The Inner-City Travel Demand in Poland. A Discrete Choice Analysis of the Preferences for Different Modes.

Pawel Strawinski

KU Leuven and Warsaw University

May 2003.

Master Thesis

Master of Science in Economics

Promotor: dr Ana Xavier

Faculteit Economische en

Toegepaste Economische Wentenschappen

Katholieke Univeriteit Leuven

Abstract

In this paper I study the Travel Demand for a private car and a public transport in urbanised areas of Poland. My case study concerns data from Gdynia. In frame of the Random Utility Model I derive the Travel Demand function as the density function of the Ordered Generalized Extreme Value Model. I found that the perception of travel times and quality indicators like distance to the nearest bus stop, travelling speed and frequency are most important factors related to the modal choice.

Acknowledgements.

I would like to thank professor Olgierd Wyszomirski from Gdansk University for providing me a data set, Mr. Krzysztof Borzyszkowski from ZKM Gdynia and Ewa Sokolowska for a help with the data, Mr. Vincent Kang Fu from UCLA for instruction how to properly interpret the results. Also I would like to thank Greetje Everaert, Jan de Loecker and other LICOS seminar participants for useful comments.

Contents

1 Introduction

In the last three decades, the growth of the cities and transport was faster than the growth of the economy in most industrialized countries. Rapid growth of transport and private car ownership has significantly changed the condition of transport within cities and urbanized areas. In the European Union countries, road traffic is growing at 3 percent per year, and is forecast to double by 2030. At the same time transport infrastructure has been also developed, but rather outside of cities. In the inner city areas usually there is no space for an increase of street capacity. This two factors leads to growing congestion and traffic jams. For policy-makers the following is the key question: how to divert traffic from private cars to public transport?

In Poland, and the others East-European post socialistic countries, the effect of growing car ownership has occurred just after fall of the system. Poland had about 130 car per 1000 inhabitants in 1989, 192 cars in 1992, and this figure increased to over 300 in 2002. At this time all existing and planned roads was not designed to carry such amount of traffic. This situation led to heavy traffic jams not only in the big cities, but also in the smaller ones, especially in places where there is a junction of two important roads. This situation can be worse when Poland joins the EU, and the growth rate of the economy will become even faster.

In my personal opinion studying the demand for a public good is an important economic issue, because it can help establish a proper policy. In this study I will compare the Multinomial Logit Model commonly used for transport research purposes with the not so widely known and used method of the Ordered Generalized Extreme Value (OGEV) model. This model extends classical methodology of the ordinal logit model by allowing for heterogeneity among consumers. To enlarge that view, I need to remark that to my knowledge studies in this subject with use of the presented methodology was done only in the United States, the United Kingdom and in the Netherlands. I now present a study based on Polish data.

The main purpose of my paper is to estimate of the travel demand function for urban areas of Poland. I will treat the demand as a probability function that describes the chance of choosing public transport as a mode of transportation against private car. I will address the question of which factors are important for travel decisions. My focus will be put on most sensitive ones. I will also try to figure out how big number of car owners for 1000 inhabitants in comparison to GDP level affect the demand for public transport. To investigate that I will check if people are sensitive to restrictions imposed on car users. It can possibly have important impact on the transport policy, because Poland GDP per capita is about 20 % of GDP of most developed EU countries while car ownership rate per 1000 inhabitants is almost at the same level! Next thing that I am going to investigate is the effect of congestion and growing cost of private transport (increased taxed and parking fees) on public transport demand. In the conclusion I will try to construct some policy advices, regarding the best way to prevent people from the extensive use of cars.

My starting points are concepts of stated and revealed preferences and the Random Utility Model (RUM). In the next part I will derive the exact form of the utility function. I begin with a general form of the model, in which I will treat utility as a function of personal socio-demographic characteristics, journey characteristics and other unobserved factors. I assume that the latter factors are random variables and therefore I can use a model from the RUM class. Then I am going to present the Independence of Irrelevant Alternatives (IIA) property, because the way of dealing with the problem of the independence of alternatives is a key point in modelling discrete choices. After that I will go straight to the most preferred model formulation and analyse the results.

The paper is organized as follows. In the next section I present a review of recent economic literature that present techniques used in transport demand analysis. In the following section I develop two concepts of preferences and their estimation. Then I illustrate the Random Utility Model and the IIA property. This section ends with formulation and presentation of the Multinomial Logit model (MNL), the Ordered Logit model (OL) and the Oredered General Extreme Value model (OGEV). In the 3rd part of this thesis I present the data set and the consecutive part is filled with results of estimations. The paper ends with conclusions and policy discussions.

2 Literature

Travel demand analysis has become important study field in transport economics during recent years. The first influential work that analysed the possibilities of measuring and forecasting transport behaviour was written by Domenchich and McFadden (1973). The first part was a summary of previous work by different authors. They also showed several techniques of dealing with transport data and discussed a problem of analysing discrete choices. This problem, with the exception of some simple models, remained unsolved until the early 1990's due to computational difficulties. Since computers became bigger, better and faster the scope of travel demand analysis could be switched from aggregate data to choices made by individual persons. Firstly, the Multinomial Logit Model (MNL) proposed by McFadden (1973) was used. However, this model requires an assumption of the Independence of Irrelevant Alternatives (IIA), which is very demanding in the context of transport and often does not hold in the real data. Choices faced by individuals can be very similar rather than completely different, and it is hard to claim that they are independent. Hence, to overcome that problem a class of more general models such as the General Extreme Value (GEV) models has been developed. Models from this class are more suitable for the analysis of data sets with categorical variables with correlated categories. There are many examples in recent literature of the application of the different models. The classic Multinomial Logit (MNL) and the Nested Logit (NL) was used by Ortuzar and Gonzales (2002) to study transport mode choice among the plane, the fast ship and the ferry on route between two islands Gran Canaria and Tenerife, looking at the impact of travel time and fare level. They showed that the nested logit performed better. The wider example of comparison between the MNL and the NL is the work of Rouwendal and Meijer (2001). They have studied the importance of various aspects of housing, employment and commuting. The more general study of the Nested Logit (NL) as a single approach was used to examine factors that have influenced household mobility among different states in the USA (Knapp et al. 2001). This model relaxes the IIA assumption between alternatives coming from different nests. This means that alternatives from the same nest can be correlated, but they are not correlated with the others. Another example is the work of Eliason and Mattsson (2001) in which they study the effect of the location and transport patterns to feasible road pricing policies. A different application of the NL was done by Parsons and Hauber (1998). They developed the Random Utility Model with NL choice structure to examine the sensitivity of the characteristics of fishing sites on actual choice. An extension of the MNL, called the OGEV was proposed by Small (1987). This model has the statistical properties of the MNL, but relaxes the IIA assumption. This model is designed for the analysis of ordered variables. It was applied by Williams et al (2002) to study response of teenagers to price and different policies of alcoholic drinks' distribution. The model that combines the OGEV with the NL is the DOGIT, proposed by Fry and Harris (2002) and used in the analysis of employment contracts.

Two extensions of the Nested Logit (NL) were proposed by Bhat (1997) to overcome the heterogeneity problem. A first approach is to allow for heterogeneity across agents in the variance-covariance matrix among the nested alternatives. The second approach proposes to allow for a non-diagonal variance-covariance matrix. Moreover, the Heteroscedastic Extreme Value (HEV) model was developed and used by Bhat (1995) to analyse inter-city transport choice mode. This model allows for more flexible cross-elasticity patterns among the alternatives than the Nested Logit, because it relaxes the assumption of the identical utility functions of consumers. Another extension of discrete analysis is to use simultaneously two kinds of data: stated preferences, which are crucial for obtaining information about attributes that are not present in the market, but whose forecasts are implausible, and revealed preferences which are limited to information available on the marketplace, but have better forecasting features. Fianlly the Mixed Logit Model (MXL) in a random coefficients setting was used by Bhat and Castelar (2002) to study congestion pricing, and the same model but in error components (random parameter) setting was used by Brownstone et al. (2000) to measure willingness to buy a car equipped with alternative fuel engine. Both applications gave more precise results than the NL model, however they need a specific way of designing and processing a survey. A more general comparison between the MNL, the NL and the MXL is presented by Rouwendal and

Meijer (2001).

