar study for Japan. At this point I would like to thank Professor Yurtoglu for his advice and support and hours of discussion.

In the original survey more than 4000 households were interviewed over the telephone. The data set includes the most important product specific information such as the chosen type of internet connection, speed and download limits, monthly expenses, and the availability of different internet connections in the respective area. It also included a set of individual specific observations such as age, household size, highest level of education etc. These socio-economic variables incorporate consumer heterogeneity and stand as proxies for taste variations. Since the answers of such surveys are not always reliable additional price, volume and download limit information was obtained from the websites of the internet service providers. Information on internet access availability according to postal code was also obtained. The responses given on the telephone interview were then compared to the official information. Observations were deleted if either the answers concerning the availability of certain internet connections did not match our information about the availability in the households respective area received from the operator, or if the answer about the price of an internet connection differed substantially from information collected on the ISP's internet website. The underlying argument for this data selection process is that a consumer's choice based on too little or wrong information cannot be fully rational and should therefore not be used for any estimations.

Table (4.2) provides an overview over the explanatory variables used in the models to come:

Table 4: Explanatory Variables

Var	Description
Survey Data	
choice	Type of internet access in the household (HH)
area	Area code of the household
sex	Head of the household is male/female
nb_usr	Number of people using the internet
nb_pc	Number of PCs in the HH
nb_lt	Number of laptops in the HH
nb_hh	Number of people living in the HH
age	Age of the head of the HH
income	Income of the head of the HH
	Highest level of education:
edu_com	Compulsory schooling
edu_hsno	High School without graduation
edu_hsyes	High School with graduation
edu_uni	University degree
price	Price per month as stated in the interview
volume	Monthly included data volume
speed1	Maximal download speed
d_flat	Flat rate dummy
dsl_av_i	DSL availability according to interview
cab_av_i	Cable availability according to interview
mob_av_i	Mobile internet availability according to interview
Other Sources	
price2	Monthly price according to the operator
speed2	Maximal download speed according to the operator
dsl_av_ac	DSL availability according to area code
cab_av_ac	Cable availability according to area code
mob_av_ac	Mobile internet availability according to area code

4.3 Descriptive Statistics

In order to give the reader some feeling for the data set I have included some descriptive statistics - both product specific as well as individual specific

- in tables (4.3) and (4.3). The household specific variables include age, household size, income, sex, and the highest level of education.

Table 5: Household Specific Statistics

	DSL	Cable	Mobile	Narrowb.	No Int.
Mean age (head of household)	45.5	43.9	41.2	46.8	61.6
Mean household size	3.0	2.8	2.8	3.0	2.0
Mean income (head of household)	2566	2524	2147	2570	1507
Gender: female	46.8%	47.8%	43.7%	48.9%	66.4%
Education:					
Compulsory school	35.1%	31.6%	37.5%	31.4%	72.3%
High school without graduation	23.1%	14.9%	22.9%	24.0%	15.7%
High school with graduation	25.4%	30.7%	29.2%	29.7%	9.2%
University degree	16.3%	22.8%	10.4%	14.9%	2.8%

The product specific variables include price, download speed, and download volume included in the monthly fixed price for non-flat-rate products. While no flatrate products exist for mobile and narrowband connections the percentage of DSL and cable users with a flat rate tariff is 9.9% and 53.2% respectively.

Table 6: Product Specific Statistics

	Mean	Std. Dev.	Min	Max
Price per month (€)				
DSL	32.02	10.50	9.9	85
Cable	39.38	14.41	10	94.9
Mobile	37.12	16.52	9.5	99
Narrowband	20.44	14.11	4	60
Download speed (kbit/s)				
DSL	1,345	1,040	0,248	6,144
Cable	2,938	2,511	0,128	16,384
Mobile	900	0	900	900
Narrowband	56	0	56	56
Download volume included (MB)	(non flat-rate)			
DSL	1973	3706	250	30000
Cable	5109	8466	100	30000
Mobile	952	961	250	4000
Narrowband	0	0	0	0

Surveys tend to be very expensive. Research budgets therefore constitute one of the greatest constraints to the size of a sample. In smaller samples one can often observe that the distribution of the alternatives in the sample differs significantly from the known distribution of alternatives in the overall population. This is problematic since a bias in the sample will be transmitted to the parameter estimates. A correction of the distribution by simply weighing the observations will do the trick. In our sample users of DSL were clearly overrepresented. Therefore I corrected for this bias by weighting my observations in order to match the distribution of Statistik Austria 2006. Table 4.3 depicts the market shares of different kinds of internet connections as well as the distribution in my sample. 'Sample weighted' lists the corrected distribution used in my estimations.

Table 7: Observations

	DSL	Cable	Mobile	Narrowb.	No Int.	Total
Maket Share	13.45%	9.97%	1.22%	16.21%	59.16%	100%
Sample unweighted	553	228	48	176	1007	2012
Sample Share	27.49 %	11.33%	2.39%	8.75%	50.05%	100%
Sample weighted	271	201	25	326	1189	2012
All connections region	27.55 %	17.59%	2.49%	7.47%	44.90%	
Area One:						
Vienna	21.77%	21.77%	1.64%	5.34%	49.49%	
Rest of Austria	31.27%	14.91%	3.03%	5.34%	41.95%	
Area Two	31.70%	-	1.53%	11.49%	55.28%	

I have differentiated between Vienna and the rest of Austria due to the fact that the capital has a significantly higher coverage and therefore a much higher market share of cable internet. This will be considered in the estimations to come. Only an insignificant number of people had more than one internet access. Most of them combined mobile internet with an other form of access. However, this number was so small that I decided not to create an own category. Another way to deal with the problem of alternatives with too small shares in the sample is choice-based sampling.

4.4 Suggestions for the Sampling Method

Whenever one alternative from the choice set is very rare, this will also manifest in the sample. The quality of the estimation will suffer from this lack of a reasonable number of observations. A common method to fix this problem is to deliberately skew the sampling process in favour of the underrepresented alternative. The survey actively includes more agents who have chosen the certain alternative in order to obtain a more balanced sample than random sampling would produce. The oversampling of one alternative necessarily creates a biased mix of observed decisions, which will have to be corrected by weighing the alternatives respectively. Manski and Lermann (1977) have derived the WESML - the weighted endogenous sampling maximum likelihood estimator - for situations like these. One necessary requirement is that the true population distribution of alternatives is known.

> *If the researcher is using a purely choice-based sample and includes an alternative-specific constant in the representative utility for each alternative, then estimating a Logit model as if the sample were exogenous produces consistent estimates for all the model parameters except the alternative specific constants. Furthermore, these constants are biased by a known factor and can therefore be adjusted so that the adjusted constants are consistent* [3]

If we are particularly interested in knowing whether mobile internet - a new and therefore not widespread technology - is a close substitute of ADSL or Cable a random sample would have to be very large in order to assure a reasonable number of households with mobile internet.

Another important sampling issue is the question whether to use transaction or holding data. The internet market is characterising by ongoing strong technological progress. The mass use of the internet is a recent development and still today constantly new applications are created and perfected and technology improved. It can therefore hardly be considered a market in equilibrium. Bearing in mind that people always lag behind technological progress - it involves time and costs to adapt and inform - one has to raise the question whether the market shares are the outcome of sound decision making at any point in time. The 'captive consumer' does not react even to a strong increase in price because he either doesn't care or is not sufficiently informed. Therefore I would suggest a transactions rather than a holdings approach. This means that only households who have either changed their internet connection or who got their first internet connection in a period of time should be included in the sample. Arguably, a transaction approach more closely reflects the decision making at a specific point in time since these households have actually expressed, through their choices, preferences for a specific product.

