# Measuring Location in Residential Location Choice: An Empirical Study on the Canton of Zurich

P.M. Schirmer[1], M.A.B. van Eggermond[2] and K.W. Axhausen[1]

[1] Institute for Transport Planning and Systems, ETH Zürich, Zürich
[2] Singapore-ETH Centre, Future Cities Laboratory, Singapore
patrick.schirmer@ivt.baug.ethz.ch; eggermond@ivt.baug.ethz.ch; axhausen@ivt.baug.ethz.ch

## Abstract

In transportation and land-use research, discrete choice models have become a common method for assessing the value or utility of discrete alternatives for an individual choice-maker and have led to the simulation of land-use developments on a microscopic level. Discrete location choice models represent relocation behavior in those simulations and generally implement three groups of variables, representing attributes of the alternative, the decision-taker and the location. With the growing availability of spatial data on a disaggregated level, a large number of location variables have been reported in these models, which reduces their comparability and their transferability to other study areas. To address this limitation, Schirmer et al. (2012) classified location variables and proposed a common set of attributes as an initial setup. In this paper, we explore the impact of these attributes on residential location choice in the Canton of Zurich.

## 1. Introduction

McFadden (1978) first introduced the discrete choice framework for use in residential location choice. Since then, this framework has gained tremendous popularity in residential location modeling – judging from the number of studies. The attractiveness of the framework is due to, among others, the discrete nature of the residential location decision, the possibility to capture trade-offs between attributes within an alternative and the possibility to differentiate between chosen and non-chosen alternatives and their attributes (Guo*,* 2004). This has led to the use of residential location choice models in a wide range of fields, such as urban planning, policy-making, sociology, and real estate development. In addition, they are also integrated in land-use models and simulations that provide future forecasts on land usage and transport demand.

With the increasing availability of highly disaggregated data, the possibilities of defining the residential alternative and its attributes have evolved. Whereas earlier studies used administrative districts or transport zones as choice alternatives (Weisbrod *et al.,* 1980, Anas*,* 1982) and aggregated characteristics of these zones as attributes, recent models consider buildings or units as choice alternatives and include building-specific attributes as well as location-specific attributes (Habib and Miller*,* 2009*,* Lee and Waddell*,* 2010a).

A broad range of hypotheses have been tested with residential location choice models, made possible by the aforementioned availability of data sources. However, most studies can only be compared in their methodological approach. Comparing household or individual behavior between studies and study regions is often not possible due to the different sets of variables included in the alternative description and the differences in attribute measurement. Moreover, the diversity of attributes that is used in residential location choice prevents the definition of a common data model for the simulation of decision processes in different study areas (Figure 1).

In this paper, we propose a common set of variables that describe the decision-maker and the complexity of the residential choice. Generally, variables are grouped into three categories, describing the decision-maker, the alternative and the location. Following Schirmer *et al.* (2012), we propose to divide the location into four categories representing the socio-economic environment of the household, the built environment, accessibility and points of interest. As a case study, we use a survey conducted among households that have recently moved within or to the Canton of Zurich, Switzerland.

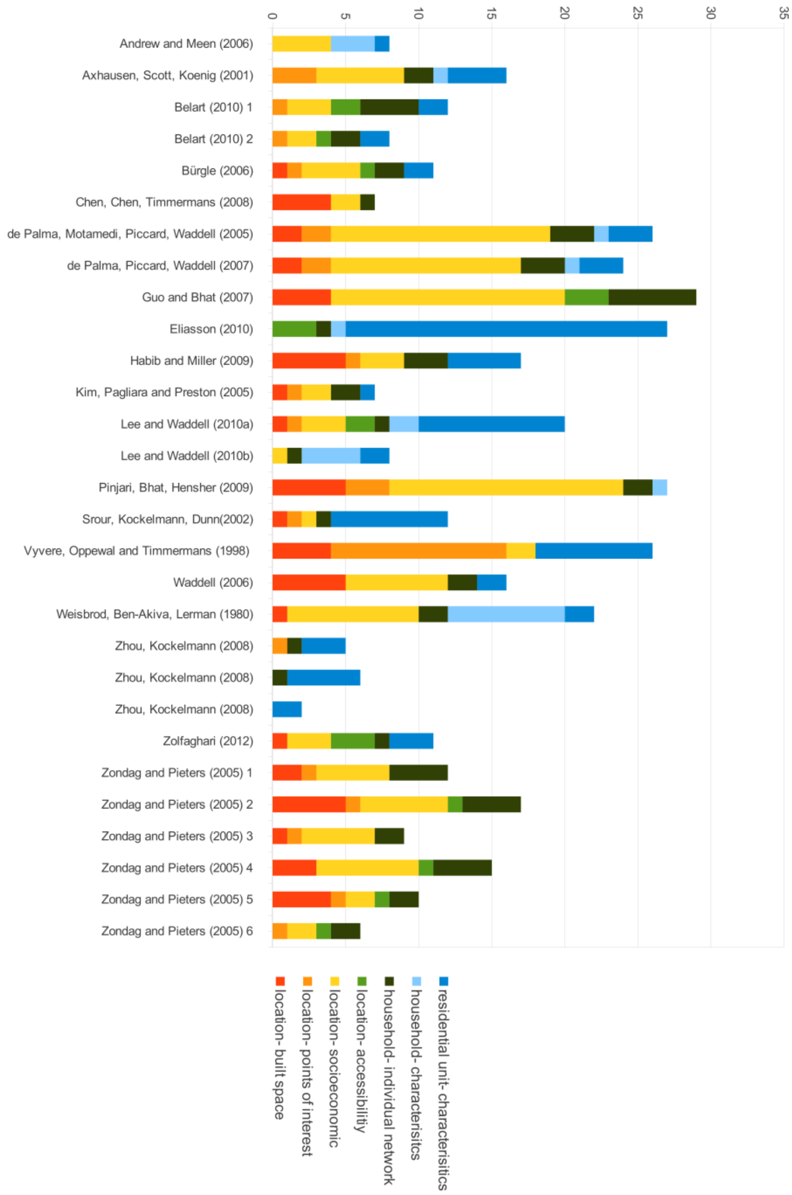


**Fig. 1.** Groups of variables in recent papers of residential choice models. Source: Schirmer *et al.* (2012)

## 2. Focus of the study: Zurich

Zurich is the largest city in Switzerland and the capital of the Canton of Zurich. The city of Zurich has a population of 376,008 (2011) and measures 87.88 km$^2$; the Canton of Zurich has a population of 1,390,124 (2011) and measures 1,729 km$^2$.

The primary data source used in this paper is a revealed preference survey conducted in August and September 2010 in the Canton of Zurich. Recently moved households were asked about their previous residential locations, household attributes, workplaces of all workers in the household, attributes of their residential unit and their social network. In total, 1,039 households responded. As stated in other works using the same data (Belart*,* 2011*,* Schirmer *et al.,* 2011a) the survey presents a representative sample of recently moved households in terms of the statistical distribution of age, sex, income and household size as compared to census data and travel surveys. With this data, very differentiated information can be obtained on the households' characteristics and the residential units that were chosen.