The literature is completed with the analysis of the human behaviour with special attention paid to transport research. The first aspect of analysing survey data is the way of dealing with the problem of stated and revealed preferences (see below). In my work I will keep in line with Brownstone et al (2000). This means that I will use both theoretical conceptions of stated and revealed preferences simultaneously. Secondly I will use theoretical concepts of Train (1998) and Parsons and Hauber (1998) to design a choice set and interpret the results. This means that I combine two types of the data. Then I will plug them into a not very complicated model to examine the outcomes. The second problem that I will address is an extensive usage of car is Polish society. In the analysis of that phenomenon I will follow the behavioural study of Vasconcellos (1997). During the analysis of the results I will also refer to the similar study done for the Lodz by a team of professor Suchorzewski (Suchorzewski 1996). I will use the result for comparison and generalisation reasons. I will also refer to previous findings. Ortuzar and Gonzalez (2002) shown that travellers who use faster mode attach higher valuation to time in comparison to person using slower one. However, it is very hard to judge how big is impact of time due to heterogeneity among travellers and travels. (Hensher 2002).

3 The Model

In this section of my work I discuss the problems that arise during the estimation of preferences. I then switch to the Random Utility Model and the Independence of Irrelevant Alternatives property. At the end of this chapter I will develop the Multinomial Logit Model, the Ordered Logit Model and the

General Ordered Logit model, which I will use later in estimation. In what follows I will mention Logit model, I have in mind its general multinomial form. I will treat the usually used form of binomial logit as a special case, in which dependent variable can have only two different values.

3.1 Stated and Revealed Preferences

Estimating the demand for products or services requires the information about consumers' preferences. In the real market every consumer faces such many different choices and usually chooses among the alternatives from the different set that is impossible to create one model, which will include all possibilities of choice. On one hand, even if there exists a data set of Revealed Preferences (RP) meaning that we have the information about consumer behaviour in real life and we know that he was choosing among all alternatives that we consider in the model, many difficulties arise during the development of the model. There is frequently high collinearity and limited variation among the attributes in the real market. Products are very similar rather than completely different. On the other hand, we can face the opposite problem. The real data could be spread so widely and describe so many choice situations that it would be very hard to find patterns of consumers behaviour among it. Without a reliable and robust method of aggregating the data it would not be possible to construct the model. Finally there may exist a self-selection problem. The sample of users is not random because people have previously decided to buy a product or take a part in the event. To overcome these problems researchers have designed the Stated Preferences (SP) experiments to measure consumer's preferences over hypothetical alternatives. A person is asked how would he behave in a hypothetical situation, which act as a model of reality. However, it is not clear if data collected in this way (SP) can be truthfully treated. On the one hand, this methods were object of criticism by some economists and other researchers because of a belief that consumers react differently to artificial experiments than they would facing the same alternatives on the real market, but on the other hand recently a Nobel prize has been awarded to John Hay for the development of the experimental economy.

Modelling of a travel behaviour is not as easy at it seems. The first problem to overcome is the measurment of the preferences. There is no unique measure of the preferences, so we have to do it indirectly, by collecting the information from several similar respondents and choice situations. Secondly, we must collect the data in such a way that we have to be sure that they are reliable. There exist at least two different ways of collecting the information that describes peoples' activity. One is more technically demanding and it leads to information about RP, the other method is much easier to apply, but it collects information about SP only. However, it is possible to reveal true preferences from stated preference information by confronting given answers with additional information about individuals' job, income and other economic characteristics.

The first straightforward way is to collect data from activities at the time of activity. For example ask passengers to fill a questionnaire at the train station or at the airport. This data is called in the literature a "revealed preferences" data. By taking part in a particular activity an individual showed his preferences, showed that this activity gave him more utility than the available alternative. Moreover, we are sure that the questioned person took part in the activity. Despite that, this method is often hard to apply. For example, it is very difficult to measure preferences of commuters within a city. It is very hard to imagine that a passenger would fill a survey questionnaire containing thirty or more questions at the time of travelling from one stop to another in the overloaded bus. If we cannot collect data at the time of activity we can always ask after, but there exists a methodological problem. People might be untruthful, especially when we ask about private life and other sensitive aspects of their lives. If the thing that they did is not politically correct, they simply may give answers that are not true. This leads to the situation in which the data that come out of the survey are stated preferences, which are not necessarily the true preferences. It is more obvious when for example by a questionnaire survey we want to find out if people will use a new bus line. In cases like this in the example obviously we are not able to collect the true data. All we can do is to estimate model on existing data, and correct result for possible errors. One potential solution to these problems is to develop and estimate joint models to exploit the advantages of each type of data and to neglect the weaknesses.

3.2 Random Utility Model

The underlying feature of discrete choice models, which distinguishes them from the other models that are built from discrete data, is the assumption about the probability function. It is assumed that it come from a known parametric family and is independent from all choice-deciding variables. A discrete choice models specifies probabilities $P(j \mid X, \theta)$ for each set of alternatives j among which the individual can choose. The exogenous variables x describes observed attributes of the agent and observed attributes of the alternatives available to him. They are treated as decision variables that affect the choice. The unobserved attributes of the alternatives and the characteristics of the decision-maker are included in the error component. The vector of the parameters θ is estimated from the available information about choices of individuals. The method of estimation depends on the functional form of the probabilistic model or on the distribution of exogenous variables X (Cosselet,1974).

A very often and widely used basis for application of discrete choice models is the Random Utility Model. It is derived as follows. Agent i faces a choice among J available alternatives. Every alternative has its own utility level. The total utility that can be reached by individual i choosing an alternative j is described by:

$$U_{i,j} = \beta' x_{i,j} + \epsilon_{i,j} \tag{1}$$

The utility from equation (1) can be easily separated into two parts. One that depends on visible parameters that are linked with vector $x_{i,j}$, and the other unknown part $\epsilon_{i,j}$. The revealed part of the utility is known to everybody. Usually it contains social and demographic characteristic of the person and known factors linked with particular choice situation. The unknown part of utility is known only to decision makers, and the scientist while constructing a model has no information about it and treats this part as a random variable from some, known distribution. It is also assumed that this random variable is independent across agents. We assume that the decision maker acts rationally and he makes a decision by maximizing his utility. Upon this assumption we are able to construct the model, in which the probability of a choice j is equal to:

$$Pr(y_i = j) = Pr(U_j > U_k \quad \forall j \neq k) =$$

$$= Pr(\beta' x_j + \epsilon_j > \beta' x_k + \epsilon_k) = Pr(\epsilon_j - \epsilon_k > \beta' x_k - \beta' x_j) \tag{2}$$

This means that an agent chooses the alternative j over k if the utility connected with choice j is greater than one of choice k. The analytical form of the model depends on the distribution of the error terms. In practice only the normal distribution, the log-normal distribution and its mixtures are taken into consideration. The other distributions are also possible to implement, but they lead to very complicated computations. The choice of the normal distribution leads to the probit model, which is complicated and computationally demanding. The Logit model is much easier and more tractable, and this is a reason why is widely used. McFadden (1973) proposes the following distribution of error terms as basis for the logit model

$$F(\epsilon) = exp(-exp(-\epsilon)) \tag{3}$$

This distribution has several names but the most common in the literature are the Type I Extreme Value distribution, or the Gumbel distribution, and the log-Weibull distribution. This error terms structure leads to an easy-to-apply utility model. It has a simple closed analytical form although, in some cases, its assumptions are demanding.

$$Pr(y_j = k|x) = \frac{e^{\beta x_k}}{\sum_{l=2}^{J} e^{\beta x_l}} \tag{4}$$

where β_0 or other β coefficient is assumed to be 0 for the identification purpose.

The standard logit model assumes that consumers are homogenous. In the model each individual has the same utility function. The model does not allow for any variation among individuals. It also exhibits the so called "independence from the irrelevant alternatives" (IIA) property. This property states that the odds of choosing alternative j over alternative $k \quad (k \neq j)$, $P_{i,j}/P_{i,k}$, are independent of all other alternatives, and of the number of available alternatives in the choice set. For a logit model, the odds ratio is given by

$$\frac{P_{i,j}}{P_{i,k}} = \frac{\exp(U_{i,j})}{\exp(U_{i,k})} = \exp[X_i'(\beta_j - \beta_k)] \tag{5}$$

The odds are determined without reference to the other outcomes that might be available. While this may appear to be an obscure mathematical detail, it has an important practical implication often can be illustrated with the famous red/blue bus example that has been attributed to McFadden, and is cited in almost every article or book about discrete choice, e.g. McFadden (1973), Long (1997), Train (2000).

