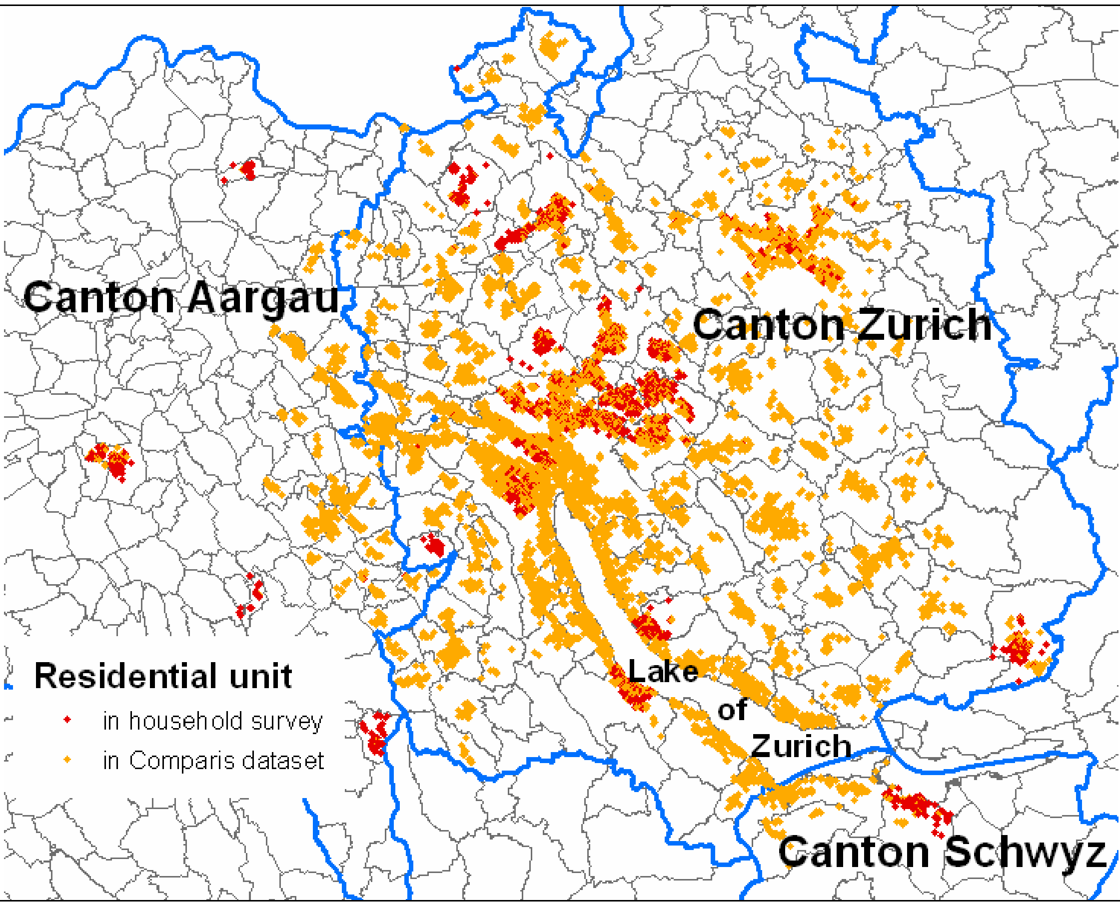


Residential location choice model for the Greater Zurich area

Michaela Bürgle, ETHZ

Conference paper STRC 2006

STRC

6th Swiss Transport Research Conference
Monte Verità / Ascona, March 15. – 17. 2006

Residential location choice model for the Greater Zurich area

Michaela Bürgle
IRL
ETH Hönggerberg
8093 Zurich

Phone: 0041 – 1 – 633 45 43
Fax: 0041 – 1 – 633 11 02
email: buergle@nsl.ethz.ch

March 2006

Abstract

In the context of a land use – transport simulation for the Greater Zurich area that is currently being developed at ETH Zurich, multinomial logit models of residential location choice have been estimated. The estimations were based on two datasets: a household survey conducted in 2005 and real estate offers collected from the Internet in 2004 and 2005. Drawing on these data sources, a number of different variable combinations were tested in order to obtain models with satisfactory explanatory power. In addition, estimation results for specific household types were compared.

A core set of variables was used as a starting point to explore the significance of various characteristics of household, dwelling and location. This core set was then tentatively extended by adding further variables. Variables that proved significant comprised mostly location-related attributes like various densities or travel time to city centre but also characteristics of municipalities like the rate of vacant rental units or the tax index and housing unit's features such as price and size. Interaction terms with sociodemographic attributes of the decision-making households were introduced to improve the explanatory power of the models.