[3] Kenneth Choice (2003)

The third and final point that I would like to criticize regards the set of chosen explanatory variables obtained from the telephone survey. In order to capture consumer heterogeneity or taste variation, socio-economic data was collected from every interviewed person. To name just a few: age, income, size of the household, number of PCs in the household,... These entities should serve as proxies for the strongly varying profile of internet use. But why were the households not asked directly about the way they use their internet access? Questions such as: What are the main applications of the internet connection in your household? Is there an online gamer in your household? How many hours do you use your internet connection on average per week? Do you download music or film from the internet?
The answers to those questions would have shed a lot more light on consumption profiles. They have arguably more explanatory power as far as taste variation is concerned. An internet user whose only application is online banking is likely to be more price elastic than a online gaming afficionado. Whereas the online banker can go to a real bank there is no close substitute for the online gamer. Therefore, an additional differentiation according to internet consumption profile would have made more sense. The quality with which we can explain consumer heterogeneity will have a great influence on our estimation results.

4.5 Estimating the Parameters of the Non-chosen Alternatives

Since we can only observe the attributes of the chosen alternative an obvious question is: What proxies should be used for the attributes of non-chosen alternatives? For the application of discrete choice models it is necessary to allocate price and other characteristics of internet connections to non-chosen options. There is, however, no standard methodology to obtain these figures. The researcher is left with his intuition. Since prices, volumes and speeds offered by the large number of different ISPs were far from uniform, one of the most delicate tasks in the process of this allocation was to identify an appropriate average for price, volume and speed for every group. The approaches are explained briefly by type of internet access in the following sections.

4.5.1 Narrowband

The allocation of narrowband prices to households that currently do not have internet access, or a broadband connection is tricky. The individual specific variables that were most significant in explaining narrowband expenses were region and age. Therefore four geographical groups were formed, within which two age subgroups were created, differentiating between those younger and older than fifty. The group price averages were then allocated to all its members. Speed and included volume does not vary for narrowband households. Therefore 56 kbit/s maximum speed and zero included megabytes were assigned.

4.5.2 Cable

There are many cable operators in Austria, however the packages they offer are similar in terms of bandwidth and included volume. Therefore the "low", "medium" and "high" usage packages of the seven biggest cable ISPs were collected and allocated to DSL, NB, Mobile, and NoInternet households in the following manner: Households without internet and narrowband households that spend on average less than 25 were assigned the "low" usage package. Narrowband households that spent more than 25 were assigned the 'medium' usage package, DSL households were assigned the package closest in bandwidth and mobile households the package closest in price.

4.5.3 DSL

DSL price, speed and volume data was allocated in the following manner: The tariffs of the three largest DSL operators were collected from their websites. In areas with more than one active operator price average was calculated weighted by market share. No internet and narrowband households were assigned the "low" user package, cable users were assigned the package closest to their own in terms of price, speed and volume. Mobile users were assigned the package closest in price.

4.5.4 Mobile

The price and volume for mobile internet was allocated according to the monthly amount paid by the household for their current internet access. Weighted price averages of the four big mobile operators were used. The download speed was assumed to be fix and 900kbit/s.

5 Selecting the Appropriate Discrete Choice Model

Even within the family of Logit models there are endless possibilities to specify the model. The main distinguishing features in my case are the choice between estimating joint coefficients or alternative specific coefficients and the choice of subdividing the choice set into nests. As far as the first question is concerned, the first approach was to estimate a price coefficient β_c for every alternative. This would allow us to calculate cross-price elasticities between all pairs of alternatives in the choice set, delivering 25 different elasticities for area one. However our data set lacked the explanatory power for so many estimated coefficients. Therefore I estimated a joint coefficient β for all alternatives in area one.

$$\beta_c = \beta \qquad \forall c \in \mathcal{C}$$

This limits the number of estimable coefficients to one own-price and one cross-price elasticity for every alternative, delivering a number of ten altogether. As a result the interpretation of β also has to be altered. The joint price coefficient β has to be interpreted as the average value of price in terms of utility.

In area two, where only DSL and NB are available, the estimation of one parameter per alternative was feasible. The most promising method to obtain possible settings for Nested Logit models is to simply look at the data in the following manner. Tabulate the alternative in the choice set distinguishing between regions where single alternatives are not available. This (completely made up) table depicts this kind of illustration.

Table 8: Market shares according to availability

Alternative	AllConnections	NoCable	NoDSL	NoNarrowb.	NoMobile
Cable	30%	-	60% (+100%)	36.6% (+25%)	34.3% (+14.3%)
DSL	40%	60% (+50 %)	-	50% (+25%)	45.7% (+14.3%)
NB	20 %	20% (+0%)	25% (+25%)	-	20% (+0%)
Mobile	10%	20% (+100%)	15% (+50%)	13.3% (+33.3%)	-

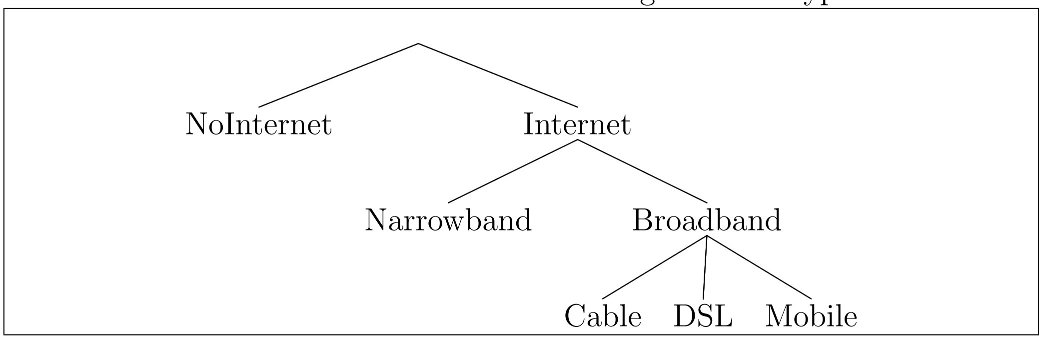
Under the assumption that consumer tastes are homogenously distributed over all regions one can draw certain conclusions. Bear in mind that two alternatives in the same nest have to satisfy the IIA assumption. This means that no matter what alternative is missing, the substitution pattern of two alternatives in the same nest has to be reasonably similar. As one can see this is the case for DSL and Cable and is clearly not the case for NB and Mobile.

The conclusion one could draw from this result is that DSL and Cable belong to the same nest and NB and Mobile do not. Unfortunately, our data does not allow this kind of analysis. Firstly there is no region is Austria where NB is not available since all it requires is telephone lines. Also information on the availability of mobile internet was not accessible at the time of the collection of the data. Secondly there seem to be strong differences in terms of socio-demographic qualities. E.g. in urban areas the rate of academics is far higher than in the hinterland. The assumption that tastes do not vary with the regions is therefore at least questionable.