3.3 The Independence of Irrelevant Alternatives (IIA)

The following example illustrates the all known problem of the IIA property. A person has two choices for commuting to work: a private car that is chosen with $Pr(car) = \frac{1}{2}$ and *red* bus with $Pr(\text{red bus}) = \frac{1}{2}$. The implied odds of taking a car versus the red bus are equal to $\frac{1}{2}/\frac{1}{2} = 1$. Now, suppose that a new bus company has started to operate, that is identical to the current service except that buses are blue. The IIA property requires that the new probabilities are: $Pr(\text{car}) = \frac{1}{3}$, $Pr(\text{red bus}) = \frac{1}{3}$, $Pr(\text{blue bus}) = \frac{1}{3}$. This is necessary so that the odds of a car versus a red bus remain at $\frac{1}{3}/\frac{1}{3} = 1$. However, if the only thing to distinguish the new bus service from the old is the colour of the bus, we would not expect car travellers to start taking the bus. Instead, the share of red bus riders would be split, resulting in: $Pr(\text{car}) = \frac{1}{2}$, $Pr(\text{red bus}) = \frac{1}{4}$, $Pr(\text{blue bus}) = \frac{1}{4}$. The new, implied odds of car versus red bus are $\frac{1}{2}/\frac{1}{4} = 2$, which violates the IIA assumption. The IIA assumption requires that if a new alternative becomes available, then all probabilities of the prior choices must adjust in precisely the amount necessary to retain the original odds among all pairs of outcomes. McFadden (1973) suggested that the IIA assumption implies that the logit model should only be used in cases where the outcome categories "can plausibly be assumed to be distinct and weighted independently in the eyes of each decision maker".

The IIA assumption implies the proportional substitution across alternatives. This property can be seen either as a restriction imposed by the model or as the natural outcome of a well-specified model that capture all sources of correlation over alternatives into representative utility such that only white noise remains. Often though, the researcher is unable to capture all sources of correlation explicitly, such that the unobserved factors of utility are correlated and the IIA does not hold. In these cases, a more general model than the standard binomial or the multinomial logit is needed. These are definde in the section that follows.

3.4 Analytical form of the models

The logistic probability model was first introduced in the context of binary choice as above where the logistic distribution is used. Its generalization to more than two alternatives is referred to as the Multinomial Logit Model (MNL). The Multinomial Logit Model is derived from the assumption that the error terms of the utility function are independent and identically Type I Extreme Value distributed, that is $\epsilon_{i,j}$ for all i,j is distributed with a distribuant:

$$F(\epsilon) = e^{-e^{-\mu(\epsilon-\eta)}}, \quad \mu > 0 \tag{6}$$

and density function:

$$f(\epsilon) = \mu e^{-\mu(\epsilon-\eta)} e^{-e^{-\mu(\epsilon-\eta)}} \tag{7}$$

where η is a location parameter and μ is a strictly positive scale parameter. This distribution is not empirically distinguishable from the normal distribution (Train 2002). Its usage is equivalent to impose the independence and the normality on error terms, so that this distribution does not produce any limit of use. In fact, the normality assumption is commonly used in a wide range of econometric models from the OLS to the most complicated ones like the VARs, autoregressive models or the multinomial probits. The more restrictive assumption is that unobserved factors are uncorrelated. The mean of this distribution is greater than zero, but this is not a problem since the model operates on differences in utilities.

The difference of two variables with a Type I Extreme Value distribution has a logistic distribution with the mean 0 and the variance $\Pi/6$(Train 2002). The probability that a given individual i chooses an alternative j within the choice set C_i is given by:

$$P(j|C_i) = \frac{e^{\mu X_{j,i}}}{\sum_{k \in C} e^{\mu X_{k,i}}} \tag{8}$$

An important property of this model is the IIA: there is no correlation between error terms and choices made by the decision maker should be easily distinguishable. However, in some cases this property is very restrictive as previously mentioned.

We can interpret the constant for each category in the MNL model as an average utility linked with particular choice. Nevertheless, the constant as the estimator of average utility is biased, because it contains beside taste variations usual error terms. The proper way of dealing with the problem of individual taste variations is to apply model from the GEV family e.g. the Random Parameter Logit model (RPL). I consider this model as potential extension of one given below.

The first way of resolving the problem of the IIA assumption is to use a slight modification of the MNL model, that is a model with ordered alternatives. In the Ordered Logit model (OL) only attributes of the alternatives that lie close to each other in the natural ordering are correlated. The derivation of the model is analogous to the MNL, but during the result interpretation we have to have in mind the specified ordering. This model is applicable when dependent variable has a natural ordering of values (e.g. education level, choice of number of cars in the household). The exact values of dependent variable are not important, what is critical is the order of categories. It is widely used in the analysis of psychological surveys when the respondents choose an answer from a given set, which is coded in the Likert's scale. The Ordered Logit model needs the same assumption of the distribution of the error terms as the MNL, but it can be used also in situations when there exists small, but statistically insignificant correlation between alternatives. It shares other properties with the MNL.

The second possible way of relaxing the IIA assumption is to create a very similar model to the Multinomial Logit Model but one that can overcome the IIA problem. The Nested Logit model, firstly proposed by Ben-Akiva(1973) , is an extension of the MNL designed to capture some correlation among alternatives. It is based on portioning the choice set C_i into M non-overlapping subsets called nests. McFadden showed that the NL model is consistent with utility maximization (Train 2002). This model can be treated as a product of two logit probabilities. It has a distribution, which is a generalization of one used in construction of the logit model. It is also the simplest model from the General Extreme Value family. Despite that, it contains shortcomings, because some correlations have to be equal to zero by the assumption.

The Ordered Generalized Extreme Value (OGEV) model proposed by Small (1987), also referred in the literature as the General Ordered Logit (GOL) model, is a different representant of the GEV family. It is a more general model that combines proprieties the NL and the OL. Like these models, the OGEV is consistent with the Random Utility Model. As the OL, this model implies that there is a natural ordering of dependent variable. It can, but not need to, impose zero correlations between random utility components. It allows for overlapping subsets of choice set, however when correlations among outcomes are small, it produces worse results than previously described models. In this feature ie. overlapping sets it may be preferred to the NL. The probability function of the model could have several different patterns depending on the number of correlated alternatives. The OGEV probabilities expand on the MNL ones of equation (8) such that IIA is no longer embodied. However, the underlying motivation for the OGEV model is to provide a suitable model for outcomes that are ordered in some sense, whilst still providing the flexibility of the MNL model. Unlike the MNL probabilities, the OGEV ones embody a correlation between outcomes in close proximity. Such a correlation appears likely for ordered data in many instances, especially where the observed outcomes are realizations of an underlying latent scale. For example, given a five-point response scale (j = 1,... , 5) of satisfaction, individuals may choose "neutral" (j = 3), but be heavily influenced by the neighbouring choices of "moderately satisfied" (j = 4) and "moderately dissatisfied" (j = 2). Although it is possible to allow the window of correlation to be arbitrarily large, this increases the number of parameters to be estimated and makes estimation cumbersome (Small 1987). Therefore I restrict attention to the standard OGEV model. The model is obtained by assuming that the vector of unobserved factors of utility $\epsilon_{i,j}$ has a cumulative distribution:

$$\ln(f(\epsilon)) = \exp(-\sum_{i=1}^{K}(\sum_{j=1}^{R} e^{-\epsilon_{i,j}/\lambda_r})^{\lambda_r}) \tag{9}$$

In Smalls notation we have R = 2 and $\lambda_r = r \quad \forall r$. However, for some applications (such as in the expectations example), it is possible that the more flexible correlation structure implied by multiple λ might be more appropriate. The standard OGEV model implies a non-zero correlation between outcomes that are near neighbours. Analogously to a moving average process, this correlation decreases the further away two outcomes j and k are and is zero when $|j-k|>2$. Although they cannot be written explicitly in closed form (Small 1987), these correlations are inversely related to the parameter λ.

This distribution is a type of General Extreme Value. It is a generalization of the distribution that is used to construct the logit model. For logit, each $\epsilon_{i,j}$ is independent with univariate Type I Extreme Value distribution. For this GEV the marginal distribution of each $\epsilon_{i,j}$ is univariate Type I Extreme Value. However, $\epsilon_{i,j}$ are correlated. The parameter λ_k represents a measure of the degree of independence in unobserved utility among different alternatives j. For $\lambda_r = 1 \quad \forall j$, representing independence among all alternatives, the GEV distribution becomes the product of the independent extreme value terms. In this case, the OGEV reduces to standard logit model. The fact of weakening restrictions imposed on correlation matrix has an impact on estimation. Likelihood function in these models is not necessarily globally concave, as it is in the MNL or the OL case. This feature makes estimation by iteration more difficult and in the extreme cases result might be unreliable. I will not derive exact form of distribution, because it is complicated due to possible different correlation patterns. The exact functional form can be found in Small (1987).