As non-chosen alternatives, we used 3,890 offers that were placed on an online real estate portal and were available during the period of the survey. As these offers represent asking prices, the prices had to be scaled to fit to the survey data. This process was described in detail by Belart (2011) and no further comments will be made here. Although sales data is available in the survey, we only considered the rental market.

Both data sets have been geo-coded and enriched with attributes of the environment. These attributes were derived from census data on population and enterprises, building data from the cantonal building register and the cantonal assurances and cadastral information, which have been merged in the context of the project SustainCity (Schirmer *et al.,* 2011b).

Based on this data, we have a highly disaggregated representation of the urban environment that allows us to model the residential location choice of households and to consider the individual persons within a household.

## 3. Measuring residential location preferences

Variables measuring residential location choice preferences can be classified into six categories (Schirmer *et al.,* 2012): variables describing the household, the residential unit, the socio-economic environment, the built environment, accessibility and points of interest. The latter four categories provide four different perceptions of the household regarding its environment. *Socio-economic environment variables* represent the non-fixed configuration of the urban landscape, i.e. they quantify the "soft" factors of the built environment. The *built environment* is represented by the geometries and volumes of spatial objects around a chosen residential unit and includes buildings, parcels, blocks and connecting networks. *Points of interest* (POIs) form the distribution of functions with relevance for the public, which can be introduced by urban planners and policy-makers or form a reaction of the market. In general terms, *accessibility* is a measurement of the spatial distribution of activities around a point, adjusted for the ability and desire of people or firms to overcome this spatial separation (Hansen, 1959).

### 3.1. Characteristics of the household

Variables describing the household are commonly included in residential location models. These socio-economic characteristics are interacted with other variables in the utility function in order to estimate taste preferences across different household segments. It is common to include disposable income after housing costs (e.g. Lee and Waddell, 2010a,b) or differences between household and zonal income to estimate a segregation effect (Pinjari *et al.*, 2011). In general, it was found that households with a higher income tend to commute longer or are insensitive to commuting while large households, as well as households with children, tend to move less often (Lee and Waddell, 2010b, Eluru *et al.*, 2009).

Life cycles influence the relocation probability and the residential location choice (Eluru *et al.*, 2009, Kim *et al.*, 2005, Lee and Waddell, 2010a). Relocation tends to be towards the city center in early life stages and away from the city in later stages. This is captured by the distance to the CBD (central business district) and through interaction with urban characteristics, such as densities. An alternative approach would be to include lifestyles in order to respect self-selection effects (Krizek and Waddell, 2002). Figure 2 depicts the Euclidean distance to the Zurich CBD for households, differentiated by the age of the head of household. It can be seen that households with a younger head tend to live closer to the CBD. As comparable definitions of lifestyles are not available and as attitudes change over time and are partly correlated with life cycles, we only implement life cycles in our models.

Few studies (de Palma *et al.*, 2005, 2007, Zondag and Pieters, 2005, Axhausen *et al.*, 2004) include the distance to the previous location as an explanatory variable, while other studies include the social network (Vyvere *et al.*, 1998, Belart, 2011). Figure 2b depicts the distance to previous residential location versus the distance to social contacts of recent movers. It can be seen that households tend to stay in the vicinity of their previous location and that the location of a household tends to correlate with the location of social contacts.

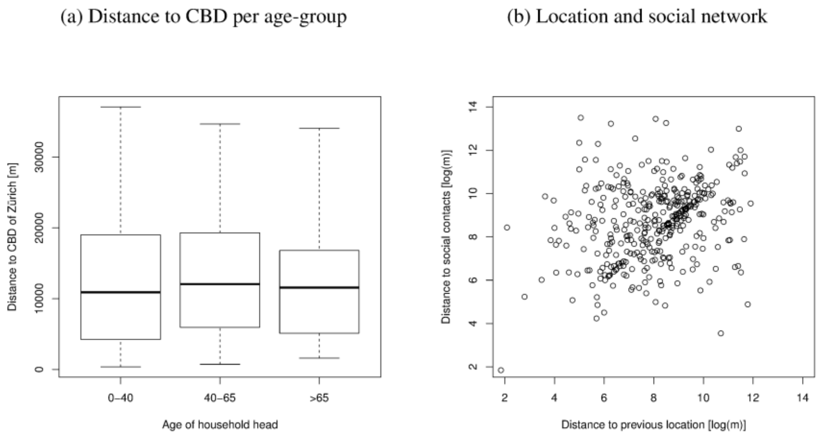


**Fig 2.** Plots on household behavior

### 3.2. Characteristics of the residential unit

Due to data limitations, a common approach is to use a zone as a residential alternative (e.g. Axhausen *et al.,* 2004, Chen *et al.,* 2008, de Palma *et al.,* 2007, Guo and Bhat, 2007, Pinjari *et al.,* 2011), while a few recent studies consider the building as an alternative (e.g. Lee and Waddell, 2010a,b, Habib and Miller, 2009, Vyvere *et al.,* 1998).

If the residential unit is considered as the choice alternative, it is possible to interact household attributes with attributes of the residential unit, e.g. size, price and number of rooms. Various studies observe a preference for larger units or a gain in number of rooms (e.g. Eliasson, 2010, Habib and Miller, 2009). Eliasson (2010) shows that a high number of rooms per person is expected to be favored by families, while singles tend to prefer larger rooms.

Price is either included as an untransformed variable (Kim *et al.,* 2005, Vyvere *et al.,* 1998), a logarithmic transformation (de Palma *et al.,* 2005, 2007, Habib and Miller, 2009, Lee and Waddell, 2010a,b) or a ratio with the household income (e.g. Habib and Miller, 2009). Figure 3a shows the rental price versus the household income per month. It can be seen that rent increases when income increases, so that the interaction of these two variables is recommended.

The type of residence has also been reported to be of relevance; singles and retired persons prefer a multifamily house, while families prefer single family houses (Lee and Waddell, 2010a,b). Figure 3b shows household composition versus building type in the Canton of Zurich. We can partly observe these preferences here, but did not include it in the modeling as this information was not available for the alternatives.

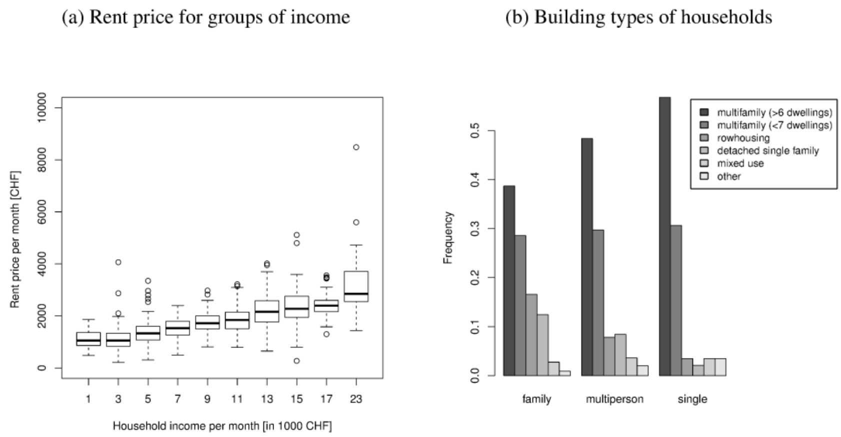


**Fig 3.** Plots on residential units

## 3.3. Socio-economic environment

Variables describing the socio-economic environment are considered in most studies as they are commonly available from census data and municipal statistics. Attributes that have often been used to characterize a location are: population density, aggregated household statistics (size, age, income, origin, children, workers) and employment rates.