5.1 Area 1

Since the bespoken method is only partially appropriate in our case the researcher has to rely largely on his intuition. Since the researcher is very likely to have an internet connection himself he can try to analyse his own decision process. One possible model of a decision making process is the following: Arguably the first decision would be whether to make a purchase or not. All other considerations are secondary. Once I have decided to make a purchase the next question would be whether to get broadband internet or not. A decision making process of this type can be illustrated in the following way:

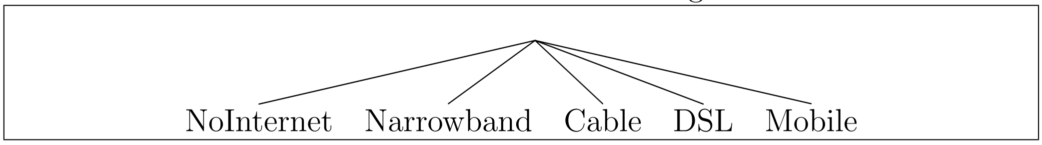
Table 9: Area 1: 3-Level Nested Logit Model Type A



There are, however, a number of different and also plausible structures to model this process. In the process of finding the model that fits the data best I have tried the following:

A Conditional Logit model.

Table 10: Area 1: Multinomial Logit Model



A two-level Nested Logit model with two nests. The only distinctive criteria is internet or not.

Table 11: Area 1: 2-Level Nested Logit Model

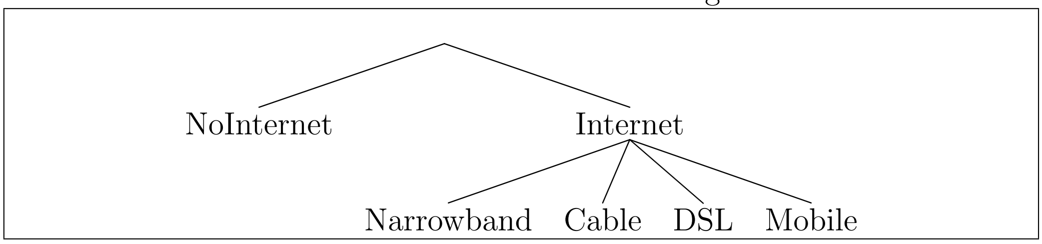
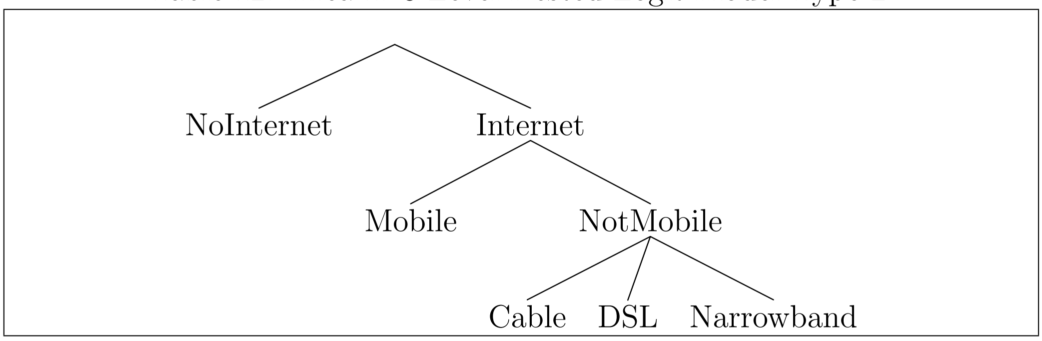


Table 12: Area 1: 3-Level Nested Logit Model Type B



5.2 Area 2

The limited choice set of this area ($\mathcal{C} = (DSL, NB, NI)$) affected the modelling decision in two ways: Firstly I decided to estimate a separate price coefficient for both DSL and NB. Secondly the number of possible settings for discrete choice models was dramatically reduced. I have tried both a Conditional Logit model and a 2-level Nested Logit model.

Table 13: Area 2: Conditional Logit Model

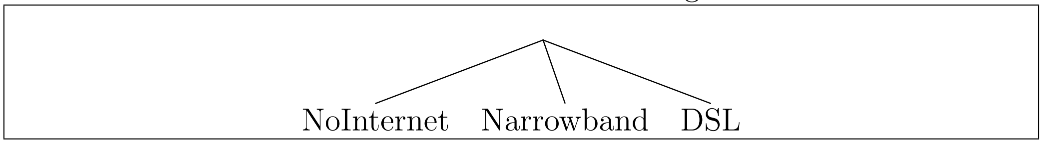
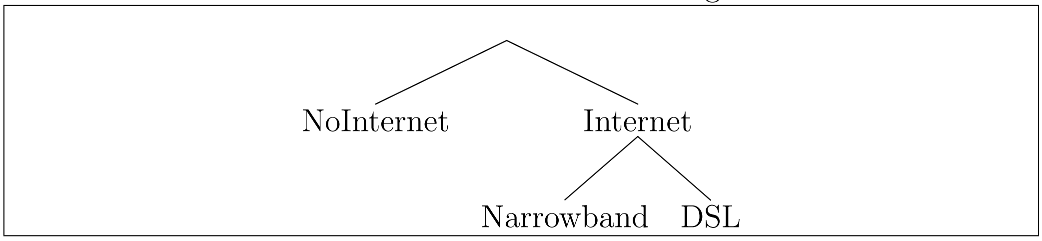


Table 14: Area 2: 2-Level Nested Logit Model



5.3 Comparing Models - Goodness of Fit

Model comparison was undertaken by means of three indicators. Sign conditions of variables, their statistical significance (t-values), and degrees of fitness such as the McFadden R^2 or Akaike's Information Criterion. Sign conditions are a very strong method for model selection. Coefficients with signs that do not match our expectations based on economic theory e.g. a positive coefficient for the price of an alternative are an excellent indicator that the model doesn't properly fit the data. A statistic that describes how well the model fits the data is the likelihood ratio index also referred to as the McFadden R^2. It can be used for every discrete choice model since it is based on the log-likelihood function which is the estimation method for all discrete choice models and is defined as

$$\rho^2 = 1 - \frac{LL(\hat{\boldsymbol{\beta}})}{LL(\mathbf{0})}$$

where $LL(\hat{\boldsymbol{\beta}})$ is the value of the likelihood function at the estimated parameters and $LL(\mathbf{0})$ is its value when all parameters are zero, indicating that our variables have no explanatory power at all. One can easily see that this statistic is confined to $[0, 1]$. Its value is 0 when our model does not do any better than the zero parameter model and 1 if our model can perfectly explain the decisions of the agents. Note that this statistic cannot be interpreted in the same fashion as the R^2 in linear regression models, which indicates the percentage of the variation in the dependent variable which can be explained by the model. It can, however, be used to compare different models with respect to goodness of fit as long as the two models were estimated using a) the same data and using b) the same set of alternatives. If this is not the case the likelihood function cannot be used to compare models.