4 Data

The operational structure of public transport in Poland differs from the common practice of the EU membership countries. Before 1989 all firms were state owned. The state was responsible for management, providing services, planning, etc. After the fall of the system tremendous and fast changes were made. At the moment in a major part of Polish cities public transport is already deregulated in such a way that the management firm is owned and governed by the local authorities and transport services are provided by firms that compete with each other. The management firm organizes a public auction and chooses one or several service providers, which are either publicly owned or private firms. Despite that, the main difference is that the cost of public transport in Poland is relatively high in comparison to the current European Union members. The simple reason that stands behind this is the low level of subsidisation given to the public transport by local governments. This leads to a situation in which income from tickets cover approximately 2/3 of the total cost. Local governments participation in costs does not exceed 25 %.

The data set that I am going to explore in my paper was collected during a joint study by Zarzad Komunikacji Miejskiej w Gdynii, which is a public transport management company in the city of Gdynia, and the Uniwersytet Gdanski (Gdansk University). The survey was done in the year 2002. The main purposes of this study were to examine mobility patterns within the city and it suburbs and gather information about citizen's attitudes in order to plan the extension of the public transport network. Because this data set, unfortunately, does not consist of all the information that is required for proper economic analysis of transport I have completed the data with information taken from other sources like the Polish Statistic Bureau and other institutions (such as the Gdynia city council).

Gdynia is an average sized Polish town. It has about 250 thousands inhabitants. Lying in the northern part of Poland at the seaside, it creates with Gdansk, Sopot and several other smaller cities and villages a great Gdansk-Gdynia conurbation in which lives nearly one million people. These cities make up the third largest city area in Poland after Warszawa (Warsaw) and the Gorny Slask (Upper-Silesia) conurbation. This conurbation is located along the Baltic sea cost, so it is quite long, approximately 40 kilometres, but not wide. The city is bounded from the northeast by the sea, and from the southwest by old ring road. Nowadays, however many new villages have been already established on the other side of this road.

There exist five different modes of public transport in the Gdansk-Gdynia area. The most important is the Szybka Kolej Miejska (Rapid Railway System), which is integrated with the other means of transport. The others are the tramway's in the city of Gdansk, the trolleybuses in the city of Gdynia and Rumia (northeast suburb) and the buses which provide the service in the rest of the area. The last kind of public transport, the taxi, which is treated by me as a private car, because characteristics of both modes do not differ in terms of speed and congestion [1]. The only exception is that the taxi driver does not go to the same destination, he just works inside his car.

The data for the survey is a randomly drawn representative sample for people aged 16 to 75 living in the northern area of Gdansk-Gdynia conglomeration. Younger persons are not included because they often do not make independent travel decisions themselves. The older people are excluded from a sample due to fact that mobility among this group is relatively low. Moreover, in Poland there exists a law, which states that every person who is aged

[1] Note that in Poland there are a few bus lines.

above 75 can use a public transport system completely free of charge. Hence, the analysis of the behaviour of these people is not very important from the point of view of the management company.

The drawn sample consists of 1895 [2] persons. Three of them do not respond to the survey. Furthermore, there are another 13 observations, which I excluded from the sample. The reasons for reducing of the sample were either errors in coding or missing answers to key questions from the point of view of the analysis. This makes the sample of 1879 observations, which I use throughout the construction of the model and estimation.

There exist several reasons that can explain why people travel around the city. Three of them are very common across the world and produce a vast share of all trips. The first purpose of travel is to get to work place. Almost all employed persons have to use a public or private transport to get to their work. Moreover, many people use the car as a tool or service during work. The other means of transport for getting to work such a bike or walk are used so rarely that can be neglected without loosing any important information [3]. The second major travel purpose is going to school. The younger persons do not work, but they also have to travel to get to educational institutions. The third factor that makes also lot of travel is shopping. Those three main factors produce very directed passenger streams that can be easily measured. The other purposes like recreation, entertainment, and visit to friends or relatives do not produce a big demand, but these travels are spread all over the city area rather than concentrated in usual paths.

My dependent variable indicates how often a person chooses a private car

[2] 1 % of population living in area covered by the survey.

[3] Taxi was used by 4 persons, bike by 7. No motorbikes travel reported. However there are additional 20 cases of travelling by a taxi, but all of them are related to visit to friends or relatives

or a public transport as a mean of transport. It is an answer to the survey question, coded by Likert's scale where 1 indicates that person always use a public transport, 2 means that generally uses public transport, but also from time to time a private car is used, 3 shows that the respondent is indifferent between car and public transport, 4 indicates that a person often uses a car, but sometimes chooses the public transport and 5 stands for using a car only. So it represents rather a ranking of available alternatives than a simple choice from equally ranked ones. There are also obviously independent variables. I can divide them into three separate groups. The first contains socio-demographic characteristic, the second one contains collected information that describes different transport modes and travel conditions and the last one is made from the answers about the preferences.

Note that, during the analysis of socio-demographic characteristic, when a particular attribute was found to be common only for a small group of people, that situation can lead to a non-significant result, that is, the impact on the dependent variable is not significant due to bad statistical properties rather than true preferences. I recoded some variables into wider groups. During the process of group construction I was trying to create sufficiently large groups to be valid in estimation and at the same time I was dividing persons according to their potential travel behaviour, that is, aggregating similar sub-groups.

Usually, like in every cross-sectional survey, the data set consist information about age, gender, and the main activity of the respondent. Given that information, I have constructed a dummy for gender, which take value of 1 for a male. I did not have exact information about age. Only reports about age groups are available, so I use them in the model. People are divided into 7 age groups from 16-20 to 70-75. The range of each group in between is

10 years. The answer to the question about the main economic activity was coded in seven categories, which I have recoded into four dummies. The first indicates that person work full time or part time, or is early retired and work part time. The second one is an indicator of studying, the third of retirement due to age or due to health problems and the fourth one is for non-working person.

Along personal profiles, I used the information about the travel activity. The survey provides a lot of travel-based activity. I considered all of them as potential explanatory variables in the model, but I decided to leave out some characteristics that have weak connection with modal choice. I have chosen only the most relevant to the question being analysed in the paper. I pick a dummy for persons having a driving licence and another one for the main car users. I also have information about number of cars in the household. I have incorporated into the model a dummy indicating that person is travelling to work or school. Besides that I created a several dummies for different purposes of travel - work, school, shopping, entertainment, meeting with friends or relatives, work related activity and visit to a public institution. Furthermore, I consider traffic conditions in my model by including dummy that indicates how often there are traffic jams.

In addition I have used information about time of travel. The travel time variable was present only for 951 respondents, only for those travelling to work or school. I assumed that travel to work is similar to travel to institution, and travel to school is similar to travel to work. To create my time variables used in the regression, firstly I dropped the outliers. As odd observations I treated the ones indicating that a person have travelled for more than two hours for schooling or work purpose and the destination was outside Gdansk-Gdynia conurbation. As a consequence I have removed seven observations from the sample to compute the group averages, but I still used them in the model, because even such long, but frequent trips affect travel behaviour within the city. This reports was done by persons that are studying or working in different towns (Torun, Poznan, Warsaw). All these persons are young and they travel to school spending between 2,5 and 4 hours.

To create time variables - proxies for travel times by car and by the public transport - I used the given answers by 951 respondents. For people that have answered I am using the original data. For those not responding I assumed that the people from the same age group that live in the same area should have fairly similar travel habits. Moreover, I know that only people who work or study have been asked a question about travel time. Knowing this I can assume that for the rest of the population the main purpose for travel is shopping or visit to a public institution, like administration or hospital. The latter is especially the case for older persons. Given that some shops are fairly close to the residents and travel causing institutions are rather in the city centre these can be classified in two kinds of travel. One, for shopping purpose, is similar to going to the school and the second to institutions is close to going to work. Therefore if I compute the average of the two variables: the duration of travel to work and to school for every previously defined subgroup I can use it as a proxy of travel time.

I constructed my time variables in this way to overcome a self-selection problem. I had an alternative of splitting the sample and construct a model upon observations with full information only. Nevertheless, I believe that the bias caused by my assumptions has a lower impact on the result than the problem arising during estimation on sub-samples only [4].

[4] I estimated the regression for sub-sample of 951 observations. The results from this model are very similar. This indicates that my results seems to be robust.

In the model I try to catch an impact of time travel by car and by public transport on mode choice. If the results are not clear I could replace these two variables by difference of time spend in both modes, as, it may be that only a difference between the two modes matter.

Table 1. Travel times by groups.