More detailed results were obtained by estimating separate models for different household types. Those types were formed regarding sociodemographic and socioeconomic features.

For choice set selection, random sampling was applied. This approach might later be extended to include stratified sampling strategies making use of similarities between chosen and non-chosen alternatives. For carrying out the sampling procedures a custom made Java programme was developed.

The models resulting from the estimation process are presented and evaluated in the paper at hand.

Keywords

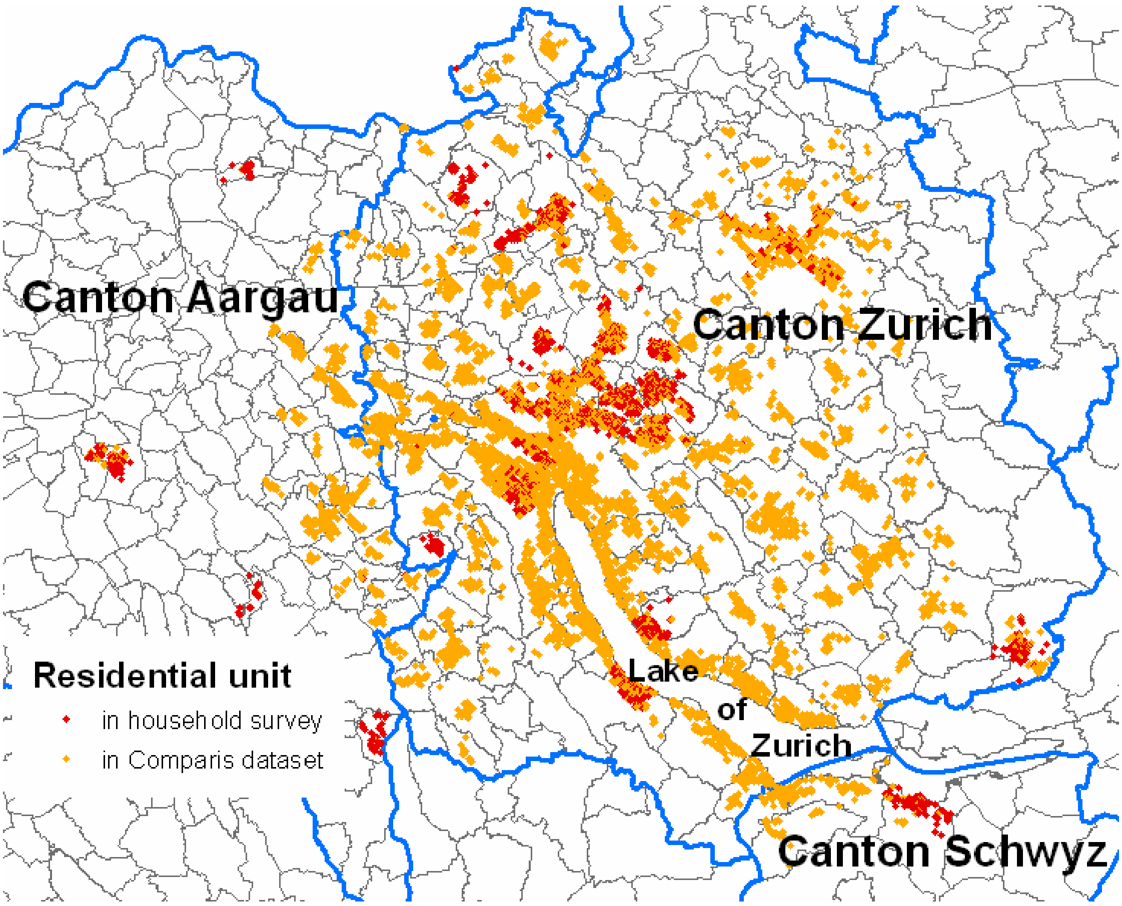
Discrete choice modelling – residential location choice – choice set determination – multinomial logit – household types

1. Introduction

Currently, research is being done at ETH Zurich on a land use – transport simulation for the Greater Zurich area (see Bürgle 2005 and NSL 2005 for detailed information). The software used for the simulation, UrbanSim (Waddell 2002), aims to model the behaviour of different actors and processes in land use and transport. The households living within the simulation area are one example for agents represented within the simulation framework. The estimation of discrete choice models for residential location choice is an integral part of setting up the simulation system. The first models estimated in the course of this work are presented in the paper at hand.

To start with, the datasets used for obtaining alternatives for the multinomial logit estimation are described. Basically, there was one dataset comprising chosen locations and another set from which other alternatives were sampled. Sources and contents of the two datasets are outlined in section 2. Figure 1 displays the data points used for sampling and demonstrates the spatial scale of the models presented in this paper.