5.4 Results: Area 1

In my first attempt I tried to estimate price coefficients for all five alternatives rendering a total of 25 elasticities. This did not prove successful since this caused insurmountable computational difficulties. Either the software program in use was not able to solve the resulting maximum likelihood estimation or the results were miles from expected values. In order to reduce the number of demand parameters I had to put some structure on my demand function. Therefore I chose to estimate one joint price coefficient for all alternatives. This delivered the following results:
The Conditional Logit model (5.1) seemed not to capture our data correctly. Not only were sign coefficients problematic (App.7) - e.g. the price coefficient was positive - but also the Hausman test - testing whether coefficients were stable when one alternative is dropped - clearly rejects the assumption of IIA. This result is very robust since when I dropped the alternatives one by one the χ^2 test value never dropped below 46.78. Not surprisingly, the fitness of this model had to be rejected.
In order to gain a feeling for how robust my estimation results were I calculated the elasticities for all the aforementioned models and variations of them and compared them. The results were not very stable. Elasticities varied by $+/-0.3$ in the most cases. One in five models delivered completely altered results such as different signs. While the absolute values of the coefficients varied to some extent, the differences between the coefficients were surprisingly robust. In almost all models the elasticity for mobile internet was significantly smaller than the elasticity for DSL and Cable. Also the elasticities for immobile broadband connections never differed much.
The model that delivered the most promising results was the three level Nested Logit model where the distinguishing criteria in the second level was broadband or not. (3.7.2) The estimation results are depicted in table

Table 15: Area 1: Nested Logit Regression

Var	**Coef.**	**Std. Err.**	**z**	$P > \|z\|$
3rd level:				
price2	-0.0428	0.0057	-7.49	0.000
volume	0.0131	0.0032	4.06	0.000
speed	0.5605	0.0588	9.53	0.000
dummy_cable	0.4890	0.1911	2.56	0.011
dummy_dsl	0.9877	0.1706	5.79	0.000
2nd level: broadband vs. narrowband				
age*bb	-0.0300	0.0044	-6.88	0.000
edu_com*bb	-0.6043	0.2295	-2.63	0.008
edu_hsno*bb	-0.5393	0.2501	-2.16	0.031
edu_hsyes*bb	-1.1289	0.2187	-5.16	0.000
1st level: internet vs. no internet				
edu_com*i	-2.0857	0.5869	-3.55	0.000
edu_hsno*i	-1.1598	0.5961	-1.95	0.052
edu_hsyes*i	-0.2283	0.5254	-0.43	0.664
nb_hh*i	0.4271	0.0709	6.02	0.000
inclusive values:				
broadband	1.2810	0.1950	6.57	0.000
internet	3.7027	0.5677	6.52	0.000

Where * indicates that the variable was interacted with a dummy variable - bb indicates the broadband dummy, i indicates the internet dummy. The price coefficient has the expected negative sign and volume and speed the expected positive sign. The negative sign for the education dummies for high school graduates seems a little bit surprising. This would indicate that a higher education reduces the probability of having internet at home.

The derivation of the elasticity formulas from the Logit and Nested Logit formulas can be found in Appendix .6. The only additional formula necessary for the calculation of the elasticities is:

$$\frac{\partial Y_a}{\partial x_a} = \hat{\beta}_a(1 - \text{share of a})$$

The resulting elasticities were:

Table 16: Area One: Elasticities

Alternative	**Cross Price Elasticity**	**Own Price Elasticity**
Cable	0.9891	-0.9077
DSL	0.6320	-0.6623
NB	1.6398	-0.3246
Mobile	0.2515	-1.1653

Economic theory would prompt us that the magnitude of these elasticities are too low. One would expect them to be higher than in regions with a limited choice set. This is, however, not the case. The differences in elasticity seem to make good sense. The NB own-price elasticity is very low. Since this is the cheapest form of internet access, people that really depend on the internet will be prepared to pay even higher prices. The own-price elasticities of all other alternatives are higher. Consumers have the possibility to downgrade their broadband connection to a narrowband connection if prices are increased. Even though the mobile internet users in the sample were only few, the result seems to be in accordance to our intuition. The own-price elasticity is the highest indicating that consumers would readably give up the mobility feature if the price was to be increased.

5.5 Results: Area 2

The results from the second area were much more stable. Even though I had to reject the Conditional Logit model on basis of the Hausman test the resulting coefficients were very similar to the ones obtained by the 2-level Nested Logit model. I assume that the smaller number of coefficients to be estimated makes the model more balanced. I was even able to estimate separate price coefficients for DSL and NB.

Table 17: Area 2: Nested Logit Regression

Var	**Coef.**	**Std. Err.**	**z**	$P > \|z\|$
2nd level:				
price2*dsl	-0.1045	0.0095	-11.00	0.000
price2*nb	-0.0448	0.0062	-7.24	0.000
speed	1.3572	0.1406	9.65	0.000
volume	0.4958	0.1503	3.30	0.001
1st level: internet vs. no internet				
age*i	-0.0533	0.0090	-5.95	0.000
edu_hsno*i	-1.3232	0.3677	-3.60	0.000
edu_hsyes*i	-0.8376	0.4122	-2.03	0.042
nb_hh*i	0.3773	0.0899	4.20	0.000
dummy_i	4.3144	0.8112	5.32	0.000
inclusive value:				
internet	7.7006	1.0599	7.27	0.000

The resulting elasticities were:

Table 18: Area Two: Elasticities

Alternative	**Own-Price Elasticity**	**Cross-Price Elasticity**
DSL	-1.5228	0.2284
NB	-0.5717	0.6385

These results seem plausible. Even though there is no other broadband connection available in these regions the own-price elasticity of DSL is high indicating that consumers readily switch to NB, or even NoInternet. The own-price elasticity of NB is smaller suggesting that people depend on their internet connection and are ready to pay even higher prices. Cross-price elasticities more or less tell the same story.

6 Conclusio

The starting point of my thesis was a simple two-stage least squares estimation which could provide price and income elasticities for the average broadband product in Europe. Its simplicity relied mainly on the assumption of homogenous goods and homogenous consumers. It cannot differentiate between different forms of internet access or different broadband packages. It also fails to capture differences in the taste of consumers. Whereas these

assumptions keeps the model simple and the data used readably available, the conclusions that can be drawn from it are very limited. Therefore models for differentiated products were developed. They represent a clear generalization of the homogenous goods market. However, this generalization and power comes with a number of disadvantages. In order to obtain demand elasticities for an n goods market one has to estimate n^2 parameters. The sheer number of parameters makes estimation difficult. To solve this problem economist have come up with the idea to base the demand function on individual demand, which can be obtained from a model of discrete choice. This approach created a very rich but also more complicated model. Models of this type circumvent the problem of price endogeneity by the (reasonable) assumption that a single consumer does not have any price setting power. This, however, comes with the disadvantage that it completely neglects supply side and market equilibrium considerations.
From a structural perspective the neoclassical model has very little structure. Only the choice of the set of variables included in the models stems from economic theory. The Discrete Choice model is highly structured. There an underlying behavioural model and a number of structural assumption for example abouththe distribution of the error terms. In Discrete Choice models both prices and unobserved product characteristics enter demand equations in a nonlinear fashion.

Discrete Choice models have become very popular because they have the following great strengths: They allow for a high degree of product differentiation and account for consumer heterogeneity by using socio-economic data standing as proxies for variations in taste. However, this comes at the cost of having to conduct a consumer survey. Industry or country level data tends to be more readably available.

The power of these two features can best be illustrated by giving some examples of possible research questions which would call for a discrete choice approach. Suppose you are a company and would like to introduce a new internet connection package. A most obvious question would concern the potential demand for this particular product given its product characteristics. One powerful feature is that the discrete choice model can estimate the potential demand for new products, especially one that are dissimilar to goods already on the market. It delivers a quantifiable measure of how much customers value certain characteristics of a good. It allows us to easily move between statements about aggregate demand and statements about consumer utility. Hence, it has the power to predict how the willingness to pay will increase when the bandwidth of the product is increased. Suppose you are a competition authority. This model can deliver estimates of demand elasticities between any two goods. It can therefore be used for a market definition, an important part in any market analysis. It can indicate, whether the market lacks competition or not. It can also predict changes in the structure of demand as a result of socio-demographic shifts in the population. How will a change in the sociodemographic structure of the country effect demand of high speed internet goods?