A general observation of this group of measurements is that households tend to cluster around similar households in terms of age, income, size, education and ethnic background (de Palma *et al.*, 2005, 2007, Guo and Bhat, 2007, Pinjari *et al.*, 2011, Weisbrod *et al.*, 1980, Zondag and Pieters, 2005). However, the causal explanation of this behavior remains unclear and demands further research in order to understand whether similar households tend to group or whether their preferences for the same location results in a grouping, commonly referred to as self-selection effect. In addition, the planned mix of household types, e.g. through the construction of social housing or a predefined mix of apartments can influence this distribution.

Including the percentage of similar households in an area as an explaining variable improves model estimation results (e.g. Pinjari *et al.,* 2011*,* Weisbrod *et al.,* 1980*,* de Palma *et al.,* 2005*,* Guo and Bhat*,* 2007).

Being a measure of urban centrality, population density is expected to be perceived as negative by most households, except young households and singles (Guo and Bhat*,* 2007*,* Zondag and Pieters*,* 2005). This also holds true for the Zurich area. In Figure 4, it can be seen that singles and families with a young household prefer areas with a higher population density. We thus aim to include the interaction between life cycle and population density.

Guo and Bhat (2007) compare the socio-economic variables with the measurements on the built environment and state that their scales have a smaller spatial extent, which we do not further explore here. Instead, to allow for comparability of the location attributes, we define the spatial extent as the walking distance with a range of 300 m.

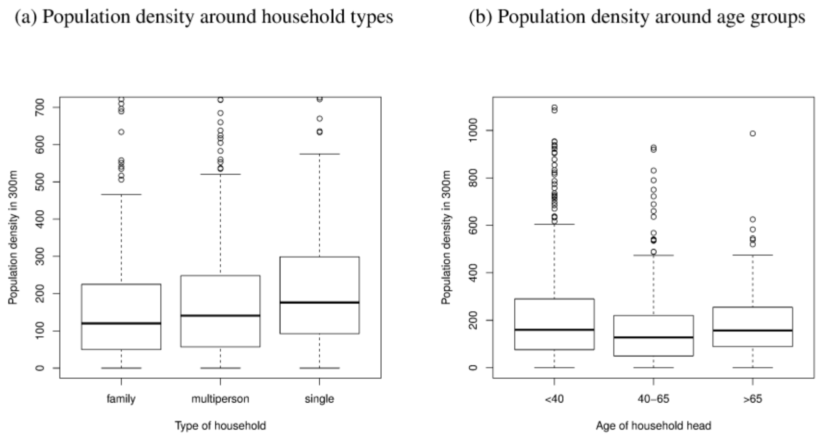


**Fig. 4.** Plots on the socioeconomic environment

### 3.4. Built environment

Built environment measures typically include land use, the share of open space, structural density, built density, network buffers and settlement areas. Truly typological measures have not been included, except for studies considering the number of single and multi-family buildings (Lee and Waddell*,* 2010b*,* a).

Dwelling density represents an urban characteristic; we expect this to correlate with population density. Dwelling density is not available for the Zurich area, therefore, we include building density as a proxy for the density of residential units. However, as can be seen in Figure 5, the correlation between population density and building density is only high for low building densities. High population densities result in larger buildings, not in more buildings.

Guo and Bhat (2007), among others, state that mixed land use is valued by young households and persons without cars, which corresponds to their preference for an urban environment. Due to missing data, we cannot model this variable as a spatial value fraction, but use the number of jobs within a radius of 300 m instead, which can be assumed to be strongly correlated with mixed land use. The number of retail jobs, respectively, service jobs versus the age of head of household for the Zurich area can be seen in Figure 6.

Proximity to the transport network provides accessibility but also generates noise and particle emissions. Contradictive results can be found in several studies (e.g. Vyvere *et al.,* 1998*,* de Palma *et al.,* 2005), depending on the additional attributes included in the model. We therefore differentiate between these two aspects by incorporating the distance to highway access points (see points of interest) and a network buffer on railways and arterials, expecting the former to have negative influence and the latter to be of value for car-owners.

In various studies, a preference was observed for proximity to open space and a high share of open space, however, the calculation method was not explained (Habib and Miller*,* 2009*,* Chen *et al.,* 2008*,* Zondag and Pieters*,* 2005). We include the distance to recreation areas measuring more than two square kilometers.

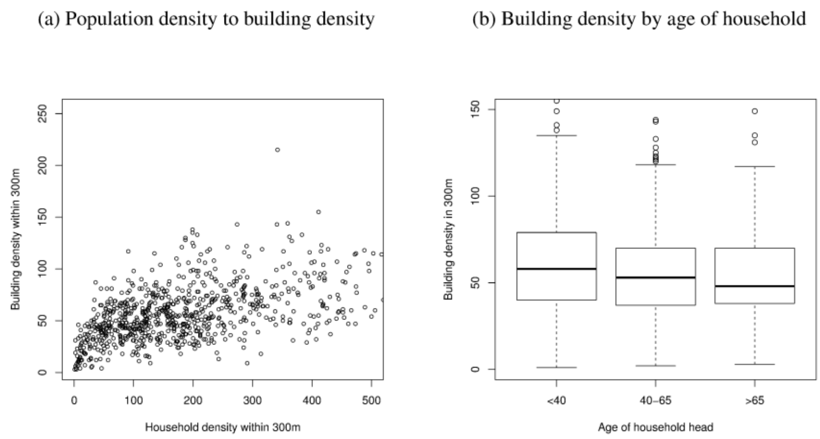


**Fig. 5.** Plots on the built environment

### 3.5. Points of interest

Points of interest (POI) can be classified into categories of education, transport, retail, services and urban centers (Schirmer *et al.,* 2012). Either proximity or the number of POIs within a certain range is considered. Educational opportunities are included in both forms and are valued positively (e.g. Axhausen *et al.,* 2004, Vyvere *et al.,* 1998). We include the distance to the nearest school, assuming that the density of schools is not relevant on a disaggregated level of residential location choice. The diverse observations on school quality (e.g. Chen *et al.,* 2008, Kim *et al.,* 2005, Weisbrod *et al.,* 1980) are expected to reflect the social structure of an area and the educational background of the parents. As no common method of evaluation for school quality is available for the Canton of Zurich, this is not included in our location choice models.

We include the proximity to service and retail using the density of jobs within a walking distance of 300 m. Previous studies have shown that density of services and proximity to retail is favored by all population groups (Vyvere *et al.,* 1998, Zondag and Pieters, 2005, Lee and Waddell, 2010a); thus far, no further differentiation of household types has been considered.