Figure 1 Spatial scale of the residential location choice models



The next section of the paper sets out the model specification, going into model structure, variable selection and choice set formation. Subsequently, the resulting models are presented in detail in section 4. One model for all types of households and corresponding estimation results for different household types are presented and discussed. These are general models of residential location decisions in the Greater Zurich area drawing on all variables available in the datasets. Models for use in the land use – transport simulation in contrast can only make use of variables available endogenously in the simulation. Such constrained models are yet to be developed. The paper closes with a summary of the insights gained in the modelling process and makes suggestions for further investigations.

The estimation was conducted using the optimisation toolbox BIOGEME (version 1.4) provided by Michel Bierlaire (Bierlaire 2006). The sampling of alternatives was done with a custom made Java programme (see 3.3).

2. Data used for modelling

To provide chosen and other alternatives for the discrete choice modelling, two different data sources were used: Revealed preference information about households in the Greater Zurich area was gathered by means of a household survey conducted in 2005. In addition, a large number of real estate offers was obtained from the Internet. All data records acquired in this manner were geocoded (Waldner et al. 2005) and subsequently augmented with spatial information by applying GIS analysis.

From both datasets, only rented property was used for estimation, as price was considered an important variable influencing the choice and the number of records with information on purchase prices was too small. The data was checked for suspicious or missing values. Outliers were not considered for estimation. This affected attributes like the rent, where prices below 6 CHF or above 60 CHF per sqm were deemed unreasonable, or the size of the housing unit, where units smaller than 20 sqm or larger than 500 sqm were not regarded. If site-related information could not be obtained for a location (e.g. regional accessibility for data records outside the range of the regional transport model), the corresponding record was also not used.

2.1 Household survey in the Greater Zurich area

In winter 2005 a survey was shipped to 9,330 households in 21 municipalities of canton Zurich and surrounding cantons plus four city districts of Zurich. The survey contained questions concerning sociodemographic features of the households, characteristics of their dwelling and housing price information (see Waldner et al. 2005 for detailed information about the conduction of the survey and Löchl et al. 2005 for the results obtained). The return rate of the survey was 36% yielding around 3,300 household records.

These household records were geocoded where possible. However, only those households that had occupied their present dwelling for no more than five years when answering the survey were considered recent movers and therefore eligible for the modelling of residential location choice. While this constraint left almost 2,000 data records for estimation, a varying amount of these could not be taken into account depending on the variables used for modelling: The use of attributes like rent or income in the models reduced the set, as the corresponding values were missing in several of the records. Such records could not be considered. Finally, a total of roughly 800 to 1,000 records was actually used, depending on the model.

As the household survey delivered information about revealed preferences, these records were used as chosen alternatives for the location choice models. The non-chosen alternatives were taken from the second data source (see 2.2).

2.2 Real estate offers

To complement the information collected in the household survey and to gather a reasonable large amount of data to build meaningful models, real estate offers were obtained from the online real estate portal “comparis” (http://www.comparis.ch). The webpages were parsed using a Java programme: Data posted on the Internet in the period from December 2003 until October 2005 was scanned to collect a comprehensive database of real estate bids for the area in which the household survey (see 2.1) had been conducted. The real estate information gained through this procedure was subsequently geocoded. The resulting database comprises around 20,000 records for the simulation area of the project plus the additional municipalities included in the household survey (see 1). As was the case with the survey data, not all available records could be used for estimation depending on the quality of geocoding and the variables considered in the models. For example, only around 10,000 of the collected records have rent price information.

2.3 Comparability of the datasets

The comparis dataset was used to sample non-chosen alternatives to the locations provided by the survey. This approach is not without pitfalls as the two datasets sport some systematic differences. The prices collected from comparis are bid prices which exceed the average rents declared in the survey: The average rent per sqm paid by the recent movers who ansered the household survey amounts to 17.50 CHF and is notedly exceeded by the average offer of 20.75 CHF. This difference can partly be accounted for by the fact that offers from cooperative building associations and low priced residences that are passed on underhand are not included. To balance this systematic difference, the bid prices were multiplied by 0.845. Another difference between the two datasets concerns the spatial distribution of data points: As apparent in Figure 1, the comparis dataset covers all of the area under consideration, while for reasons of cost and expenditure the survey was only conducted in representative municipalities (see 2.1). Furthermore, the focus of the survey was on the Glattal, a suburban conglomerate of municipalities northeast of Zurich, which is also the centre of interest for the land use – transport simulation, the intended use of the residential choice model. One of the consequences is that a high proportion of survey data is located near Zurich airport which has implications on the use of aircraft noise as an explanatory variable (see 4.1).