Group	Obs	Mean(car)	SD(car)	Mean(pub)	SD(pub)
1	541	18.23	10.92	35.04	17.89
2	569	17.24	8.55	33.19	15.37
3	231	18.85	15.66	35.71	22.46
4	284	19.55	14.16	37.53	22.61
5	254	19.34	12.17	41.19	25.40
Average	1879	18.36	11.72	35.77	19.90

In the survey, several questions about suggestions and the preferences have been asked. From my point of view only three of them are relevant and can bring more explanatory power into my model. The respondents were asked to choose from a closed list with one open point up to three factors that make them use a private car in city travels, that make them use a public transport in those travels and what characteristics of public transport are important for them. From given answers I have constructed 29 dummies, which describe preferences. Four of them concern time of the journey, another four cost of travelling, five the level of congestion. There are also indicators of travel quality, that is, travel related infrastructure quality. Although, later in the results section I present only the factors that are relevant for choice between modes and are statistically significant. Unfortunately, due to a large variation in opinions many of these factors may be unimportant for modal choice decision.

Table 2. Descriptive statistics

Characteristic	No of obs	Percent	Characteristic	No of obs	Percent
Gender			Activity		
Female	993	52.85%	Work	948	50.45%
Male	886	47.15%	Study	349	18.57 %
Age groups			Retired	440	23.42%
Age 16-20	181	9.63%	Non-working	238	12.67 %
Age 21-30	371	19.74%	Car related		
Age 31-40	275	14.64%	Access to car	1088	57.90%
Age 41-50	385	20.49%	Licence	725	38.57%
Age 51-60	341	18.15%			
Age 61-70	232	12.35%			
Age 71-75	94	5.00%			

The sample is designed in such a way to be representative for the population of Gdynia and the suburbs. As it is common for Poland, there are more females than males. Age structure has two peaks. First is effect of post war boom of the population growth in the 40's and 50's and second one is in late seventies. Work and study rate are high in comparison to the rest of the country, but it is usual pattern for big city.

5 Results

The model that I will explore describes a function of the aggregate demand for travel. It is built from the information about individual choices. This approach allows me to overcome a problem of the heterogeneity and the taste variations by treating the whole population as a single decision-maker, who makes a certain decision with some probability. However, this model allows for a limited degree of heterogeneity among travellers. Then I would be able to calculate probabilities of each choice conditional on the decision-maker characteristics. One of the most important short-term travel decisions is a choice of a specific mode. It is important to note that short-run decisions are conditional on long-term and mobility ones such as choice of becoming a car owner and choice of the residence and work location. In the model I will treat these factors as given to the decision-maker and I will not consider them in a process of making decision, although they may affect the choice.

In the model I also do not consider two other factors that potentially may have an influence on the results. The first is the spatial location of household. I do not differentiate between people living close to the city centre and those who live in the suburbs. This can be justified in the case of Gdynia since Gdynia shipyard and affiliate firms, which were located in one particular area, and created lot of traffic, are no longer a major employee. Working places are rather uniformly spread over the city, but with higher density in the city centre. The second element that I neglect is a separation between people living inside the city boundaries and outside the city. In a recent study (van Ommoren 1998) it was shown that the critical time for commuting for work or school (except university) is between 30 and 45 minutes in the Netherlands. In Poland I expect this number to be higher, particularly for work based travels due to the high unemployment rate and therefore the low possibility of changing a job.

The set of data is a cross-sectional survey. It contains, in general, information about stated preferences. There is some additional information about revealed preferences, but available only for some observations and therefore I cannot plug them into a model. So I decided to use stated preferences data as common in the area..

The dependent variable of my model is categorical. There is a discussion in the literature (Train 2002, Greene 2003), concerning the way in which the most proper discrete choice model should be estimated. As a starting point I have chosen the Multinomial Logit model. Then I compare this model to the Ordered Logit Model and then extend the latter to the OGEV form. I was also considering the Nested Logit model as a potentially more general model, but first of all it was hard to apply due to the problem of constructing the nests. These are not obvious in the context of this paper. Despite that, I have tested several nested logit models with different nest structures and they fit to the data poorly.

The MNL model is a very general form of the discrete choice models. It assumes that every possible outcome of the dependent variable is independent from the others. The estimated model looks very nice. It has ρ statistic above 50% [5] so the model has a good explanatory power. All coefficients have expected signs or they are not significant. Nevertheless, there is a problem of irrelevant alternatives. The IIA test fails: although, the Hausmann test suggests that the IIA property holds, the asymptotic assumptions of the IIA test are not met. Moreover, the results of the Small-Hsiao test are negative indicating that the IIA does not hold. According to the recent literature when

[5] rho statistic is equal to 1 minus the ratio of log likelihood of the full model divided by the log likelihood of the model with intercept only.

asymptotic assumptions are not met, all the IIA test in general less powerful (Fry&Harris 1996). This suggests that the IIA does not hold. The first way of resolving the problem is to check a possibility of combining categories of the dependent variable. The Wald test shows that any pair of categories cannot be collapsed to one category (for the result see Appendix A1).

In this situation I have decided to estimate the OL model, but this model has worse statistical properties. First of all, the standard comparison LM-test rejects the OL in favour of the MNL. Secondly the AIC and the BIC criteria also suggest that there is something wrong with this model. So, knowing that the MNL is not correct, I decided to combine both models and estimate the OGEV model. They are compared in Table 3.

Table 3. The model comparison.

Statistic Model	MNL	OL	OGEV
LogLik	-1425.484	-1777.990	-1402.726
LogLik intercept	-2878.041	-2878.041	-2878.041
Pseudo-R^2	0.5047	0.3822	0.5126
McFadden R^2	0.439	0.365	0.447
AIC*n	3226.968	3655.980	3181.425
BIC	-1518.324	-1853.404	-1563.840

The OGEV model has better properties than the OL and the MNL. It has the highest log-likelihood and pseudo-R^2, and the lowest Akaike criterion. Only the Bayesian Information criterion suggests that the OL is better than the OGEV, which may be caused by the non-concavity of the log-likelihood function, which was reported during the OGEV estimation. In this context the BIC criterion may be not advisable.

To present all full estimated models a lot of space must be provided. Therefore, full results are presented in the appendix A.2 in terms of coefficients and in the appendix A.3 in terms of odds-ratios for the OGEV estimation, the most preferred model.

First of all it is important to note that most variables are significant for choice between "public transport always or often" vs "car always or very often or indifferent" and "public transport always or often or indifferent" vs "car very often or always". There is nothing wrong with that results, because this is strongest choice situation. The coefficients for the other choices vary greatly indicating that most of people are in fact using both modes of transport. The main role of the explanatory variables is to distinguish between different modes, so e.g. if variable is significant for choices connected with the car, and not significant for those related to public transport, and statistically not significant for all choices, but for the needs of my model this it is sufficient when is significant just for one particular choice situation. Therefore for checking the relevance of variable I perform F-test.

The travel times across all the groups are fairly similar (see Table 1), and have one general feature: the car is on average two times faster than a public transport. Despite that, standard deviations of car travels are higher, which inevitably means that cars are used for wider range of trips than the public transport. Nevertheless, for some of people, journey time is not the main reason for the modal choice. In fact, the mean differs very little across groups.

Table 4. Modal choice by gender

Gender	Public only	Public+Car	Indifferent	Car+Public	Car only
Male	213	244	101	139	189
Female	328	325	130	145	65

The gender has an impact on car usage rate. It is very common that males are using cars rather more commonly than females and females are using a public transport. Although, we have to remember that very often two adults person are going together to the work and the husband is giving a lift to his wife. In the survey only the driver was treated as a car user, but on many occasions there is more than one person in the car. It is also very common that parents provide a lift to school to their children.

The age also has an impact on modal choice. The drivers are in general middle aged (41-50). Young and older persons are more frequent users of a public transport than others. However, they sometimes use a car. It is easy to explain this phenomenon. Simply they do not have enough funds to maintain a car and maybe for that reason they become aim of the public transport. During the analysis of the coefficient we have to remember that the reference group are people above 70 and 83 % of them are using a public transport.

The main economic activity is also an important factor of diversification of people to different transport schemes according to the literature. Workers tend to have higher valuation of time than the others and in many situations they choose a car, but the results from the model are not significant. Despite that, if the sample would be split into contract workers and self-employed, I expect that persons running their own business choose a car while contract workers are rather indifferent between modes. Similar behaviour was observed in Lodz (Suchorzewski 1995). My results indicate this phenomenon because odds ratio for choice between "public transport always or often" and "other alternatives" suggests that in many cases people are really indifferent. The coefficient is close to one, and next coefficient is far larger than 1.