Transportation facilities form nodes and hubs of the transport network in the urban landscape and have been shown to have a significant impact on residential location choice (de Palma *et al.,* 2005, Vyvere *et al.,* 1998, Habib and Miller, 2009). According to de Palma *et al.* (2005), the differentiation of the facilities into noisy train stations and access-oriented subways should have a rather negative and a positive impact, respectively. As the Canton of Zurich lacks a subway network, trains are a major transport facility for commuting. We thus include the Euclidean distance to a station in our models and expect this to be valued by households not owning a car.

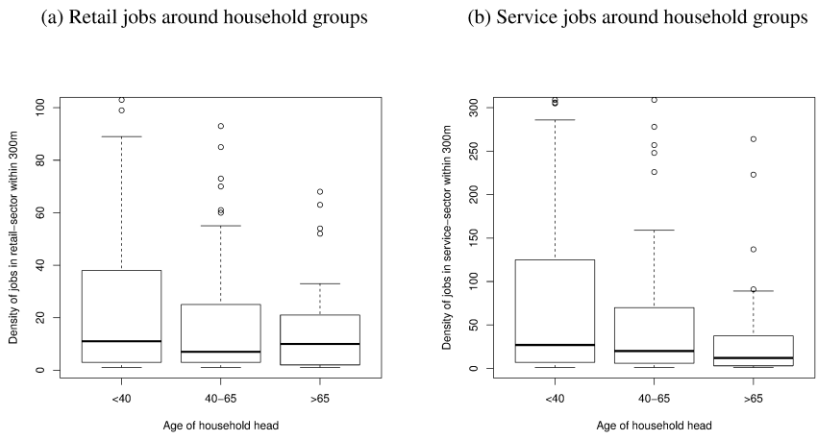


**Fig. 6.** Plots on the points of interest

Proximity to urban centers has often been stated as a relevant variable (de Palma *et al.,* 2005*,* 2007*,* Kim *et al.,* 2005*,* Axhausen *et al.,* 2004*,* Belart*,* 2011). Although we expect that urban characteristics can be represented by other variables in a more generic and reproducible way, we include the distance to the central business districts (CBD) of Zurich and Winterthur in our models for comparability. In accordance with previous studies on Zurich (Belart*,* 2011), we manually defined "Bürkliplatz" as the CBD for Zurich and the "Bahnhofplatz" as the CBD of Winterthur.

Based on the observed tendency of young households to move toward the city and families to move out of the city, we differentiate those two groups of households in our variable specification.

## 3.6. Accessibility

Accessibility can include commuting time, cost or distance to work, accessibility to employment or recreational opportunities. Some studies rely on an aggregated zonal accessibility measure (Guo and Bhat*,* 2007). Eliasson (2010) considers the direct utility from the optimal activity pattern. Besides accessibility to employment, he considers accessibility to services and shops and concludes that it is necessary to include them. Chen *et al.*

(2008) include accessibility to open space and finds that this is positive for households with children. However, the role of accessibility is smaller than that of other factors, such as income and other household related factors.

Commute times by transit are found to be more important than commute times by private transport (de Palma *et al.,* 2007). Pinjari *et al.* (2011) found heterogeneity between households towards commuting times. Chen *et al.* (2008) found that preferences in commuting time depend on the experience with commuting time from the previous residential location; households with commuting experience don't mind commuting in their next location. Habib and Miller (2009) found that households consider travel costs and level of service more important than commute travel time.

In our model, we include commuting distance as a Euclidean distance to the workplace for every person in the household. The accessibility is included as a distance-weighted logsum of jobs, using travel time by car or transit for the distance function.

**Table 1.** Household attributes, location attributes, attributes of the residential unit and their interactions used for the modeling approach

| Classification | Specification in initial model | Function | Radius | Expected sign** | Mean* | Median* | s.d.* | Name in models |
|---|---|---|---|---|---|---|---|---|
| **HOUSEHOLD** | | | | | | | | |
| Commuting time | euclidean distance to workplace ^ eta | distance | | - | 43815.86 | 3251.59 | 542495.90 | HH_DIST_WORK, HH_ETA_WORK |
| Distance to previous location | euclidean distance to previous location ^ eta | distance | | - | 29418.60 | 3626.60 | 385233.10 | HH_DIST_PREVLOC, HH_ETA_PREVLOC |
| **ACCESSIBILITY** | | | | | | | | |
| Car | accessiblity of jobs by car * household has a car | accessibility | | + | 9.07 | 9.06 | 0.61 | LA_MIVACC_CAR |
| Public transport | accessiblity of jobs by public transport * household has no car | accessibility | | + | 11.15 | 11.25 | 1.09 | LA_PTACC_NOCAR |
| **BUILT ENVIRONMENT** | | | | | | | | |
| Built density | density of buildings | density | 300 | + | 53.10 | 51.00 | 34.03 | LB_BUILDG_DENS |
| | density of buildings * is a family (nonsingle household with children) | | | | 52.06 | 52.00 | 33.43 | LB_BUILDG_DENS_x_FAMILY |
| | | density | 300 | - | | | | |
| | denisty of buildings * is a young household (age < 40) | density | 300 | + | 56.42 | 54.00 | 36.58 | LB_BUILDG_DENS_x_YOUNG |
| Open space | open space within 2km radius | spatial ratio | 2000 | + | 5806.19 | 6494.22 | 2483.30 | LB_OPENSPACE |
| | distance to lake (lakesize > 1km²) | distance | | - | 6032.54 | 3866.97 | 5543.24 | LB_LAKE_DIST |
| Network/noise | is within buffer on arterials (50m) and railways (100m) (dummy) | buffer | 50/100 | - | | | | LB_NETWORK_BUF |
| **POINTS OF INTEREST** | | | | | | | | |
| Education | distance to school | distance | | - | 446.98 | 359.93 | 357.36 | LP_SCHOOL_DIST |
| Service and retail | density of jobs in retail ( NOGA-Code G47) | density | 300 | + | 16.46 | 1.00 | 56.91 | LP_RETAIL_DENS |
| | density of jobs on service (NOGA-Codes K to N) | density | 300 | + | 61.10 | 5.00 | 206.60 | LP_SERVICE_DENS |
| Transport | distance to station * household has no car | distance | | - | 723.72 | 651.92 | 450.85 | LP_STATION_DIST_x_NOCAR |
| | distance to highway-access * household has a car | distance | | - | 2333.56 | 1858.27 | 1696.85 | LP_HIGHWAY_ACCESS_DIST_x_CAR |
| Urban center | distance to CBD Zürich * is a young household (age < 40) | distance | | - | 11743.23 | 10931.44 | 7971.44 | LP_CBD_ZH_DIST_x_YOUNG |
| | distance to CBD Winterthur * is a young household (age < 40) | distance | | - | 17588.70 | 19469.00 | 8419.60 | LP_CBD_WIN_DIST_x_YOUNG |
| **SOCIOECNONOMIC STRUCTURE** | | | | | | | | |
| Population density | density of population | density | 300 | - | 192.17 | 147.00 | 185.79 | LS_POP_DENS |
| | density of population * is a young household | density | 300 | + | 213.43 | 160.00 | 204.62 | LS_POP_DENS_x_YOUNG |
| | density of population * is a single household | density | 300 | + | 223.24 | 176.00 | 198.79 | LS_POP_DENS_x_SINGLE |
| Household type | share of households with same age (3 groups) | spatial ratio | 300 | + | 0.38 | 0.40 | 0.13 | LS_SAME_HH_AGE_SHARE |
| | share of households with same size (7 groups) | spatial ratio | 1km | + | 0.04 | 0.04 | 0.01 | LS_SAME_HH_SIZE_SHARE |
| | share of households with children * household has children | spatial ratio | 300 | + | 0.60 | 0.61 | 0.14 | LS_HH_CHILDREN_SHARE_x_CHILD |
| | share of households with same nationality (swiss/nonswiss) | spatial ratio | 300 | + | 0.74 | 0.81 | 0.21 | LS_SAME_HH_ETHNIC_SHARE |
| **RESIDENTIAL UNIT** | | | | | | | | |
| Size | sqm per room | value | | + | 28.32 | 27.00 | 9.13 | RU_SQM_ROOM |
| | rooms per person * is not a single household | value | | + | 1.60 | 1.50 | 0.53 | RU_ROOMS_PERSON_x_NONSINGLE |
| | rooms per person * is a single household | value | | - | 2.58 | 3.00 | 0.99 | RU_ROOMS_PERSON_x_SINGLE |
| Costs | ratio of rent to income | value | | - | 5.34 | 4.89 | 3.15 | RU_RENT_INCOME_RATIO |
| Age | Log (age of building) | value | | - | 48.00** | 32.00** | 93.29** | RU_LOG_BUILDING_AGE |
| | buidling is a new constructed building (<10 years) | value | | + | 1 | *(dummy)* | *(dummy)* | RU_NEW_BUILDING |

