Reviewing measurements in residential location choice models

Patrick Schirmer, IVT ETH Zurich
Michael A.B. van Eggermond, FCL Singapore ETH Centre
Kay W. Axhausen, IVT ETH Zurich
Reviewing measurements in residential location choice models

Patrick Schirmer
Institute for Transport Planning and Systems (IVT)
ETH Zurich
Wolfgang-Pauli-Strasse 15
8093 Zurich
Phone: +41 44 633 62 58
E-mail: patrick.schirmer@ivt.baug.ethz.ch

Michael A.B. van Eggermond
Future Cities Laboratory (FCL)
Singapore ETH Centre
1 CREATE Way
#06-01 CREATE Tower
Singapore 138602
E-mail: eggermond@ivt.baug.ethz.ch

Kay W. Axhausen
Institute for Transport Planning and Systems (IVT)
ETH Zurich
Wolfgang-Pauli-Strasse 15
8093 Zurich
Phone: +41-44-633 39 43
E-mail: axhausen@ivt.baug.ethz.ch

Abstract

With the increase of the availability of disaggregated data, the modelling of residential location choices on a disaggregated level, has recently gained more and more popularity. All of these studies report on a variety of methodological approaches, but ignore a systematic selection of the variables to include.

Within the project SustainCity (www.SustainCity.org) an integrated land use transport simulation environment is being set up for the study areas of Paris, Brussels and Zurich. The implementation of location models into this simulation is limited through the data model of the simulation software. In order to understand these limitations we want to explore the objects and attributes that commonly have proved to be of relevance to residential location choice models.

In this paper we explore this topic and compare the variables that have been reported in recent literature of disaggregated residential location choice. The focus shall be an overview of the main findings, to understand common behaviour, and what therefore is necessary to represent in the data structure of urban simulation.

Keywords
residential location choice, household location choice, discrete choice modelling, UrbanSim, SustainCity, land use simulation

Preferred Citation
1. Introduction

The result of individual residential location choices made by households influences, amongst others, employment conditions, economic vitality of a neighbourhood, traffic conditions and social segregation. Understanding residential location choice is thus of relevance for urban planners, policy-makers and researchers.

Long-term decision such as residential mobility and location choice are closely interrelated with short-term travel decisions. Researchers and urban planners therefore have developed integrated transport land-use models in order to model this interaction between land use and transport. One of the first spatial-interaction models to have been developed was the Pittsburgh model by Lowry (1964). Subsequently, a series of land use models has been developed. Comprehensive reviews can be found in Wegener (2004) and Iacono et al. (2008).

Within the EU project “SustainCity” (www.sustaincity.org) such an integrated the land use transportation model for three case-studies, one being the Canton of Zurich, is being implemented. The open source software package UrbanSim (www.urbansim.org) is used in conjunction with the activity-based transportation model (www.matsim.org).

UrbanSim (Waddell et al., 2003) is an open-source software, simulating land use development in cities based on the choices of households, businesses, land owners and developers, interacting in urban real estate markets. A previous implementation of UrbanSim has been performed on the Greater Zurich Area in 2007 (Löchl et al., 2007) within the project “The future of urbanized landscapes” (Zukunft Urbaner Kulturlandschaften, ZUK). This initial model used the grid cell as spatial resolution for its choice models; the latest version of the UrbanSim allows simulating at the parcel level.

The increase in level of resolution leads to a number of challenges, one of them being the modelling of residential location choice. First, there is the increase in number of alternatives faced by individuals. Also, it is possible to include a more detailed specification of the utility function. Most studies investigating residential location choices agree on importance of income and other household related factors on residential location. However, with the increase of level of resolution, in residential choice model, confounding parameter estimates have been discussed Zondag and Pieters (2005) and Molin et Al (2003) find the role of accessibility modest, whereas Lee and Wadell (2010) find accessibility to play a role measured with a time-space prism. The question arises which factors need to be incorporated in our residential location choice model for the Zurich region, which data is available and which data structures need to be added to UrbanSim.