3. Model specification

3.1 Model structure

Multinomial logit was deemed an appropriate approach to the estimation task at hand in accordance with McFadden (1978) who showed that unbiased parameters can be produced in the face of a large number of alternatives by using a random sample of the universe of the available choice set for alternatives.

The multinomial logit formulation of a discrete choice problem is based on random utility theory. It is assumed that a decision maker, when offered several alternatives, will select the alternative that offers the highest subjective utility to him or her. In our case, utility functions were specified for the alternative housing locations in order to assess the utility of each alternative for the decision-making household. These functions are based on formula (1) which expresses the utility of alternative j for household q as

$$U_{jq} = u_{jq} + \varepsilon_{jq} \qquad (1)$$

where u_{jq} denotes a systematic part of the overall utility and ε_{jq} a random part that cannot be explained with the aid of the observations made. The systematic part in turn can take the form of a linear combination of the selected explanatory variables (see Ortúzar and Willumsen 2001, p. 223). The basic linear formulation can be extended by transforming variables or inserting interaction terms. Under the assumption that the random part is independently identically distributed, the following multinomial logit function can be derived as the probability of household q choosing alternative i out of all possible alternatives

$$P_{iq} = \exp(u_{iq}) / \Sigma_j \exp(u_{jq}) \qquad (2)$$

with P_{iq} the probability and u_{jq} the utility of alternative j for household q as denoted in formula (1).

3.2 Explanatory variables

There is a range of publications available indicating what types of variables to use for the estimation of residential location choice. Generally, discrete choice models are based on the assumption that the probability for a decision maker to choose a given alternative is a function of his socioeconomic characteristics and the relative utility of the alternative (Ortúzar and Willumsen 2001, p. 220). The attractiveness of a residence in turn can be ascribed to attributes of the dwelling itself and attributes of its location.

The selection of explanatory variables was geared to the following working hypotheses and to the data availability:

- Factors influencing residential location choice depend on the type of household making the location decision
- Households prefer to spend as little as possible of their income on housing
- Households with employed persons prefer housing locations close to their place of employment
- Households with children prefer to live in areas with many children
- Young households without children prefer locations with high population density
- Older and retired households prefer locations with a high proportion of open spaces
- Municipality characteristics like the tax index or the rate of vacant housing units influence residential location choice
- Households tend to avoid locations with heavy noise immissions
- Environmental site characteristics like proximity to bodies of water or sunshine exposure may increase the utility of a residential location
- Households generally value a good local supply of retail trade
- The accessibility by individual or public transport in the Greater Zurich area does not show differences big enough to significantly influence residential location choice but good accessibility by public transport is important for households without a car

At first, models containing only the variables considered as core variables were used and it was determined, which of those variables contributed significant explanatory power. Different interaction terms with sociodemographic variables were tried. Based on the first results, non-linear formulations were tested for rent and distance variables. Then followed the stepwise introduction of additional variables to test if there was an added explanatory power. Subsequently, selective experiments were made with different formulations of the rent variable. The list of variables that were used is given in Table 1, the data sources for rental objects are described in section 2. Additional attributes were created by GIS analysis, making use of generally available statistical data and by the application of transport models.

Table 1 Selection of variables considered for residential choice estimation

Variable	Description	Mean	Median	Std.Dev.	Unit
Rent	Total monthly rent[1]	1578.87	1411.15	823.99	CHF
Rent/sqm	Rent per sqm[1]	17.42	16.35	4.91	CHF/ sqm
Sqm	Sqm of the housing unit	106.39	98.00	50.65	Sqm
Distance to place of employment	Distance between residential location and place of employment	10.42	6.77	14.06	Km
Density of children	Average number of children per hectare measured in a radius of 500m	7.65	7.22	3.72	Persons/ ha
Population density	Average number of inhabitants per hectare measured in a radius of 1km	28.03	21.46	23.29	Persons/ ha
Density of open space	Average sqm of open space per hectare measured in a radius of 2 km	6103.92	6741.36	2251.36	Sqm/ ha
High noiselevel	Proximity to major road or high railroad noise level	0.38	0.00	0.48	Boolean
Rental vacancy rate	Vacancy rate of rented housing (municipality level)	1.03	0.70	1.18	Percent
Tax index	Ratio of tax rate to the cantonal average weighted with total tax payers multiplied by total tax burden (municipality level)	91.45	96.15	14.14	
Travel time to Bürkliplatz	Car travel time to Zurich centre (Bürkliplatz) based on regional transport model	30.70	31.00	8.99	Minutes
Mean sunshine index	Index of sunshine exposure (mean of nine points of time per year)	9.09	9.14	0.74	
Density of jobs	Density of jobs in retail trade per hectare measured in a radius of 1 km	0.47	0.20	0.98	Jobs/ ha
Public transport accessibility	Public transport accessibility to population based on regional transport model	71.20	29.00	98.12	