** values stated here are based on the survey data of 2010*
*** expectations based on Schirmer et al (2012)*
**** values are non-logarithmic expression of building age*

## 4. Modeling

### 4.1. Econometric modeling methodology

Within the discrete choice framework, a decision-maker chooses from a set of alternatives. Each alternative is assumed to have a number of attributes; each attribute has a level of utility or disutility, which captures the costs and benefits of an alternative. The utility $U$ of an alternative $i$ for a decision-maker $q$ is defined by:

$$U_{iq} = V_{iq} + \varepsilon_{iq} = f(\beta_i x_{iq} + \varepsilon_{iq}) \quad (1)$$

with a deterministic part $V_{iq}$ that consists of a function $f$ of the vector $\beta_i$ of taste parameters and the vector $x_{iq}$ of attributes of the alternative, the decision-maker and the choice situation. The non-deterministic and non-observable part of the utility function is captured by $\varepsilon_{iq}$ . The decision-maker $q$ will choose the alternative from set $C$ with the highest utility:

$$P(i|C_q) = P[U_{iq} \geq U_{jq} \forall j \in C_q] = P[U_{iq} max_{j \in C_q} U_{jq}] \quad (2)$$

The most commonly used discrete choice model is the multinomial logit model (MNL), due to its ease of estimation and simple mathematical structure. It is based on the assumption that random terms, often called error terms or disturbances, are identically and independently Gumbel distributed (i.i.d.). The choice probability of each alternative can be calculated as:

$$P(i|C_q) = \frac{e_{iq}^{V}}{\sum_j e_{jq}^{V}} \quad (3)$$

Nested logit models are applied when the assumption on independence of irrelevant alternatives (IIA) is expected not to hold and a correlation among the error components is expected. The decision to relocate and the location decision or the choice to own or rent in combination with location choice. Nested models do not contain any assumptions on the temporary sequence of decisions. In this study, nested logit models are not considered as it is assumed that no nested decisions are present. The decision to move has already been taken because we only include observations from recent movers.

Model estimation results can be interpreted and evaluated in several ways. Parameter estimates of the same models estimated on different data sources can only be compared if the parameter estimates of one model are scaled by the ratio of the variance of both models.

### 4.2. Base model

The base model includes only the variables of the household, the residential unit and accessibility (Table 2). The choice set includes 49 randomly sampled alternatives from a total of 3,890 non-chosen alternatives. Effects of the location attributes are observed by comparing different model specifications with the base model.

Including square meters per room instead of square meters per person performs better; in addition, correlation is avoided with the variable “number of rooms per household member”. The parameter estimate has the expected positive sign. Rooms per household member are perceived negatively, which is according to expectation for singles, but not for non-singles. This variable only shows minor differences for singles and non-singles; non-singles prefer more rooms than singles do.

Building age has been specified in both absolute and logarithmic form. Both specifications yield the unexpected effect of older buildings being preferred by households, with a better model fit for the logarithmic transformation. Including a dummy variable for historic buildings only yielded a significant parameter in combination with absolute building age, while the inclusion of a dummy variable for new buildings improves the model for both variants and yields the expected positive sign. We assume that building age refers to preferences for different architectural construction styles, but did not investigate this further.

With an adjusted rho square of above 0.5, this base model performs surprisingly well and clearly shows the necessity of modeling location choice of households by specifying the residential unit as alternative and including the number of rooms, size and price as attributes.