Car owners, especially when affected by physical or mental disability whenever they have the chance of choice, they choose a car in most cases. There are some exceptions for going to the city centre, due to lack of parking spaces. Negative coefficients (odds below 1) for people with driving licences are caused by quite large share (20 %) of young and very old people in the population who are able to drive a car, but either they do not always have access to one or they are afraid to drive or find it costly to use a car. I must remark that this question was asked only to the people who live in the household with at least one car. And those two variables may be highly correlated.

This fact is confirmed by observation that the people who travel to school and work rather choose a public transport, except perhaps the previously mentioned group of self-employed. The main reason for choosing a bus for this kind of journey is the lack of parking spaces in Polish towns and even if they are available they are highly charged[6]. Despite that, the variable is only significant for "public transport always" vs. "other alternatives" and in overall it is only significant at 10 % level.

The following group of questions was asked to car users only. These are the questions regarding usage of the car in the city. Frequent car users, who find that traffic jams disturb the flow of people, are more likely than the other

[6] I do not have the exact figure for Gdynia, but in Warsaw the first hour of parking costs the same as one bus ticket.

drivers to use a private car. For the other groups we observe the opposite effect. Traffic jams are reason for using a public transport. Also perception of a time spent in a car and in the public transport in a very important factor. The people who feel that the car is much faster than a bus or other modes of transport definitely want to use a car. The strongest effect is observed for current car users. Also people who think that travelling by car is cheaper obviously choose that mode, but the effect is only slightly significant. On the other hand, is significant for both modes users, and therefore important from policy making point of view.

For persons who are using both modes of transportation the distance to the nearest bus stop and the waiting time are important factors. Long distance and necessity of waiting makes them more likely to use a private cars. Also the comfort of travel has a high valuation for them. They do not like to travel in overcrowded buses.

The people who use a car as a working tool very frequently choose a car as a mean of transport, but this does not mean that they in any situation use car. Even for people stating that they always use a car odds of "always using a car" vs. "other alternatives" is like 3.5 to 1. However, the variable is significant only for car users. This suggests that they use the same car for getting to work and during workday.

The next group of questions was asked to people who possess a car, but they in general use a public transport. The idea behind these questions was to find out what makes these people use a public transport. The first finding is that is definitely not the road congestion, because they also suffer from this. Also bad condition of the road has no explanatory power to the modal choice. The next two questions are devoted to the availability of parking places. It seems that it is rather "no parking at home" than "no parking space at place of destination" that makes these people use public transport. However both factors has a negative impact on car usage rate. That confirms the previous finding about modal choice for going to work or school. Then, people choose the public transport because there is a shortage of parking spaces in the city centre, and this is not a problem in the suburbs.

The bad technical condition of the car in a view of users has an impact on frequency of use. The answers to a related question indicate that higher exploitation costs decrease the probability of car usage. The reason may be that if a person has a car in bad condition then he/she wants not to use it. The growing costs of maintaining the car has also negative impact on car usage, and they make people more likely to use a public transport. Car owners also are not afraid to use it because of the possibility of having a car stolen or damaged in an accident. The latter factor has no implication on the modal choice of travelling.

There exist two main factors that prevent people from using their cars. The first is that the car is used by another member of the family. The second, people with disabilities sometimes are cautious and do not want to drive a car. For them it is indifferent to go by car or by a public transport, but if they can they use a car because in Poland not all buses are designed in a way to carry disabled people.

The most important fact from an economic point of view is that if prices are perceived as high they do still attract people to use a public transport. The impact of this issue will be further discussed in the conclusion.

The last set of questions was about important characteristic of public transport from consumer point of view. Although I could have expected that they will have a decisive role in modal choice, almost all of them do not play a role at all. Only two: low cost and comfort are significant at 10% level.

Nevertheless, they are not decisive. Despite that I think that their analysis can light up people's perception of public transport. The most frequently chosen answer (one person was allowed to mark 3 responses) was bus stop in close neighbourhood (accessability). The second was punctuality. Only this two answers were named by over 50 % of the respondents. Other important things were high frequency of buses on the line and straight connection from starting point to the destination. Given answers mean that passengers do not like to walk for long distances to the bus stop. The second and third answers are related, because it is natural and was shown in other studies for Poland (e.g. Suchorzewski 1995), that when frequency of buses is high passengers do not notice when bus is late, because they do not have to wait for a long time.

6 Summary and conclusions

Looking at the results of the model it is easy to observe that problem of travel demand within a city can be easily separated into two subproblems. The first is how public transport should be directed not to loose current customers of which the vast share is made of young and old persons and woman's of all age. And the second problem is how to make current car users get out of their cars and use public transport.

In the survey many people claim that the reason for not using a public transport is the long distance to the nearest bus stop, the waiting time and the travel time. These results are confirmed by the model. Beside that speed and high frequency were pointed out, but those factors have little impact in the modal choice and have to be taken with caution. This means that people want to have as many bus stops as possible, but on the other hand they want buses to ride fast. This means that there should be two kinds of bus lines. Slow local line calling at every bus stop, and fast one with few stops. From the policy making point of view is hard to find a balance between passengers expectations and rationality. Because people want to have straight connection to all point that they travel, but from the network designer point of view this makes no sense either operational or economically. It is better to design a net with hubs, in which there will be exchange of the people between different lines. This provides a better functioning of the network as a whole, shorter interval between buses on the main lines and better bus coverage when one vehicle have to return to the garage due to technical problems.

Moreover, fact that only low cost and travel conditions were found to be significant for travel decisions indicates that Polish society are rather poor in comparison to the most developed countries in the EU. The similar features were reported during the study of travel demand of citizens of Sao Paolo (Vasconsellos 1997). This study also concluded that males frequently are using a car to show their physical and welfare status. The car is perceived as essential to perform their desired daily activities and to ensure their social reproduction. For that reasons car is used more frequently than in other developed countries. This implies that information campaign should be taken into consideration. The citizens should be informed about environmental issues caused by the car and advantages of public transport in the city.

Also another factor has to be taken into consideration. In the society opinion cost of public transport is relatively high. But people do not realize how big is economic cost of car usage. There are many externalities which are not seen by the people. So they are underestimating cost of the cars, and due to that they see public transport as expensive. The cost information should be also included in information campaign.

For that reason, in my opinion, it would be very hard to convince people to get out of their cars. Nevertheless, some steps should be taken. First of all policy-maker should think about possibilities of increase the average travel speed by public transport. To achieve that it is enough to made simple adjustment like to give a priority to the public transport, set up the traffic light system in favour of city communication. Furthermore, politicians and policy makers should not afraid the car lobby, and as it is in the other European cities make the special bus lanes on every dual or more carriageway road. These policy adjustments definitely will not be popular at the time of introduction, but in the longer perspective society will benefit from them. There are of course other possibilities like extending the tramways or railways network in most congested directions, but this are hard decisions for at least three reasons. First of all, investment time is about ten years, from the beginning of the study to opening a new line, secondly it is very expensive to do so. Thirdly, to maintain a tram line stream of 20.000 passengers per peak hour is needed. If there are not enough passengers buses are preferred for economic reasons.

Nevertheless, little policy adjustments may not solve the main problem: How to convince people that public transport is as fast or even faster, in some circumstances, than car. This is a big policy or rather political question. Politicians are not likely to do unpopular things and in Poland every policy that is against the car user is unpopular. But without tackling this problem, in my opinion, nothing can be done in the long term. There must be put some effort to change perception of the public transport. This is a better way of dealing with the problem than radical solution like e.g. banning the car usage in the center have to be taken into consideration.

I was trying to get as many features of functioning of the public transport from the given data as possible. The striking thing in the given model is the lack of cost analysis. This shortcoming is caused by the lack of available data on individual's income. In Poland people are afraid, or simply do not want to give an answer to the income related question. Even if they do, the reliability of the data is low. Inclusion of income information is a first possible extension of the model. Also in the future, a separate analysis of travel behaviour of different social groups can be done. In my model impact of social characteristic is not clear, but definitely has an impact on the modal choice.

Of course, there is a possibility to design a different model, which possibly could give more directions to the policy makers. One way is to design more sophisticated model like the Random Parameter Logit or the Mixed Logit, and other way is to design a study more in direction of economic analysis. The shortcomings of my approach are coming from the fact that this study was made from a transport engineering point of view. And for traffic designer economical aspects of mobility are only in the background of the research.

There exits many possible ways to extend this study. The First I mention above - design a other model. The second way is to look at other factors and from different point of views like separate analysis of particular travel purpose (work, school), separate study of travel behaviour by location or by income group.