Another important result was the observation that with an increasing number of parameters to estimate the results became less and less robust. It would be very interesting to analyse in what way the robustness depends on the number of goods in the choice set, which determines the number of parameters to be estimated.

In this work I could observe the following: In regions where only two internet connections were offered the model delivered more reliable results. In those regions where all four connection types were available I had to restrict the number of price elasticities to make the model feasible at all. Either the results were completely off track or the software failed to compute the results. Even after the restriction the results turns out to be rather instable. Small differences in the specification of the model, such as adding one or more explanatory variables, sometimes completely alters the outcome.

One downside of structured models is that they are based on a series of maintained and therefore non-testable assumptions. This blurs the troubleshooting phase of research. If your estimation results are completely off the track there is no real way to find out where the problem lies or what assumption was not met.

6.1 Limitations of Discrete Choice Models

One very important limitation to the model is that it is static. It cannot incorporate switching costs between different internet connection types. Even though installation fees and set-up fees are free of charge for most broadband products other forms of switching costs could be significant. A model that can capture switching behaviour between different alternatives will likely deliver more reliable predictions.

Another important limitation to this demand structure, which is caused by the very nature of discrete choice models, is that these models rule out choosing several alternatives. This does not appear too limiting in the case of internet products - people very rarely have more than one internet connection. This could, however, pose a serious problem, when modelling car purchases. We have seen, that this problem can sometimes be solved elegantly by including multiple good alternatives to the choice set. This trick is not always possible. Further, the number of alternatives in the choice set should not be too large in order to produce a computationally feasible model given a certain sample size.

7 Appendix

.1 Derivation of the Logit Probabilities

$$
\begin{aligned}
\mathbf{P}_a &= \int_{\epsilon_a} \left(\prod_{b \neq a} e^{-e^{-(\epsilon_b + V_b - V_a)}} \right) e^{-\epsilon_a} e^{-e^{\epsilon_a}} d\epsilon_a = \\
&= \int_{-\infty}^{\infty} \left(\prod_{b} e^{-e^{-(\epsilon_b + V_b - V_a)}} \right) e^{-e^{\epsilon_a}} d\epsilon_a = \\
&= \int_{-\infty}^{\infty} \left(exp \left(\sum_b -e^{-(\epsilon_b + V_b - V_a)} \right) \right) e^{-e^{\epsilon_a}} d\epsilon_a = \\
&= \int_{-\infty}^{\infty} \left(exp \left(-e^{\epsilon_b} \sum_b e^{-(V_b - V_a)} \right) \right) e^{-e^{\epsilon_a}} d\epsilon_a = \\
&= |t = e^{-\epsilon_a}, \frac{d\epsilon}{dt} = -e^{-\epsilon_a}| = \\
&= \int_0^{\infty} exp \left(-t \sum_b e^{-(V_b - V_a)} \right) dt = \\
&= \frac{exp \left(-t \sum_b e^{-(V_b - V_a)} \right)}{\sum_b e^{-(V_b - V_a)}} |_0^{\infty} = \\
&= \frac{1}{\sum_{c \in \mathcal{C}} e^{-(V_c - V_a)}} = \frac{e^{V_a}}{\sum_{c \in \mathcal{C}} e^{V_c}}
\end{aligned}
$$

.2 Derivation of Logit Elasticities

In this section we will derive both own elasticities and cross elasticities from logit choice probabilities. Elasticities basically answer the question: To what extent do probabilities change as a result of a change in an observed attribute (e.g. how much will demand for ADSL increase when the price is reduced, or the bandwidth increased?).

Own elasticity ($\mathcal{E}_{a,a}$): The change in probability that agent i picks alternative a given a change in the price x_a of alternative a can be calculated:

$$
\begin{aligned}
\frac{\partial P_a}{\partial x_a} &= \frac{\partial (e^{V_a} / \sum_{c \in \mathcal{C}} e_c^V)}{\partial x_a} \\
&= \frac{e^{V_a}}{\sum_{c \in \mathcal{C}} e^{V_c}} \frac{\partial V_a}{\partial x_a} - \frac{e^{V_a}}{(\sum_{c \in \mathcal{C}} e^{V_c})^2} e^{V_a} \frac{\partial V_a}{\partial x_a} \\
&= \frac{\partial V_a}{\partial x_a} (P_a - P_a^2) \\
&= \frac{\partial V_a}{\partial x_a} P_a (1 - P_a) \\
&= \beta_x P_a (1 - P_a) \qquad \text{"if } V_a \text{ is linear in coefficients"}
\end{aligned}
$$

Therefore the own price elasticity $\mathcal{E}_{a,a}$ takes the form:

$$\begin{aligned}
\mathcal{E}_{a,a} &= \frac{\partial P_a}{\partial x_a}\frac{x_a}{P_a} \\
&= \frac{\partial V_a}{\partial x_a} P_a(1-P_a)\frac{x_a}{P_a} \\
&= \frac{\partial V_a}{\partial x_a} x_a(1-P_a) \\
&= \beta_x x_a (1-P_a \qquad \text{"if } V_a \text{ is linear in coefficients"}
\end{aligned}$$

Cross price elasticity ($\mathcal{E}_{a,b}$): The change in probability that agent i picks alternative a given a change in the observable attributes of alternative b can be calculated:

$$\begin{aligned}
\frac{\partial P_{i,a}}{\partial x_b} &= \frac{\partial(e^{V_a}/\sum_{c\in\mathcal{C}} e_c^V)}{\partial x_b} \\
&= -\frac{e^{V_a}}{(\sum_{c\in\mathcal{C}} e^{V_c})^2} e_b^V \frac{\partial V_a}{\partial x_b} \\
&= -\frac{\partial V_a}{\partial x_b} P_a P_b \\
&= x_b P_a P_b \qquad \text{"if } V_a \text{ is linear in coefficients"}
\end{aligned}$$

Cross elasticities will look like:

$$\begin{aligned}
\mathcal{E}_{a,b} &= \frac{\partial P_a}{\partial x_b}\frac{x_b}{P_a} \\
&= -\frac{\partial V_a}{\partial x_b} x_b P_b \\
&= -\beta_x x_b P_b \qquad \text{"if } V_a \text{ is linear in coefficients"}
\end{aligned}$$

.3 Proof: Logit Model is a Member of the GEV family

Consider $G := \sum_{c\in\mathcal{C}} e^{V_c}$. This function satisfies all properties from (11) since 1) the sum of positive terms is positive, 2) $\sum_{c\in\mathcal{C}} \rho e^{V_c} = \rho \sum_{c\in\mathcal{C}} e^{V_c}$ 3) $e_c^V \to \infty \Rightarrow C \to \infty$ and finally 4) $\frac{\partial G}{\partial e_c^V} = 1 \quad \forall c \in \mathcal{C}$ and $\frac{\partial^k G}{\partial e^{V_{i_1}},\ldots,\partial e^{V_{i_k}}} = 0 \quad \forall k \geq 1$. By simply inserting the defined function in (11) we obtain:

$$P_a = \frac{e^{V_a} G_c}{G} = \frac{e^{V_c}}{\sum_{c\in\mathcal{C}} e^{V_c}}$$

.4 Derivation of the Choice Probabilities of the Nested Logit Model

Starting point is the function $G = \sum_{N_j \in \mathcal{N}} \left(\sum_{c \in N_j} (e^{V_c})^{1/\lambda_j} \right)^{\lambda_j}$. First we will show that this function satisfies the four necessary qualities of the GEV family. G is clearly nonnegative since it consists of a sum of nonnegative e^x terms. G is homogenous of degree 1 since $\sum_{N_j \in \mathcal{N}} \left(\sum_{c \in N_j} (\rho x)^{1/\lambda_j} \right)^{\lambda_j} = \rho \sum_{N_j \in \mathcal{N}} \left(\sum_{c \in N_j} x^{1/\lambda_j} \right)^{\lambda_j}$. G also satisfies $e^V_{i,c} \to \infty \Rightarrow G \to \infty \quad \forall c \in \mathcal{C}$. The first derivative with respect to a

$$\frac{\partial G}{\partial e^{V_{i,a}}} = \left(\sum_{c \in N_j} (e^{V_{i,c}})^{1/\lambda_j} \right)^{\lambda_j - 1} (e^{V_{i,a}})^{1/\lambda_j - 1} \geq 0 \text{ because } e^{V_{i,c}} \geq 0 \qquad \forall c \in \mathcal{C}$$