**Table 2.** Initial model of residential location choice and the estimation for different groups of location variables

| MEASUREMENTS | | Mod_BASIC | | Mod_ALL | | Mod_LS | | Mod_LP | | Mod_LB |
|---|---|---|---|---|---|---|---|---|---|---|
| **Household** | | | | | | | | | | |
| HH_DIST_PREVLOC | - | -6.480 ** | - | -7.140 ** | - | -6.910 ** | - | -6.790 ** | - | -6.800 ** |
| HH_DIST_WORK | - | -4.620 | - | -3.330 * | - | -4.930 | - | -3.230 * | - | -4.600 * |
| HH_ETA_PREVLOC | + | 0.173 ** | + | 0.163 ** | + | 0.166 ** | + | 0.169 ** | + | 0.168 ** |
| HH_ETA_WORK | + | 0.131 * | + | 0.199 ** | + | 0.129 * | + | 0.202 ** | + | 0.143 ** |
| **Accessibility** | | | | | | | | | | |
| LA_MIVACC_CAR | - | -0.545 ** | - | -0.410 ** | - | -0.480 ** | - | -0.396 ** | - | -0.384 ** |
| LA_PTACC_NOCAR | + | 0.242 ** | + | 0.505 ** | + | 0.340 ** | + | 0.497 ** | + | 0.401 ** |
| **Built environment** | | | | | | | | | | |
| LB_BUILDG_DENS | | | + | 0.003 | | | | | - | -0.003 |
| LB_BUILDG_DENS_x_FAMILY | | | - | -1.820 | | | | | + | 1.090 |
| *(transformed impact)* | | | - | *-0.002* | | | | | - | *-0.007* |
| LB_BUILDG_DENS_x_YOUNG | | | - | -0.139 | | | | | - | -0.480 |
| *(transformed impact)* | | | + | *0.002* | | | | | - | *-0.002* |
| LB_LAKE_DIST | | | - | 0.000 | | | | | + | 0.000 ** |
| LB_NETWORK_BUF | | | - | -0.349 * | | | | | - | -0.301 * |
| LB_OPENSPACE | | | - | 0.000 | | | | | + | 0.000 * |
| **Points of Interests** | | | | | | | | | | |
| LP_CBD_WIN_DIST | | | - | 0.000 | | | - | 0.000 | | |
| LP_CBD_WIN_DIST_x_YOUNG | | | - | -0.082 | | | + | 0.125 | | |
| *(transformed impact)* | | | - | *0.000* | | | - | *0.000* | | |
| LP_CBD_ZH_DIST | | | + | 0.000 ** | | | + | 0.000 ** | | |
| LP_CBD_ZH_DIST_x_YOUNG | | | + | 0.530 | | | + | 0.278 | | |
| *(transformed impact)* | | | + | *0.000* | | | + | *0.000* | | |
| LP_HIGHWAY_ACCESS_DIST_x_CAR | | | - | 0.000 ** | | | - | 0.000 * | | |
| LP_STATION_DIST_x_NOCAR | | | - | 0.000 | | | - | 0.000 | | |
| LP_RETAIL_DENS | | | - | -0.003 ** | | | - | -0.003 ** | | |
| LP_SCHOOL_DIST | | | + | 0.000 | | | + | 0.000 * | | |
| LP_SERVICE_DENS | | | - | -0.001 * | | | - | -0.001 * | | |
| **Socioeconomic Environment** | | | | | | | | | | |
| LS_HH_CHILDREN_SHARE_x_CHILD | | | + | 0.918 | + | 1.210 * | | | | |
| LS_POP_DENS | | | - | -0.002 * | - | -0.002 ** | | | | |
| LS_POP_DENS_x_SINGLE | | | - | -0.343 | - | -0.508 | | | | |
| *(transformed impact)* | | | - | *-0.001* | - | *-0.001* | | | | |
| LS_POP_DENS_x_YOUNG | | | - | -0.444 | - | -0.268 | | | | |
| *(transformed impact)* | | | - | *-0.001* | - | *-0.001* | | | | |
| LS_SAME_HH_AGE_SHARE | | | + | 0.559 | + | 0.439 | | | | |
| LS_SAME_HH_ETHNIC_SHARE | | | + | 0.039 | + | 0.029 | | | | |
| LS_SAME_HH_SIZE_SHARE | | | - | -0.968 | - | -1.570 | | | | |
| **Residential Unit** | | | | | | | | | | |
| RU_LOG_BUILDING_AGE | + | 0.345 ** | + | 0.363 ** | + | 0.354 ** | + | 0.361 ** | + | 0.356 ** |
| RU_NEW_BUILDING | + | 0.631 ** | + | 0.542 ** | + | 0.523 ** | + | 0.544 ** | + | 0.546 ** |
| RU_RENT_INCOME_RATIO | - | -4.130 ** | - | -3.440 ** | - | -4.030 ** | - | -3.390 ** | - | -3.700 ** |
| RU_ROOMS_PERSON_x_NONSINGLE | - | -0.569 ** | - | -0.703 ** | - | -0.631 ** | - | -0.688 ** | - | -0.659 ** |
| RU_ROOMS_PERSON_x_SINGLE | - | -0.591 ** | - | -0.679 ** | - | -0.619 ** | - | -0.674 ** | - | -0.640 ** |
| RU_SQM_ROOM | + | 0.000 ** | + | 0.000 ** | + | 0.000 ** | + | 0.000 ** | + | 0.000 ** |
| log_init | | -2679.736 | | -2679.736 | | -2679.736 | | -2679.736 | | -2679.736 |
| log_final | | -1296.684 | | -1253.640 | | -1286.481 | | -1260.816 | | -1286.199 |
| $R^2$ | | 0.516 | | 0.532 | | 0.520 | | 0.530 | | 0.520 |
| adj $R^2$ | | 0.512 | | 0.519 | | 0.513 | | 0.522 | | 0.513 |
| final gradient norm | | 0.142 | | 0.275 | | 0.102 | | 0.126 | | 0.063 |
| iterations | | 43 | | 110 | | 77 | | 65 | | 65 |

## 4.3. Models measuring location

The base model is used as a starting point for the comparison of different sets of location variables. Table 2 shows the estimation results when adding the location attributes. Due to the high number of variables and expected correlations, we created dedicated models for each category of location attributes. In the final section of this chapter, these variables are reduced to a set of variables measuring location.

Including all the variables of our review reduces the significance of many location variables, likely due to the low number of observations. The attributes of the base model remain significant at the 95% level and do not change in sign in all models tested. This emphasizes the importance of this initial variable set.

Comparing the dedicated models for each category of location measurements, we find the points of interest (POIs) improve the model estimates most.

The expected preference of younger households and singles for urban environments, in comparison to other households, is modeled by the interaction with building density, population density and distance to the urban centers of Zurich and Winterthur. The interaction is computed with the expression $(1 + \beta_{\text{young}}) * \beta_{\text{all}}$ and mentioned as *transformed impact* in Table 2. Estimation results confirm the hypothesis and show the expected negative sign for population density for general households, while being less negative for young households and singles. The same difference is found for building density with a low level of significance. The distance to the urban center of Zurich is valued more by young households than by general households, which may reflect the higher house prices in the urban center.

A closer look at the data shows that the variable "young households" might not be correctly specified in our model. In line with other studies (Lee and Waddell, 2010b), young households are defined as households with a head younger than 40, and we can expect that mainly young professionals and students are seeking a vibrant urban environment in their vicinity. In future research, younger households will include only persons younger than 30.

The distance to CBD (Zurich) continues to have a significant impact even when other variables on urban characteristics are included, but we do find a correlation between distance to lake and population density. Latter variables are significant when distance to CBD is not included. This indicates that distance to an urban center can also be represented by other urban characteristics that allow computation in a reproducible way (without prior knowledge), even though we did not succeed in reaching the same model quality by replacing it with population density. The negative estimates for retail and service density show the preference of households to live in less urban areas as well.

The differentiation between car owners and non-car owners yields consistent model estimates with regard to accessibility variables and POIs describing the transport network. Access to highway is perceived positively by car-owners, as is the access to public transport by non-car-owners. The accessibility by car does not show the expected positive sign, which is expected to reflect car owners' preferences for less central areas, where car use is an advantage when compared to transit and when ample parking space is available.

Unit price captures many attributes representing POIs and a few attributes describing the built environment. When these spatial variables are included, the negative influence of ratio income to price is significantly reduced.

## 4.4. Definitive model

As discussed in the introduction, we aimed for a model that included a wide range of variables describing residential location choice of households and individuals. We also aimed for a common data model for residential location choice models. Including all variables is not possible due to the correlation between variables.