In this paper I showed one possibility how to tackle a problem of measuring the travel demand. I started form the review of recent related literature. Then I have explored the problem of stated and revealed preferences. I showed that a best solution is to combine them. Afterward I derived my model, which is based upon a random utility model. In the closing sections

I presented my results and I discussed them. At the end I showed policy implications and possible extensions of the model.

References

[1] Ben-Akiva, Moshe, Bierlaire, Michael (1998) *Discrete Choice Methods and Their Applications to Short-Run Travel Decisions* in *Handbook of Transportation Science*

[2] Ben-Akiva M., McFadden D., Train K., Walker J., Bhat C., Bierlaire M., Bolduc D., Boerch-Supan M., Brownstone D., Bunch David S., Daly A., De Palma A., Gopireth D., Karlstrom A., Munizaga Marcela M. (2002) "Hybrid Choicee Models: Progress and Challenges", *Marketing Letters*, 13 vol. 3, pp.163-175.

[3] Bhat, Chandra R. (1995) "A Heteroscedastic Extreme Value Model of Intercity Travel Mode Choice", *Transportation Research Part B*, vol. 29, pp. 471-483.

[4] Bhat, Chandra R. (1997) "Covariance Heterogenity in Nested Logit Models: Econometric Structure and Application to Intercity Travel", *Transportation Research, Part B*, vol. 31, pp. 11-21.

[5] Bhat, Chandra R., Castelar S. (2002) "A Unified Mixed Logit Framework for Modelling Revealed Preferences and Stated Preferences: Formulation and Application to Congestion Pricing Analysis in San Fracisco Bay Area", *Transportation Research, Part B*, vol. 36, pp. 593-616.

[6] Bly P.H., Dasgupta M. (1995) *Urban Travel Demand: A Growing Business, but Can it Last?* in Suchanov L.J [red] *Urban Transport and the Enviroment for the 21st century*, Computational Mechanics Publications, Southampton.

[7] Brownstone D., Bunch D., Train K. (2000) "Joint Mixed Logit Models of Stated and Revealed Preferences for Alternative-fuel Vehicles", *Transportation Research B*, vol. 34, pp. 315-338.

[8] Domencich, Thomas A., McFadden, Daniel (1975) *Urban Travel Demand. A Behavioral Analysis*, Charles River Associates Research Study, North Holland.

[9] Eliasson J., Mattsson L-G. (2001) "Transport and Location Effect on Road Pricing: A Simulation Approach", *Journal of Transport Economic and Policy*, vol. 35/3 pp. 417-456.

[10] Fry, Tim, Harris, Mark (1996) "A Monte Carlo Study on Test for the Independence of Irrelevant Alternatives Property", *Transpotration Research Part B*, vol.30 pp.19-31.

[11] Fry, Tim, Harris, Mark (2002) "The DOGIT Model", Working Paper 7/2002, University of Meulbourne, Department of Econometrics and Busines Statistics, Australia.

[12] Glowny Urzad Statystyczny (2003) *Statistical Yearbook of Poland*, Glowny Urzad Statystyczny.

[13] Greene, William H. (2003) *Econometric Analysis*, Upper Saddle River, New Jersey, 5th edition.

[14] Hausman, Jerry, McFadden, Daniel (1984) "Specification Test for the Multinomial Logit Model", *Econometrica*, vol. 52/5.

[15] Henscher David (2001) "Measurment of the Valuation of Travel Savings", *Journal of Transport Economics and Policy*, vol. 35/1, pp. 71-98.

[16] Knapp, Thomas A., White, Nancy E., Clerk, David E. (2001) "A Nested Logit Approach to Household Mobility", *Journal of Regional Science*, vol. 41/1, pp. 1-22.

[17] Long, Scott J. (1997) *Regression Models for Categorical and Limited Dependent Variables*, Thousand Oak.

[18] Maddala G.S (1999) *Limited Dependent and Qualitative Variables in Econometrics*, Cambridge University Press.

[19] Mahtrian P., Salomon I. (2001) "How Derived is Demand for Travel? Some Conceptual and Measurment Considerations", *Transportation Research, Part A*, vol. 35 pp. 695-719

[20] Mansky, Charles F., McFadden, Daniel [red] (1974) *Conditional Logit Analysis of qualitative choice behaviour* in Paul Zarembka [red] *"Frontiers in Econometrics"*, London Academic Press.

[21] Mansky, Charles F. [red], McFadden, Daniel [red] (1990) *Alternative Estimators and Sample Design for Discrete Choice* in Mansky, Charles F., McFadden, Daniel *"Stuctural Analysis of Discrete Data with Econometric Applications"*, MIT Press, 4th edition.

[22] McFadden, Daniel, Train, Kenneth (2000) "Mixed MNL Model for Discrete Response", "Journal of Applied Econometrics", vol. .

[23] van Ommeren, Jos (1998) "On-the-Job Search Behaviour: Importance of Commuting Time", *Land Economics*, vol 74/4.

[24] Ortuzar, Juan, Gonzalez, Rosa (2002) "Inter Island Travel Demand Response with Discrete Choice Models", *Journal of Transport Economic and Policy*, vol 36, pp. 115-138.

[25] Parsons, George R., Hauber, Brett (1998) "Spatial Boundaries and Choice Set Definition in Random Utility Model of Recreation Demand", *Land Economics*, vol. 74/1, pp. 32-48.

[26] Rouwendal, Jan, Meijer, Erik (2001) "Preferences for Housing, Jobs and Commuting: A Mixed Logit Analysis", *Journal of Regional Science*, vol. 41/3, pp. 475-505.

[27] Small, Kenneth, Hsiao, Cheng (1985) "Multinomial Logit Specification Test", *International Economic Review*, vol. 26/3, p. 419-429

[28] Small, Kennteh (1987) "A Discrete Choice Model for Ordered Alternatives", *Econometrica*, vol. 55/2 pp. 409-424.

[29] Suchorzewski, Wojciech (1996) *Badanie zachowan komunikacyjnych mieszkancow Lodzi 1995*, Urzad Miasta Lodzi.

[30] Train, Kenneth E., McFadden, Daniel, Ben-Akiva Moshe (1987) "The Demand for Local Telephone Service: A Fully Discrete Model of Residential Calling Patterns and Service Choices", *RAND Journal of Economics*, vol. 18/1, p. 109-123.

[31] Train, Kenneth E. (1998) "Recreation Demand Models", *Land Economics*, vol. 74/2, p. 230-240.

[32] Train, Kenneth (2002) *Discrete Choice Methods with Simulation*, Berkeley University Press.

[33] Vasconcellos, Eduardo A. (1997) *The Demand for Cars in Developing Countries*, Transportation Research Part A 31, p. 245-258.

[34] Wang, D., Borges, A., Oppewal, H., Timmermans, H. (2000) "A Stated Choice Approach to Developing Multi-faced Models of Activity Behavior", *Transportationn Research Part A*, 34, pp. 625-643.

[35] Williams, Jerry, Chaloupka, Frank J., Weichsler, Henry (2002) "Are There Differential Effects on Price and Policy on College Students' Drinking Intensity?" *NBER Working Paper*, nr 8702.

A Appendixes

A.1 The IIA Tests

```
**** Hausman tests of IIA assumption

 Ho: Odds(Outcome-J vs Outcome-K) are independent of other
alternatives.

 Omitted |       chi2   df   P>chi2   evidence
---------+------------------------------------
       1 |     -0.000    3      ---    for Ho
       3 |     -0.000    8      ---    for Ho
       4 |      0.000    9    1.000    for Ho
       5 |     -0.000    5      ---    for Ho
----------------------------------------------
 Note: If chi2<0, the estimated model does not
 meet asymptotic assumptions of the test.

Evidence for Ho is problematic. See Fry&Harris "A Monte
Carlo study of test for the IIA"

**** Small-Hsiao tests of IIA assumption

 Ho: Odds(Outcome-J vs Outcome-K) are independent of other
alternatives.
 Omitted |  lnL(full)  lnL(omit)    chi2   df   P>chi2   evidence
---------+-------------------------------------------------------
       1 |   -533.434   -436.167 194.535   31    0.000   against Ho
       3 |   -701.661   -492.674 417.974   31    0.000   against Ho
       4 |   -797.160   -572.727 448.866   31    0.000   against Ho
       5 |   -825.422   -665.912 319.021   31    0.000   against Ho
-----------------------------------------------------------------

Indeed Small-Hsio test reject the null.
```

```
**** Wald tests for combining outcome categories

 Ho: All coefficients except intercepts associated with given pair
     of outcomes are 0 (i.e., categories can be collapsed).

Categories tested |       chi2    df   P>chi2
------------------+------------------------
       1-       3 |    236.504    30    0.000
       1-       4 |    140.341    30    0.000
       1-       5 |    167.579    30    0.000
       1-       2 |    191.629    30    0.000
       3-       4 |    183.448    30    0.000
       3-       5 |    208.290    30    0.000
       3-       2 |    745.733    30    0.000
       4-       5 |    117.273    30    0.000
       4-       2 |   2197.257    30    0.000
       5-       2 |    263.647    29    0.000
-------------------------------------------
```

A.2 The OGEV Estimation Results

In the following two tables are result of estimation. The first table consist parameters for every equation. In the second one there are odds ratio. Both are computed in the way described below.