[1] With prices from real estate database weighted as described in 2.3

3.3 Choice set

The choice set is the set of alternatives the decision maker is offered when making his choice. As the projected land use – transport simulation for Greater Zurich was the context in which the residential location choice was to be estimated, the choice set of alternative household locations theoretically comprised all housing units in this area plus a 2 km buffer zone around it. Alternatives that could be used for estimation were determined by the data records available through the household survey (see 2.1) and real estate bids (see 2.2). In practice, the choice set was formed by collocating each of the housing units distinguished as chosen alternative by appearing in the revealed preference survey with a randomly chosen sample of housing units from the real estate bids. The bid data was used for sampling because of its greater size and better coverage of the area. However for comparison, the estimation of models using only revealed preference data will be carried out (see also 5). The number of alternatives provided for estimation was increased from an initial 10 to 40 per record to obtain more stable estimation results.

Two sources of estimation bias can be rooted in the way the choice set is formed. Firstly, there might be choices available to the decision maker in reality which are not used for estimation. Secondly, choices might be used for estimation that are not considered by the decision maker in reality.
The first problem can only be alleviated by using a sample that is big enough and representative for the area under consideration. In order to achieve this, the comparis dataset was added to the alternatives known through the revealed preference survey.
The second question has not been tackled yet in the context of this research. Ideas for the selective improvement of choice set composition are brought up in section 5.

For collocating the alternatives used for model estimation, a sampling programme was coded in Java. The programme compiles a dat-file containing several attributes for a specified number of alternatives as required by the estimation software BIOGEME. An XML-file is used for specifying the settings of each sampling pass. Possible settings include the specification of attributes to be included in the output file and the number of alternatives to be sampled. It is also possible to indicate attributes that are only available for the chosen alternative (e.g. sociodemographic variables from the household survey). Extensions for the conduction of more sophisticated sampling strategies like stratified samples taking into account similarity measures of alternatives have partly been implemented but have not been extensively used yet.

4. Results

4.1 General model

Table 2 shows the model estimation results for a general model of residential location choice in the Greater Zurich area. This model was estimated with a sample of 878 recent movers from the survey data. The signs of the estimated betas remained robust over a series of estimations with varied variable sets. The significance of some variables was not always given in the process of model evolution, while others always made part of the model. Most of the variables that made their way into the final model are related to the location of the dwelling. The housing unit itself influences the location decision through rent and size. Sociodemographic characteristics of the decision-making household were used in interaction terms.

Table 2 Model parameters for residential location choice in the Greater Zurich area[1]

Variable	Beta
Distance to work	-4.302
Exponent of distance to work[2]	+0.201
Tax index of municipality	-0.028
Rental vacancy rate for municipality	-0.162
Travel time to Zurich Bürkliplatz	+0.053
Ln of accessibility to population by public transport * no car dummy	+0.551
Population density * young household dummy	+0.006
Density of children * family with young children dummy	+0.042
Proximity to major roads or high railroad noise level	-0.177
Rent/ income - ratio	-2.159
Floorspace divided by square root of household size	+0.006

[1] all of the variables listed here are significant at the 5% level, rho-square is 0.19 (explanations of variables see Table 1)

[2] non-linear formulation of distance to place of employment: BETA * (distance) ^ EXPONENT

The distance to place of employment shows a negative sign. This result confirms the hypothesis that households prefer residential locations close to the place of employment. The explanatory power of this variable could distinctly be increased by chosing a non-linear formulation. The result shows that great distances to place of employment have a relatively weaker negative effect on the utility of a location than small distances.

Considering the influence of municipality characteristics on location choice, it can be stated that both tax index and vacancy rate of rental floor space proved significant and negative in all models. These findings are in line with the expectations: A higher tax index indicates a greater financial burden for the decision making household. The vacancy rate can be regarded as a general measure of attractiveness of a municipality as residential location: Higher values point to unattractive features.

Interestingly enough, the travel time to Zurich centre was always significant and positive. In this finding, a trend to urban sprawl seems to become manifest. The sign remained positive in estimations for distinct household types (see 4.2).
The accessibility by private or public transport showed no significant influence on residential location choice as had been anticipated. In the case of accessibility by individual transport, even the sign of the parameter was extremely unstable. This endorses the hypothesis formulated in 3.2. Introducing an interaction term representing the accessibility to population by public transport and the absence of cars in the household however yields a significant result with positive sign. These findings confirm the assumption that accessibility only has an impact on residential location choice in connection with the availability of mobility tools in the decision-making household.
The other interaction term with car availability that was tried for the density of jobs in retail trade per hectare also yielded a positive sign as expected but did not prove significant in the final model. A general influence of the density of jobs in retail trade on households' location choice according to the hypothesis formulated in 3.2 could therefore not be confirmed.