The unexpected negative influence of proximity to schools is expected to reflect a preference of households without children. However, including an interaction for households with children did not improve model estimations, which leads to the conclusion that school locations are sufficiently dense in Zurich and represent noisy areas or further unobserved urban characteristics. This variable was kept in our final model for this reason, as well as the high significance.

Although building density represents an essential urban characteristic, including this variable as absolute value or its logarithmic expression does not show any significant impact on the residential location choice. As mentioned earlier, using the dwelling density is expected to give better results.

Removing distance to CBD of Zurich and Winterthur was captured by the distance to lake. This became positive with a high t-value after having the prior expected negative sign, due to the proximity of the center of Zurich to the lakeside (our definition of Zurich's CBD is Bürkliplatz, which is just 200 m away from the lake, see section 3.5). We thus kept the distance to CBD in the model and removed all interaction terms, including urban characteristics that had a minor impact (see section 4.3).

By removing the differentiation between rooms per person for singles and non-singles, the model results improved. Attributes representing building density, lake distance and social characteristics were also removed due to their low level of significance.

The analysis resulted in the final model presented in Table 3. To determine the sensitivity of model estimates with regard to the generated choice sets, model estimations were carried out with three different choice sets, each containing 49 alternatives. Model estimations were not carried out with standardized data due to dummy variables and distance variables. These would differ per choice-set or would have to be included unstandardized and thus would not allow for comparison. However, by multiplying the parameter estimates with the median of the observed value in the sample, we get an impression of the weight of each variable in overall utility within an alternative.

Using this approach, some variables have a very high impact, namely, the distance to the previous residential location and workplace, both private and public transport accessibility and the rent income Additionally including the rooms per person, the logarithm of building age and the distance to Zurich's CBD also strongly increase model performance. We recommend the use of these variables as the initial set-up for location choice models.

**Table 3.** Final model of residential location choice

| MEASUREMENTS | MOD_FINAL | | | | | | Average impact (ß*median[choice]) | | | Values of observed choices | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | dat1 | | dat2 | | dat3 | dat1' | dat2' | dat3' | mean* | median* | s.d.* |
| **Household** | | | | | | | | | | | | |
| HH_DIST_PREVLOC | - | -5.150 ** | - | -7.070 ** | - | -8.740 ** | | | | | *(interaction)* | |
| HH_DIST_WORK | + | 6.380 ** | - | -3.220 * | - | -3.880 * | | | | | *(interaction)* | |
| HH_ETA_PREVLOC | + | 0.205 ** | + | 0.163 ** | + | 0.135 ** | | | | | *(interaction)* | |
| HH_ETA_WORK | - | -0.060 ** | + | 0.203 ** | + | 0.166 ** | | | | | *(interaction)* | |
| *(interaction distance to previous location)* | | | | | | | *-42.450* | *-26.892* | *-35.057* | *29418.60* | *3626.60* | *385233.10* |
| *(interaction distance to workplace)* | | | | | | | *3.349* | *-16.627* | *-22.874* | *43815.86* | *3251.59* | *542495.90* |
| **Accessibility** | | | | | | | | | | | | |
| LA_MIVACC_CAR | - | -0.165 | - | -0.302 ** | - | -0.187 | -1.513 | -2.758 | -1.714 | 9.17 | 9.13 | 0.62 |
| LA_PTACC_NOCAR | + | 0.590 ** | + | 0.541 ** | + | 0.547 ** | 6.222 | 5.877 | 5.769 | 10.55 | 10.86 | 2.21 |
| **Built environment** | | | | | | | | | | | | |
| LB_NETWORK_BUF | - | -0.090 | - | -0.304 * | - | -0.247 | -0.090 | 0.075 | -0.247 | 1 | *(dummy)* | *(dummy)* |
| **Points of Interests** | | | | | | | | | | | | |
| LP_CBD_ZH_DIST[1] | + | 0.052 ** | + | 0.067 ** | + | 0.068 ** | 0.629 | 0.760 | 0.824 | 12.10 | 11.43 | 7.83 |
| LP_HIGHWAY_ACCESS_DIST_x_CAR[1] | - | -0.102 ** | - | -0.092 ** | - | -0.088 * | -0.228 | -0.159 | -0.196 | 2.23 | 1.73 | 1.62 |
| LP_STATION_DIST_x_NOCAR[1] | - | -0.325 * | - | -0.230 | - | -0.168 | -0.003 | -0.171 | -0.156 | 0.93 | 0.74 | 0.72 |
| LP_RETAIL_DENS | - | -0.003 ** | - | -0.003 ** | - | -0.002 ** | 0.004 | 0.000 | 0.000 | 0.02 | 0.00 | 0.06 |
| LP_SCHOOL_DIST[1] | + | 0.213 | + | 0.298 ** | + | 0.342 ** | 0.000 | 0.107 | 0.153 | 0.45 | 0.36 | 0.36 |
| LP_SERVICE_DENS | - | 0.000 | - | -0.001 ** | - | -0.001 * | 0.000 | 0.000 | 0.000 | 0.06 | 0.01 | 0.21 |
| **Socioeconomic Environment** | | | | | | | | | | | | |
| LS_SAME_HH_AGE_SHARE | + | 0.792 ** | + | 0.684 ** | + | 0.634 * | 0.304 | 0.271 | 0.243 | 0.38 | 0.40 | 0.13 |
| **Residential Unit** | | | | | | | | | | | | |
| RU_LOG_BUILDING_AGE | + | 0.324 ** | + | 0.360 ** | + | 0.350 ** | 0.981 | 1.248 | 1.059 | 3.03 | 3.47 | 1.43 |
| RU_NEW_BUILDING | + | 0.614 ** | + | 0.578 ** | + | 0.634 ** | 0.614 | 0.578 | 0.634 | 1 | *(dummy)* | *(dummy)* |
| RU_RENT_INCOME_RATIO | - | -3.070 ** | - | -3.400 ** | - | -3.580 ** | -16.391 | -16.609 | -19.114 | 5.34 | 4.89 | 3.15 |
| RU_ROOMS_PERSON | - | -0.638 ** | - | -0.677 ** | - | -0.637 ** | -1.206 | -1.124 | -1.204 | 1.89 | 1.66 | 0.83 |
| RU_SQM_ROOM | + | 0.000 ** | + | 0.000 ** | + | 0.000 ** | 0.000 | 0.000 | 0.000 | 0.03 | 0.03 | 0.01 |
| log_init | | -2679.736 | | -2679.736 | | -2679.736 | | | | | | |
| log_final | | -1343.573 | | -1261.079 | | -1276.291 | | | | | | |
| $R^2$ | | 0.499 | | 0.529 | | 0.524 | | | | | | |
| adj $R^2$ | | 0.492 | | 0.522 | | 0.517 | | | | | | |

*** 5%-Significance*
** 10%-Significance*
*[1] Variable has been scaled from m to km*

## 5. Conclusions

In this paper, we estimate the residential location choice of households in order to evaluate the impact of different location attributes representing the neighborhood. The households consist of individual persons and their characteristics, while the residential location is represented through the residential unit with information on size and price. The neighborhood of the location is characterized by variables on the social environment, the built environment and points of interest, as proposed by Schirmer et al. (2012).