In the eq1 are coefficients of categories treated jointly: public transport often, indifferent, car often, car always compared to public transport always. In the appendix A.3 are the same result in terms of odds.

In the eq2 are coefficients of categories treated jointly: indifferent, car often, car always compared to treated jointly public transport always and public transport often. In the appendix A.3 are the same result in terms of odds.

In the eq3 are coefficients of categories treated jointly: car often, car always compared to treated jointly: public transport always, public transport often, indifferent. In the appendix A.3 are the same result in terms of odds.

In the eq4 are coefficients of categories: car always compared to treated jointly: public transport always, public transport often, indifferent, car often. In the appendix A.3 are the same result in terms of odds.

Variable	eq 1	eq 2	eq 3	eq 4	F-test
gender	-.077772	-.088559	.043172	.83580+	
age 16-20	-.321091	-.103899	-.27223	-4.71151*	*
age 21-30	.183680	.279003	.341688	-1.14889	*
age 31-40	.254357	.786049	.525642	-.656260	*
age 41-50	.079878	1.10081*	1.42688+	.011439	*
age 51-60	-.380799	.619191	.681071	-.24552	*
age 61-70	-.186465	1.01616*	1.06409	-.591480	*
worker	.066306	-.201248	1.12765	.437685	
student	.348616	.116074	-.170437	1.89017	
retired	-.790516+	-.51272	1.41714	1.25071	
non-worker	-.357988	-.184241	1.43574	.845244	
car time	-.014826	.009253	.033568*	-.00453	*
public time	-.003108	-.000942	-.018485*	-.00373	*
car owner	2.17717**	2.48601**	2.54929**	2.56393**	**
driving licence	-.501351**	-.929019**	-1.44234**	-4.63329**	**
tra work/school	-.698505**	-.423399	-.202492	.094510	+
jams frequency	-.211516**	-.043283	-.204550	.594717**	**
distance to bus	33.6409	.563410	2.18877**	1.26658**	**
no waiting	-19.106	.241448	2.23899**	.707463+	**
travel comfort	3.62418	1.90134**	1.94548**	1.67725**	**
car time percep	17.2199	1.43619**	3.04986**	1.11438**	**
car is cheap	.050669	.5556154	1.42978**	.884856+	*
car is tool	18.8808	.1795384	3.118672**	1.25834*	**
car is safe	(dropped)	20.5416	3.03443**	.621283	**

Variable	eq 1	eq 2	eq 3	eq 4	F-test
road congestion	.583124	-.046346	-.504425	-20.6671	
road condition	-.582621	.161762	-1.33311*	-5.71377	
no home parking	-.461565	-.598556	-3.06823**	-16.7530**	
no parking	1.57350	.228413	-1.29527**	-20.0087**	
car conditiont	18.2868	.584509	-.17457	-5.51800**	**
car costs	-.630147	.048934	-1.39616**	-22.1314	+
car save	18.0440	-.016307	-.618093	-19.1365	
steal prevent	-.090057	-.133973	.413847	-15.7843	
public cost	.684478	-.545692+	-2.01970**	-19.4055	**
public quality	.928982	-.451209	-1.62113**	-19.182	*
disability	-.200824	-.370109	.908542*	-22.1553	+
someone else	.576036	-.056096	-1.41671**	-6.81703**	**
accessibility	.099880	-.707605	.586616	-.189397	
frequency	-.033975	-.787910+	.124835	.631747	
punctuality	-.202538	-1.04113*	.463909	.499968	
reliability	.241069	-.598628	.441794	1.40577	
straightforward	-.269412	-.712998	.415073	.635732	
speed	-.097173	-.669669	1.08570	.545941	
low cost	.146133	-.580903	.968361	-.577460	+
comfort	-.225101	-.969783*	.765468	-.395642	+
information	.510209	-.657679	-2.40571+	3.18436	
rhythm	.044677	-1.52767**	-.321407	-.302240	
constant	.984205	-1.19116	-8.38242**	-9.86407	**

+ indicates significance at 10% level; * significance at 5% level; ** significance at 1% level. In the columns (2)-(5) t-tests. Last shows joint F-test for validity. The variables recoded into dummies were tested jointly.

A.3 The OGEV Estimation Results (OR)

Variable	OR 2 vs 1	OR 3 vs 2	OR 4 vs 3	OR 5 vs 4	F-test
gender	0.92518	0.91525	1.04412	2.30666+	
age 16-20	0.72536	0.90132	0.76168	0.00900*	*
age 21-30	1.20163	1.32181	1.40732	0.31699	*
age 31-40	1.28963	2.19471	1.69155	0.51879	*
age 41-50	1.08315	3.00659*	4.16568+	1.01151	*
age 51-60	0.68332	1.85743	1.97599	0.78231	*
age 61-70	0.82989	2.76257*	2.89819	0.55351	*
worker	1.06855	0.81771	3.08840	1.54912	
student	1.41711	1.12308	0.84330	6.62048	
retired	0.45361+	0.59887	4.12531	3.49283	
non-worker	0.69908	0.83174	4.20276	2.32855	
car time	0.98528	1.00930	1.03414*	0.99549	*
public time	0.99690	0.99906	0.98169*	0.99627	*
car owner	8.82128**	12.0133**	12.7980**	12.9867**	**
driving licence	0.60571**	0.39494**	0.23637**	0.00972**	**
tra work/school	0.49733**	0.65482	0.81669	1.09912	+
jams frequency	0.80936**	0.95764	0.81501	1.81252**	**
distance to bus	0.00000	1.75665	8.92426**	3.54868**	**
no waiting	0.00000	1.27309	9.38387**	2.02884+	**
travel comfort	37.4939	6.69487**	6.99700**	5.35080**	**
car time percep	3.0e+07	4.20463**	21.1124**	3.04768**	**
car is cheap	1.05197	1.74301	4.17778**	2.42264+	*
car is tool	1.6e+08	1.19666	22.6163**	3.51958*	**
car is safe	1.00000	8.3e+08	20.7890**	1.86131	**

Variable	OR 2 vs 1	OR 3 vs 2	OR 4 vs 3	OR 5 vs 4	F-test
road congestion	1.79163	0.95471	0.60385	0.00000	
road condition	0.55843	1.17558	0.26366*	0.00330	
no home parking	0.63033	0.54960	0.04650**	0.00000**	
no parking	4.82350	1.25660	0.27382**	0.00000**	
car condition	8.8e+07	1.79411	0.83982	0.00401**	**
car cost	0.53251	1.05015	0.24754**	0.00000	+
car save	6.9e+07	0.98383	0.53897	0.00000	
steal prevent	0.91388	0.87462	1.51263	1.00000	
public cost	1.98274	0.57944+	0.13269**	0.00000	**
public quality	2.53193	0.63686	0.19768**	0.00000	*
disability	0.81806	0.69066	2.48070*	0.00000	+
someone else	1.77897	0.94545	0.24251**	0.00110**	**
accessability	1.10504	0.49282	1.79789	0.82746	
frequency	0.96660	0.45479+	1.13296	1.88089	
punctuality	0.81666	0.35306*	1.59028	1.64867	
reliability	1.27261	0.54956	1.55550	4.07868	
straightforward	0.76383	0.49017	1.51448	1.88840	
speed	0.90740	0.51188	2.96150	1.72623	
low cost	1.15735	0.55940	2.63363	0.56132	+
comfort	0.79844	0.37917*	2.15000	0.67325	+
information	1.66564	0.51806	0.09020+	24.1519	
rhythm	1.04569	0.21704**	0.72513	0.73916	

+ indicates significance at 10% level; * significance at 5% level; ** significance at 1% level. In the columns (2)-(5) t-tests. Last shows joint F-test for validity. The variables recoded into dummies were tested jointly.