The population density was found to be significant when differentiating between young, childless households and other types of households and introducing a corresponding dummy variable. More detailed results are presented in 4.2. The result is again compliant with the expectations.
Another density measure with significance is the density of children per hectare for households with children under 12 years old. Again a positive sign can be observed demonstrating that households with children prefer to settle in areas where other families live. Again, more detailed information is presented in 4.2.
For the influence of the proportion of open space on decisions of retired households, a positive tendency could be observed. However the number of observations available was not sufficient to yield stable results.

The hypothesis about the influence of proximity to bodies of water could not be verified with the two datasets used, as the spatial distribution of the respective data points in respect to Greater Zurich's larger lakes was too uneven. The sunshine index only showed significance when connected to density of open space, a result that is difficult to interpret and together with the other findings points to the conclusion that open space might be the more useful of the two variables.

With regard to noise, the proximity to major roads or railway lines showed a negative influence on residential location choice as was to be expected. Aircraft noise was not included in the noise variable. The reason for this is that a big proportion of the surveyed households used as chosen locations are situated relatively close to Zurich airport, a fact that resulted in a positive sign of noise as soon as aircraft noise was regarded.

As to characteristics of the housing unit itself, the ratio of rent price to household income was always negative and appeared significant in most estimation runs, also in the final model. The negative sign shows that households indeed tend to minimise the fraction of their income spent on rents. Several different linear and non-linear formulations and income-elasticity were tried out in the hope of obtaining more explanatory power. The rent-income ratio however yielded the best results.
A second attribute of the dwelling that was found significant for residential location choice was its size, but only in relation to the square root of household members. The related parameter sports a positive sign indicating that bigger housing units are preferred with the number of household members showing the biggest influence for smaller households.

4.2 Results for different household types

To obtain more detailed information about residential location choice decisions, models were estimated for different types of households, leaving out the sociodemographic interaction terms used in the general model but apart from this keeping the model specification of the general model. The types of households considered are listed in Table 3. These types do not describe disjoint sets of households.

Table 3 Household types for location choice models

Description	Characteristics	Cases
Families with small children	At least one child in the household, youngest child is less than 12 years old	193
Young adults	No children, all members under 35 years old	332
Older adults without children	No children, not “Young adults”	353
Single household	One adult only	341
Pair household	Two adults, no children	287

Building on knowledge gained in the course of the previous estimations about sociodemographic variables that might influence decisions, these household types were parted into car owners and households without cars. In a second step, single households, family households and older households without children were separated into three income classes (low: less than 4000 CHF, medium: at least 4000 but less than 9000 CHF, high: 9000 CHF or more). Separate models were estimated for all subtypes, but for many of those subtypes the estimation results were not stable or the case numbers rendered it impossible to make meaningful statements. For single and family households e.g. the case numbers with high incomes were too low to estimate separate models. Here, medium and high income households were merged. The results are presented in Table 4.

The distance to place of employment has a negative sign for all types of households and is significant for all major types. The parameter values are mostly higher for households without a car. Family and single households without a car show the highest values, the latter in conjunction with a low exponent value. This indicates that a relatively low distance to the place of employment already has a strong negative influence on the utility of a location which does not grow much stronger with increasing distances. For singles with car and high income on the contrary, the interrelation of distance and impact of the variable is closer to linear.

The municipality's tax index is significant and negative for all types of households but shows no clear tendencies for the subtypes.
The vacancy rate of rental objects in the municipality is significant for all but pair households and always negative for those subtypes where it is significant. For those subtypes a stronger influence on the utility for medium or high income households can be observed.

The travel time to Bürkliplatz in the centre of Zurich is significant for all types except families and always positive. A stronger impact of this factor is suggested for higher income households with cars and for households without cars.
The accessibility by public transport is only significant for the childless household types (except singles) without car. This confirms the interaction term with car ownership that was introduced to the general model.

Population density is significant and positive for younger and pair households, with a stronger impact on those without car. Older, childless households with a car but with low income seem to prefer residential locations with low population densities, while households of the same group without a car seek higher population densities.
The density of children in the immediate vicinity is only significant for families. This is a result that is reflected in the general model by introducing the appropriate interaction term.

The noise exposure was found to be significant only for families where it shows the expected negative sign.

The ratio of rent to income was negative and significant for all household types under consideration. A stronger influence on young childless and on family households becomes apparent. The significance for different income groups is not always given but the negative influence on utility for households with cars seems to decrease with higher income. It also appears to be relatively strong for households without cars.