The model estimates show a very good model fit for the reference model that only includes household-specific interactions with the residential unit, the distance to individuals' workplaces and accessibilities (distance-weighted logsum of jobs). In this simple model, the location of the residence is only captured by accessibilities and price. These results show the benefit of modeling household location choice as individual persons in a household seeking a residential unit with a given size and price, and we highly recommend it for use in land-use simulations.

Our final model points out that a residence and its location is mainly rated by individuals on a set of five variables: the distance to the previous residential location, the distance to workplace, the ratio of rent price to income and the accessibility of jobs. Distance to previous location has rarely been reported in other studies. The survey data show the correlation with the household's social network for the area of Zurich under study, and we also assume it captures preferences of the household in the form of self-selection effects. This explains its relatively high negative influence.

The three groups of location attributes provide rather surprising results. Although many studies report on clustering of household types, the socio-economic environment does not show a high impact in our household location choice models, except for the share of households of similar age. Meanwhile, many points of interest and some attributes representing the built environment influence the residential location choice. This leads to the recommendation that a planning support system, incorporating a residential location choice model, should include points of interest and the built environment rather than solely modeling the social environment.

However, it should be mentioned that this finding might be specific to Zurich, where we did not observe clustering of migrants (non-Swiss persons) and which has a relatively homogeneous mix of housing types across the study area. Social grouping in terms of age, ethnic background or household size did not lead to any significant impact in our study, however, we did not extend the study to subgroups of foreign persons. Other study areas might have a different social structure, areas of social housing, gated communities or other homogeneous areas that could lead to different results.

Many studies have stated that young persons are attracted by urban centers, while families and elderly persons rather dislike these. We could partly observe this behavior in our models, but find the difference to be of minor importance for our study area. We expect the high connectivity within the Canton of Zurich, the polycentric structure and the relatively small extent of the city center of Zurich have reduced this effect.

Distance to the CBD was a variable that could not be captured by other attributes describing urban characteristics, e.g. population density or the density of jobs, and is not objectively reproducible. Further research will have to explore alternative ways to find center structures within an urban area in a reproducible way, also accounting for local centers.

This paper explored the validity of a basic set of attributes to be used in residential location choice models as proposed in our earlier review (Schirmer *et al.,* 2012). We assume that these variables also work for other study areas and attempted to include only those attributes that can be reproduced in other study areas through algorithms.

Although we excluded attributes that are specific to Zurich, a previous comparison with other studies indicates that some of the observations in this study can nonetheless be expected to be specific to our study area. This applies to all variables in the models and demands a comparison of their statistical distributions with other study areas.

Additionally, it should be stated that these location attributes only represent a few characteristics of urban spaces. Further research is needed to explore alternative attributes of urban characteristics because these have not been or have rarely been stated previously.

## References

Anas A (1982*) Residential Location Markets and Urban Transportation.* Academic Press, New York.

Axhausen KW, Scott D, König A and Jürgens C (2004) Locations, commitments and activity spaces. In: Schreckenberg M and Selten R (eds) *Human Behaviour and Traffic Networks.* Springer, Berlin, pp 205–230.

Belart B (2011) Wohnstandortwahl im Grossraum Zürich. Master Thesis, IVT, ETH Zurich.

Chen J, Chen C and Timmermans HJP (2008) Accessibility trade-offs in household residential location decisions. *Transportation Research Record* **2077:**71-79.

de Palma A, Motamedi K, Picard N and Waddell P (2005) A model of residential location choice with endogenous housing prices and traffic for the Paris region. *European Transport\Transporti Europei* **31:**67–82.

de Palma A, Picard N and Waddell PA (2007) Discrete choice models with capacity constraints: An empirical analysis of the housing market of the greater Paris region. *Journal of Urban Economics* **62**(2): 204–230.

Eliasson J (2010) The Influence of Accessibility on Residential Location In F. Pagliara, Preston J and Simmonds D (eds) *Residential Location Choice: Models and Applications.* Advances in Spatial Science. Springer, Berlin, pp 137–164.

Eluru N, Sener IN, Bhat CR, Pendyala RM and Axhausen KW (2009) Understanding residential mobility: A joint model of the reason for residential relocation and stay duration. *Transportation Research Record* **2133: 34-74.**

Guo JY (2004) Addressing spatial complexities in residential location choice models. Ph.D. thesis, University of Texas.

Guo JY and Bhat CR (2007) Operationalizing the concept of neighborhood: Application to residential location choice analysis. *Journal of Transport Geography* **15**(1): 31–45.

Habib KMN and Miller EJ (2009) Reference-dependent residential location choice model within a relocation context. *Transportation Research Record* **2133:**92–99.

Hansen W (1959) How Accessibility Shapes Land Use, *Journal of the American Institute of Planners* **25**(2): 73–76.

Kim J, Pagliara F and Preston J (2005) The intention to move and residential location choice behavior. *Urban Studies* **42**(9): 1621-1636.

Krizek KJ and Waddell P (2002) Analysis of lifestyle choices: Neighborhood type, travel patterns, and activity participation. *Transportation Research Record* **1807:**119–128.

Lee BHY and Waddell PA (2010a) Reexamining the influence of work and nonwork accessibility on residential location choices with a microanalytic framework. *Environment and Planning A* **42**(4): 913–930.

Lee BHY and Waddell PA (2010b) Residential mobility and location choice: A nested logit model with sampling of alternatives. *Transportation* **37**(4): 587–601.

McFadden, D. (1978) Modeling the choice of residential location. In: Karlqvist A (ed) *Spatial Interaction Theory and Residential Location*, North-Holland, Amsterdam, 75–96.

Pinjari, A. R., R. M. Pendyala, C. R. Bhat and P. A. Waddell (2011) Modeling the choice continuum: an integrated model of residential location, auto ownership, bicycle ownership, and commute tour mode choice decisions. *Transportation* **38**(6): 933–958.

Schirmer P, Belart B and Axhausen KW (2011a) Location Choice in the Greater Zurich Area – an Intermediate Report. 11th Swiss Transport Research Conference (STRC), Ascona, May 2011.

Schirmer P, van Eggermond MAB and Axhausen KW (2012) Reviewing measurements in residential location choice models. *13th International Conference on Travel Behaviour Research (IATBR)*, Toronto, July 2012.

Schirmer P, Zöllig C, Müller K, Bodenmann B and Axhausen KW (2011b) The Zurich case study of UrbanSim. 51st European Congress of the Regional Science Association (ERSA), Barcelona, September 2011.

Vyvere Y, Oppewal H and Timmermans HJP (1998) The validity of hierarchical information integration choice experiments to model residential preference and choice. *Geographical Analysis* **30**(3): 254–272.

Weisbrod G, Ben-Akiva ME and Lerman SR (1980) Tradeoffs in residential location decisions: transportation versus other factors. *Transport Policy and Decision Making* **1** (1).

Zondag B and M Pieters (2005) Influence of accessibility on residential location choice. *Transportation Research Record* **1902:**63–70.