The cross derivation

$$G_{a,b} = \frac{\lambda_j - 1}{\lambda_j} \left(\sum_{c \in N_j} (e^{V_{i,c}})^{1/\lambda_j} \right)^{\lambda_j - 2} (e^{V_{i,a}} e^{V_{i,b}})^{\lambda_j - 1} \leq 0$$

$$\text{if } \lambda_j \in [0,1] \text{ and if } b \in N_j \ni a$$

and $G_{a,b} = 0$ if $b \notin N_j, a \in N_j$. Higher derivates will alternate in sign due to the negative exponent of the sum. (without proof). Since all conditions are satisfied we can simply insert G in (11) whereby we compute the choice probabilities of the nested logit model.

$$\begin{aligned} P_a &= \frac{e^{V_a} G_a}{G} \\ &= \frac{e^{V_a} (e^{V_a})^{1/\lambda_j - 1} \left(\sum_{c \in N_j} (e^{V_c})^{1/\lambda_j} \right)^{\lambda_j - 1}}{\sum_{N_j \in \mathcal{N}} \left(\sum_{c \in N_j} (e^{V_c})^{1/\lambda_j} \right)^{\lambda_j}} \\ &= \frac{e^{V_a/\lambda_j} \left(\sum_{c \in N_j} e^{V_c/\lambda_j} \right)^{\lambda_j - 1}}{\sum_{N_j \in \mathcal{N}} \left(\sum_{c \in N_j} e^{V_c/\lambda_j} \right)^{\lambda_j}} \end{aligned}$$

.5 Proof: RU2 UMNL is Consistent with Random Utility Theory

In this section all three components of the NLM probabilities (16) will be recalculated after altering the deterministic utility component in the following manner $\tilde{V}_c := V_c + \delta \quad \forall c \in \mathcal{C}$.

$$\begin{aligned}
\tilde{\mathbf{P}}_{a|N_j} &= \frac{\exp(\tilde{V}_a/\lambda_{N_j})}{\sum_{c\in N_j}\exp(\tilde{V}_c/\lambda_{N_j})} = \\
&= \frac{\exp((V_a+\delta)/\lambda_{N_j})}{\sum_{c\in N_j}\exp((V_c+\delta)/\lambda_{N_j})} = \\
&= \frac{\exp(V_a/\lambda_{N_j})\exp(\delta/\lambda_{N_j})}{\left(\sum_{c\in N_j}\exp(V_c/\lambda_{N_j})\right)\exp(\delta/\lambda_{N_j})} = \\
&= \frac{\exp(V_a/\lambda_{N_j})}{\sum_{c\in N_j}\exp(V_c/\lambda_{N_j})} = \\
&= \mathbf{P}_{a|N_j}
\end{aligned}$$

$$\begin{aligned}
\tilde{I}_{N_j} &= \ln\textstyle\sum_{c\in N_j}\exp(\tilde{Y}_c/\lambda_{N_j}) = \\
&= \ln\left(\left(\textstyle\sum_{c\in N_j}\exp(Y_c/\lambda_{N_j})\right)\exp(\delta/\lambda_{N_j})\right) = \\
&= \tfrac{\delta}{\lambda_{N_j}} + \ln\textstyle\sum_{c\in N_j}\exp(Y_c/\lambda_{N_j}) = \\
&= \tfrac{\delta}{\lambda_{N_j}} + I_{N_j}
\end{aligned}$$

$$\begin{aligned}
\tilde{\mathbf{P}}_{N_j} &= \frac{\exp(\tilde{W}_{N_j} + \lambda_{N_j}\tilde{I}_{N_j})}{\sum_{N_l\in\mathcal{N}}\exp(\tilde{W}_{N_l} + \lambda_{N_l}\tilde{I}_{N_l})} \\
&= \frac{\exp(W_{N_j} + \delta + \lambda_{N_j}(I_{N_j} + \delta/\lambda_{N_j}))}{\sum_{N_l\in\mathcal{N}}\exp(W_{N_l} + \delta + \lambda_{N_l}(I_{N_l} + \delta/\lambda_{N_l}))} \\
&= \frac{\exp(2\delta)\exp(W_{N_j} + \lambda_{N_j}\tilde{I}_{N_j})}{\exp(2\delta)\sum_{N_l\in\mathcal{N}}\exp(W_{N_l} + \lambda_{N_l}I_{N_l})} = \\
&= \mathbf{P}_{N_j}
\end{aligned}$$

.6 Derivation of the Nested Logit model elasticities

In this section we will derive the required elasticities for both the 2-level and the 3-level nested logit models. Own price elasticity ($\mathcal{E}_{a,a}$) can be interpreted as the change in probability an agent picks alternative a given a change in an attribute i.e. the price x_a of alternative a.

In the 2-level case the elasticities are calculated:

$$\mathcal{E}_{a,a} = \frac{\partial P_a}{\partial x_a}\frac{x_a}{P_a}$$

The probability of choice a is

$$\mathbf{P}_a = \mathbf{P}_{a|N_j}\mathbf{P}_{N_j}$$

Its derivative with respect to x_a is

$$\frac{\partial \mathbf{P}_a}{\partial x_a} = \frac{\partial P_{a|N_j}}{\partial x_a}\mathbf{P}_{N_j} + \frac{\partial P_{N_j}}{\partial x_a}\mathbf{P}_{a|N_j}$$

The derivatives of (13) and (14) are

$$\frac{\partial \mathbf{P}_{N_j}}{\partial x_a} = \mathbf{P}_{N_j}(1-\mathbf{P}_{N_j})\mathbf{P}_{a|N_j}\frac{\partial Y_a}{\partial x_a}$$

$$\frac{\partial \mathbf{P}_{a|N_j}}{\partial x_a} = \mathbf{P}_{a|N_j}(1-\mathbf{P}_{a|N_j})\frac{\partial Y_a}{\partial x_a}\frac{1}{\lambda_{N_j}}$$

Own price elasticity therefore equals

$$\mathcal{E}_{a,a} = \frac{\partial Y_a}{\partial x_a}x_a\left((1-\mathbf{P}_{N_j}) + (1-\mathbf{P}_{a|N_j})\frac{1}{\lambda_{N_j}}\right)$$