The ratio of living space to the square root of household members is significant for all but young childless households and positive for all except single households. When differentiating between different income groups, especially for older childless households articulate discrepancies become manifest: Low income households seek relatively small housing units, while for households of high income the sign of the parameter turns positive, indicating that larger housing units are preferred.

Generally, the explanatory power of the model applied to separate household types is larger than that of the general model using interaction terms. The distinction between households with and those without a car improves the results drastically in some cases.

Table 4 Model parameters for different household types[1]

Household type	Single household				Pair household			Young household		
Car availability	All	Yes		No	All	Yes	No	All	Yes	No
Income class		Low	Not Low							
Distance to work	-2.281*	-3.921*	-1.108*	-12.036*	-1.990*	-0.124*	-1.752*	-4.577*	-0.827*	-4.899
Exponent of distance to work	0.295*	0.190*	0.485*	0.063*	0.339*	0.449*	0.345*	0.175*	0.529*	0.136
Tax index	-0.028*	-0.042*	-0.021*	-0.024	-0.029*	-0.027*	-0.021	-0.029*	-0.030*	-0.020
Rent vacancy rate	-0.164*	-0.106	-0.269*	0.014	-0.108	-0.121	0.062	-0.151*	-0.015	-0.089
Travel time to Bürkliplatz	0.039*	0.026	0.041*	0.050*	0.075*	0.065*	0.141*	0.059*	0.058*	0.094*
Accessibility /public transport	0.031	0.044	-0.046	0.032	0.113	0.050	0.551*	0.097	0.007	0.482*
Population density	0.000	-0.016	-0.012	0.013	0.011*	0.000	0.041*	0.009*	0.000	0.023*
Density of children	0.018	0.061	0.042	-0.010	0.020	0.045	-0.067	0.007	0.003	-0.034
Proximity to road or railway	-0.069	-0.140	-0.094	0.023	-0.173	-0.128	-0.356	-0.097	-0.226	0.255
Rent / income	-1.673*	-3.370	-0.115	-2.203*	-1.824*	-0.686	-1.638*	-2.314*	-0.419	-1.075
Floorspace / household size	-0.007*	-0.013	-0.000	-0.020*	0.015*	0.012*	0.004	-0.001	-0.005	-0.018*
Rho-square	0.167	0.225	0.157	0.28	0.164	0.15	0.307	0.142	0.12	0.246
Number of cases	341	94	132	100	287	229	58	332	237	95

[1] Significance at the 5% level is indicated by '*'

Table 4 (continued) [1]

Household type	Family				Older household without children				
Car availability	All	Yes		No	All	Yes			No
Income class		Low	Not Low			Low	Medium	High	
Distance to work	-4.055*	-0.919*	-0.667*	-7.537	-3.965*	-2.288*	-3.001*	-1.662	-2.550*
Exponent of distance to work	0.250*	0.599*	0.686*	0.655*	0.217*	0.304*	0.304*	0.306*	0.285*
Tax index	-0.017*	0.006	-0.023*	-0.010	-0.027*	-0.027*	-0.025*	-0.019	-0.028*
Rent vacancy rate	-0.201*	-0.347	-0.144	-0.278	-0.149*	-0.182	-0.261*	0.142	0.118
Travel time to Bürkliplatz	0.015	0.000	0.036	0.073*	0.049*	0.028	0.067*	0.014	0.091*
Accessibility /public transport	-0.021	0.190	-0.017	0.360	0.059	0.063	-0.043	0.311	0.390*
Population density	-0.011	-0.029	-0.006	-0.004	-0.000	-0.031*	-0.017	-0.015	0.028*
Density of children	0.084*	0.147*	0.062	0.107	0.017	0.055	0.061	0.059	-0.062
Proximity to road or railway	-0.506*	-0.421	-0.283	-0.167	-0.045	0.295	-0.057	0.037	-0.229
Rent / income	-2.944*	-0.286	-0.108	-0.381	-1.630*	-1.792*	-0.514	0.703	-1.045
Floorspace / household size	0.024*	-0.012	0.018*	0.006	0.008*	-0.012*	0.009*	0.018*	-0.008
Rho-square	0.279	0.297	0.248	0.242	0.179	0.227	0.214	0.175	0.259
Number of cases	193	41	104	38	353	59	173	43	78

[1] Significance at the 5% level is indicated by '*'

5. Summary and Outlook

The estimation results for residential location choice in the Greater Zurich area confirmed most of the working hypotheses formed beforehand. It showed clearly that in many cases household characteristics are necessary to formulate meaningful variables while other site-related attributes' influence on location choice does not depend significantly on the type of decision-making household. Significant variables of the final model comprise mostly location-related attributes like various densities or the travel time to city centre but also characteristics of municipalities like the rate of vacant housing units or the tax index and the housing unit's features price and size.