Cross-price elasticity equals

$$\mathcal{E}_{a,b} = \frac{\partial P_a}{\partial x_b}\frac{x_b}{P_a}$$

$$\frac{\partial \mathbf{P}_a}{\partial x_b} = \frac{\partial P_{a|N_j}}{\partial x_b}\mathbf{P}_{N_j} + \frac{\partial P_{N_j}}{\partial x_b}\mathbf{P}_{a|N_j}$$

$$\frac{\partial \mathbf{P}_{N_j}}{\partial x_b} = \mathbf{P}_{N_j}(1-\mathbf{P}_{N_j})\mathbf{P}_{b|N_j}\frac{\partial Y_a}{\partial x_b}$$

$$\frac{\partial \mathbf{P}_{a|N_j}}{\partial x_b} = \mathbf{P}_{a|N_j}\mathbf{P}_{b|N_j}\frac{\partial Y_b}{\partial x_b}\frac{1}{\lambda_{N_j}}$$

Crossprice elasticity therefore equals:

$$\mathcal{E}_{a,b} = \mathbf{P}_{b|N_j}x_b\frac{\partial Y_b}{\partial x_b}\left((1-\mathbf{P}_{N_j}) + \frac{1}{\lambda_{N_j}}\right)$$

For 3-level nested logit models the elasticities are calculated:

$$\mathcal{E}_{a,a} = \frac{\partial P_a}{\partial x_a}\frac{x_a}{P_a}$$

The probability of choice a is

$$\mathbf{P}_a = \mathbf{P}_{a|M_k|N_j}\mathbf{P}_{M_k|N_j}\mathbf{P}_{N_j}$$

Its derivative with respect to x_a is

$$\frac{\partial \mathbf{P}_a}{\partial x_a} = \frac{\partial \mathbf{P}_{a|M_k|N_j}}{\partial x_a}\mathbf{P}_{M_k|N_j}\mathbf{P}_{N_j} + \mathbf{P}_{a|M_k|N_j}\frac{\partial \mathbf{P}_{M_k|N_j}}{\partial x_a}\mathbf{P}_{N_j} + \mathbf{P}_{a|M_k|N_j}\mathbf{P}_{M_k|N_j}\frac{\partial \mathbf{P}_{N_j}}{\partial x_a}$$

The three derivatives with respect to x_a are

$$\frac{\partial \mathbf{P}_{a|M_k|N_j}}{\partial x_a} = \frac{\partial V_{a,M_k,N_j}}{\partial x_a}\frac{1}{\lambda_{M_k,N_j}}(\mathbf{P}_{a|M_k|N_j} - \mathbf{P}^2_{a|M_k|N_j})$$

$$\frac{\partial \mathbf{P}_{M_k|N_j}}{x_a} = \frac{\partial V_{a,M_k,N_j}}{\partial x_a}\frac{1}{\lambda_{N_j}}\mathbf{P}_{a|M_k|N_j}\mathbf{P}_{M_k|N_j}(1 - \mathbf{P}_{M_k|N_j})$$

$$\frac{\partial \mathbf{P}_{N_j}}{\partial x_a} = \frac{\partial V_{a,M_k,N_j}}{\partial x_a}\mathbf{P}_{a|M_k|N_j}\mathbf{P}_{M_k|N_j}\mathbf{P}_{N_j}(1 - \mathbf{P}_{N_j})$$

Where $\frac{\partial V_{a,M_k,N_j}}{\partial x_a}$ equals β_x if $V_{i,a}$ is linear in coefficients. Therefore the own price elasticity is equal to

$$\mathcal{E}_{a,a} = \tfrac{\partial V_{a,M_k,N_j}}{\partial x_a} x_a \Bigg((1 - \mathbf{P}_{N_j})\mathbf{P}_{M_k|N_j}\mathbf{P}_{a|M_k|N_j} + \\ + (1 - \mathbf{P}_{M_k|N_j})\tfrac{1}{\lambda_{N_j}}\mathbf{P}_{a|M_k|N_j} + (1 - \mathbf{P}_{a|M_k|N_j})\tfrac{1}{\lambda_{M_k,N_j}} \Bigg)$$

Cross price elasticity ($\mathcal{E}_{a,b}$) can be interpreted as the change in probability an agent picks alternative a given a change in an attribute i.e. the price x_b of alternative b.

$$\mathcal{E}_{a,b} = \frac{\partial \mathbf{P}_a}{\partial x_b}\frac{x_b}{\mathbf{P}_a}$$

The derivative of the probability $\mathbf{P}_a = \mathbf{P}_{a|M_k|N_j}\mathbf{P}_{M_k|N_j}\mathbf{P}_{N_j}$ of alternative a with respect to the price of alternative b is

$$\frac{\partial \mathbf{P}_a}{\partial x_b} = \frac{\partial \mathbf{P}_{a|M_k|N_j}}{\partial x_b}\mathbf{P}_{M_k|N_j}\mathbf{P}_{N_j} + \mathbf{P}_{a|M_k|N_j}\frac{\partial \mathbf{P}_{M_k|N_j}}{\partial x_b}\mathbf{P}_{N_j} + \mathbf{P}_{a|M_k|N_j}\mathbf{P}_{M_k|N_j}\frac{\partial \mathbf{P}_{N_j}}{\partial x_b}$$

The only derivative which is significantly different is

$$\frac{\partial \mathbf{P}_{a|M_k|N_j}}{\partial x_b} = \frac{\partial V_{b,M'_k,N'_j}}{\partial x_b}\mathbf{P}_{a|M_k|N_j}\mathbf{P}_{b|M_k|N_j}$$

The resulting cross price elasticity is equal to

$$\mathcal{E}_{a,b} = \tfrac{\partial V_{a,M_k,N_j}}{\partial x_b} x_b \Bigg((1 - \mathbf{P}_{N_j})\mathbf{P}_{M_k|N_j}\mathbf{P}_{a|M_k|N_j} + \\ + (1 - \mathbf{P}_{M_k|N_j})\tfrac{1}{\lambda_{N_j}}\mathbf{P}_{a|M_k|N_j} + \mathbf{P}_{a|M_k|N_j} \Bigg)$$

.7 Estimation results: Conditional Logit Model

```
clogit zugangsart preis2 volumen_gb downloadrate_gb alter_sb alter_sb
bildung_hs_sb bildung_hsom_sb d_wien_catv d_wien_dsl if dsl_vf_plz==1
& catv_vf_plz==1, group(serial)
```

Table 19: Condititial Logit Model

Variable	Coeff.	Std. Error	z	$p > \|z\|$
Price	.0379044	.0057103	6.64	0.000
Volume in GB	-.00748	.0050109	-1.49	0.136
Speed in MBit/sek.	.2989016	.0768916	3.89	0.000
Age*NoInt.	.0588009	.0028251	20.81	0.000
Age*Nb	.0197845	.003906	5.07	0.000
Edu_HS*Nb	-1.328497	.3168221	-4.19	0.000
Edu_HSwG*NB	-.4950805	.3224985	-1.54	0.125
Dummy_vienna_cable	.6083347	.1563382	3.89	0.000
Dummy_vienna_dsl	.6831593	.1530855	4.46	0.000
Log likelihood = -1141.3753, Pseudo R2 = 0.3083				

```
clogit zugangsart preis2 volumen_gb downloadrate_gb alter_sb alter_sb
bildung_hs_sb bildung_hsom_sb d_wien_catv d_wien_dsl if dsl_vf_plz==1
& catv_vf_plz==1 & choice!=''dsl'', group(serial)
est store partial
suest full partial
test [ full_zugangsart = partial_zugangsart ], common
```

Delivered a χ^2 of 46.78 which yields Prob > chi2 = 0.0000.

References

[1] McFadden, D., *Conditional logit analysis of qualitative choice behaviour.* In P.Zarembka(Ed.) Frontiers in econometrics (pp.105-142). New York: Academic Press (1974)

[2] McFadden, D., *Conditial logit analysis of qualitative choice behavior.* In P. Zarembka (ED.), Frontier in econometrics (pp. 105-142) New York: Academic Press.

[3] Pinelopi Koujianou Goldberg, *Product Differentiation and Oligopoly in International Markets: The Case of the U.S. Automobile Industry*

[4] Kenneth E. Train, *Discrete Choice Models with Simulation.* Cambridge University Press

[5] Manski 1997.

[6] Greene, W.H., *Econometric Analysis.* (5th edition). New Jersey: Prentice Hall

[7] Silberhorn, Boztug, Hildebrandt, *Estimation with the Nested Logit Model: Specification and Software Particularities*

[8] Cardona, M., Schwarz, A., Yurtoglu, B.B., and Zulehner, C., *Demand Estimation and Market Definition for Broadband Internet Services*

[9] Ida, T., & Kuroda, T., *Discrete choice analysis of demand for broadband in Japan*, Journal of Regulatory Economics, 29(1), 5-22

[10] Goel, R.K., Hsieh, E.T., Nelson, M.A., and Ram, R., *Demand Elasticities for internet services.* In Applied Economics, 2006, 38,975-980.

[11] RTR (Austrian Regulatory Authority), *TRT Telekom Monitor, 3.Quartal 2007.* Quartal 2007. http://www.rtr.at/de/komp/TKMonitor:Q32007/TM3-2007.pdf

[12] Reiss, P. C., & Wolak, F.A., *Structural econometric modeling: rational and examples from industrial organization.* Prepared for the *Handbook of Econometrics*, Volume 6.

A Deutsche Zusammenfassung

In dieser Arbeit diskutiere ich Möglichkeiten für eine Schätzung der Nachfrage für Internetzugang in Österreich. Es werden zwei unterschiedliche Modelle, (A) und (B), vorgestellt, die an den entgegengesetzten Enden des Spektrums möglicher Modelle stehen. Modell A baut auf einer Reihe stark vereinfachender Annahmen auf, worunter jedoch die Aussagekraft leidet. Dieses dient daher in erster Linie als Motiviation für das viel reichere und aussagekräftigere Modell B. Die Kernannahme, durch die sie sich unterscheiden, ist, ob die Güter des betrachteten Marktes homogen sind oder nicht. Diese Frage kann apriori nicht eindeutig geklärt werden, ist für die Modellierung des Marktes jedoch von entscheidenster Bedeutung. Die Annahme homogener Güter (A) führt auf folgendes Gleichungssystem, das sich mit Hilfe von Instrumental Variable Estimation schätzen lässt:

$$Q = f(P, ...)$$

$$P = g(Q, ..)$$

Modell B nimmt heterogene Güter an. Ein Markt mit n Gütern benötigt konsequenterweise eine Angebots- und eine Nachfragefunktion für jedes dieser n Güter.

$$\begin{aligned}
P_1 &= f(Q_1, ..., Q_n, ...) \\
&\vdots \\
P_n &= f(Q_1, ..., Q_n, ...) \\
Q_1 &= g(P_1, ..., P_n, ...) \\
&\vdots \\
Q_n &= g(P_1, ..., P_n, ...)
\end{aligned}$$

Ein Resultat der Annahme von differenzierten Gütern ist eine viel größere Anzahl an zu schätzenden Parametern. Ohne jegliche einschränkende Annahmen würde ein n Gütermarkt also n^2 zu schätzende Parameter bedeuten. Mit Preis- und Mengendaten allein ist dies schwer zu bewältigen. Daher wurden Modelle entwickelt, die Surveydaten von Haushalten, einer viel reicherer Quelle, verwenden. Das Aggregationsniveau und (als Konsequenz) die Datengrundlage stellt einen weiteren wesentlichen Unterschied der beiden Modelle dar. Während Modell A Preis- und Mengeninformation der OECD Länder heranzieht - d.h. ein Makromodell ist - verwendet Modell B

Haushaltsdaten. Dabei wird die Gesamtnachfrage als die Summe über die Konsumwahrscheinlichkeiten über alle Haushalte modelliert:

$$Q_c = \sum_{i \in \mathcal{I}} \mathbf{P}(\text{Haushalt } i \text{ kauft Produkt } c) = \sum_{i \in \mathcal{I}} \Psi(p_1, ..., p_n, \boldsymbol{x}_1, ..., \boldsymbol{x}_n, \boldsymbol{\omega}_i, \theta)$$

Die Konsumwahrscheinlichkeiten eines Individuums oder eines Haushaltes liefert ein sogenanntes Discrete Choice Modell. Dieses kann unter Verwendung der beobachtbaren Konsumentscheidungen und einer Reihe von sozio-ökonomischen Daten, wie Einkommen, Bildungsstand etc. die Wahrscheinlichkeiten, daß ein gewisser Konsument ein bestimmtes Gut wählt, berechnen. Hier offenbart sich eine weiter Stärke des zweiten Modells. Während Modell A von homogenen Konsumenten ausgeht, berücksichtigt Modell B individuelle Präferenzen, die mit Hilfe der sozio-ökonomischen Daten erklärt werden.

Diese Arbeit beinhaltet einerseits eine Gegenüberstellung der genannten Modelle und analysiert sie auf Stärken und Schwächen, andererseits wird in den folgenden Kapiteln die Nachfrage für Internetgüter in Österreich konkret mit Hilfe eines Nested Logit Modells geschätzt. Hierbei ergaben sich eine Vielzahl delikater Modellierungsfragen. Diese können durch die folgende Frage zusammengefasst werden: Wie können die idiosynkratischen Eigenschaften des betrachteten Marktes und die Verfügbarkeit von Daten im Rahmen eines Nested Logit Modells bestmöglich berücksichtigt werden?

B Curriculum Vitae

Fabian Schroeder
Curriculum Vitae

Contact

Email: fabian@schroeder.com
Phone: +43 699 122 655 02
Address: Rudolfsplatz 2/32, 1010 Vienna, Austria

Personal Details

Date of birth: 5th of November, 1983
Place of birth: Vienna, Austria
Citizenship: Austrian

Education

Since 10/2006	B.Sc. for Statistics and Mathematics for Economics, Vienna University of Technology (B.Sc. Expected Summer 2009)
10/2005–01/2006	Exchange Student at the Université Paris Dauphine
10/2003–03/2009	Mag. soc. oec. der Volkswirtschaftslehre Equivalent to: M.Sc. in Economics University of Vienna Specialisation: Econometrics
2000	Exchange Year at Box Hill School, Surrey U.K.
1993–2001	Döblinger Gymnasium BGXIX

Thesis

Demand Estimation with a Model of Discrete Choice; Supervisor: Prof. Dr. B. Yurtoglu.

Working Experience

Since Mai 2003	Personal and Research Assistant to Prof. John Naisbitt, Futurist, Author www.naisbitt.com
05/2002–05/2003	Holocaust Memorial Service at the Canadian Jewish Virtual Museum and Archives, Montreal, Canada www.cjvma.org
02/2002–04/2002	Holocaust Memorial Service at Yad Vashem, Jerusalem, Israel www.yadvashem.org

Skills

Languages	English, German(native), French(advanced)
Computer	MSOffice, HTML, Photoshop, LaTeX, MatLab, STATA
Hobbies	Saxophon, Travelling, Beachvolleyball

Vienna, 17. Februar 2009