The structure of the estimated models depends to some degree on data availability issues. The different structure of the two datasets used for alternative sampling might have precluded variables that could prove significant in a residential location choice model. This assumption will have to be verified by drawing samples from the survey dataset only. Like this, more dwelling-related variables could be tested that are not available in the comparis dataset. Also the differences between the two datasets described in 2.3 prevent some location-related variables from being used. On the other hand, using only the survey data greatly reduces the number of alternatives available for sampling.

Another problem connected to the data used has been described among others by Biggiero and Pagliara (2000, p. 7): The revealed preference data does not necessarily reflect the households' true preferences but also the market conditions. In addition, real world data is often strongly correlated, making it difficult to separate influences of different factors. First experiments with residential location choice models drawing only on survey data have confirmed this proposition by providing less explanatory power.

One field of work that future research will focus on is the formation of choice sets. Stratified sampling making use of similarities between the chosen alternative and the other alternatives collated for estimation is one possibility of arriving at more realistic estimations. Behind this strategy stands the concept that a given household will not admit every single housing unit in the region under consideration in its choice set but will probably search for housing in a subregion or among certain types of house. To reflect this heuristic search strategy, a selective sampling will be considered. This implicates the sorting and classification of alternatives in respect to some distance measure. Random samples can then be drawn separately for each of the specified distance classes.

The models presented in this paper are general models of residential location choice in the Greater Zurich area. As pointed out in section 1, the major incentive for performing the modelling task was the need for such a model within the land use – transport simulation of Greater Zurich. A model for the simulation however will have to manage with whatever area-wide data is endogenously available in the simulation database. For some of the attributes contained in the available data sources (see 2) this requirement does not hold. Variables from the models presented here that cannot be used in a confined model include the distance to place of employment and the size of the housing unit. While households with children can be identified, it might be hard to distinguish those with young children. The young households (age below 35, no children) cannot be safely identified because within the currently envisaged simulation only the age of the respective head of household is known. Therefore a separate confined model will have to be estimated for use within the simulation framework. Future estimation runs will tell what is left of the model's explanatory power when adjusting it to these constraints.

The simulation framework on the other hand opens possibilities to introduce additional information. It utilises a synthetic population of households that has been created area-wide for the simulation area (see Bürgle et al. 2005). This makes it possible to calculate e.g. percentages of high-, middle- or low-income households within a certain perimeter, an approximated information that is not publicly available through statistics and has not been used yet for the general model.

6. References

Bierlaire, M. (2006) BIOGEME 1.4 homepage <http://roso.epfl.ch/biogeme> [accessed in March 2006]

Biggiero, L. and F. Pagliara (2000) Calibrating residential location choice model: SP and RP data, *Trasporti Europei*, **16**, 5—9.

Bürgle, M., M. Löchl, U. Waldner und K.W. Axhausen (2005) Land use and transport simulation: Applying UrbanSim in the Greater Zürich area, paper presented at CUPUM, London, June 2005.

Löchl, M., U. Waldner und M. Bürgle (2005) Haushaltsbefragung zur Wohnsituation im Grossraum Zürich – Tabellenband, *Arbeitsberichte Polyprojekt Zukunft urbane Kulturlandschaften*, **2**, NSL, ETH Zürich, Zürich.

McFadden, D. (1978) Modelling the choice of residential location. In Karlquist A., L. Lundquist, F. Snickars and J.W. Weibull (eds.) *Spatial Interaction Theory and Planning Models*, 75 – 96, North-Holland, Amsterdam.

NSL (2005) Project homepage "Infrastructure, Accessibility and Spatial Development" <http://www.nsl.ethz.ch/zuk/verkehrUndRaum> [accessed in February 2006]

Ortúzar J. and L.G. Willumsen (2001) *Modelling Transport*, John Wiley & Sons Ltd, Chichester.

Waddell, P. (2002) UrbanSim: Modeling Urban Development for Land Use, Transportation and Environmental Planning, *Journal of the American Planning Association*, **68** (3), 297—314.

Waldner, U., M. Löchl, M. Bürgle und K.W. Axhausen (2005) Haushaltsbefragung zur Wohnsituation im Grossraum Zürich – Feldbericht, *Arbeitsberichte Polyprojekt Zukunft urbane Kulturlandschaften*, **1**, NSL, ETH Zürich, Zürich.

Waldner, U., H. Machguth, S. Simonet und K.W. Axhausen (2005) Geokodierung von Adressen: Methodik und Erfolgsquoten bei unterschiedlichsten Datenquellen, *Arbeitsbericht Polyprojekt Zukunft urbane Kulturlandschaften*, **3**, NSL, ETH Zürich, Zürich.