This papers aims to investigate systematically which factors influence residential location choice by means of a comprehensive literature review. The purpose of the literature review is to classify used variables according to their use in simulation process, find common attributes between studies that proved to be relevant, understand the various interaction terms and
summarize the main findings in literature. Eventually, this analysis of variables will result in a
default set-up for our residential location models, making comparison between regions more
convenient.

The following section further discusses the modelling of residential relocation and focusses
on discrete choice methodology and choice set formation. Section 3 continues with a
systematic comparison of attributes considered in a wide range of residential location
choice studies. Section 4 concludes with a summary and outlook.
2. Modelling residential location choice

The roots of residential location modelling can be traced back to the first advances in land-use modelling by Von Thuenen (1826). Von Thuenen made the first attempt to explain the effect of transport costs on the location of activities and the functioning of the land market by means of a single-market in an agricultural region where land-owners are willing to rent their properties to the highest bidder: the bid-rent concept. Alonso (1964) applies this bid-rent concept to residential location and considers a monocentric city with employment opportunities. Individuals and households base their residential location choice of households on maximizing a utility function that depends on the expenditure in goods, size of the land lot, and distance to the city centre. Parallel to Alonso, Lowry (1964) applies the gravity model for residential location. Lowry assumes an initial set of basic employment centres per zone. Households are allocated to zones based on a deterrence function describing the number of workers employed in zone \( i \) and living in zone \( j \). Residential attractiveness is measured by the amount of land available for development in a particular zone.

The discrete choice framework was first introduced to residential location choice by McFadden (1978). Initial studies considered households which move to a certain zone (Weisbrod et al., 1980; Anas, 1982). Each zone was attributed characteristics, such as housing price, employment level, crime rate and accessibility to other zones or employment. Households select the location that maximizes their utility, with prices being determined exogenously through a hedonic model or by given prices. The choice approach with hedonic prices is valid under the rather stringent assumption that prices are properly estimated under equilibrium approaches. To overcome this assumption Hurtubia and Bierlaire (2011) propose a model that is able to account for the bid-rent process and location process simultaneously in a micro-simulation context.

The remainder of this section will discuss discrete choice models and their application to residential location choice and two specific aspects of discrete choice models: the utility function and choice set formation. Section 3 will continue with specific attributes considered in a wide range of residential location studies.

2.1. Discrete Choice Models

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 capture the costs and benefits of an alternative; the utility \( U \) of an alternative \( i \) for a decision-maker \( q \) is defined by:

\[
U_{iq} = V_{iq} + \epsilon_{iq} = f(\beta_i x_{iq} + \epsilon_{iq})
\]

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. In addition, socio-demographic attributes of decision-maker $q$ can be included in the deterministic part of the utility function. The non-deterministic, non-observable part of the utility function is captured by $\varepsilon_{iq}$. Four types of errors can be recognized: unobserved alternative attributes unobserved individual characteristics, measurement errors and proxy variables. In order to reflect this fact uncertainty is modeled as a random variable.

Decision-maker $q$ will choose the alternative from set $C$ with the highest utility:

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

The most commonly used discrete choice model is the Multinomial Logit (MNL) model due to its ease of estimation and simple mathematical structure (McFadden 1974). It is based on the assumption that the random terms, often called error terms or disturbances, are identically and independently (i.i.d.) Gumbel distributed. The choice probability of each alternative can be calculated as:

$$P(i \mid C_q) = \frac{e^{V_i}}{\sum_j e^{V_j}}$$

The Independence of Irrelevant Alternatives (IIA) property states that the ratio of the choice probabilities of any two alternatives is entirely unaffected by the systematic utilities of any other alternatives (Ben-Akiva & Lerman 1985). This property stems from the fact that the distribution of the disturbances are assumed to be mutually independent and requires that the sources of errors contributing to the disturbances do so in a way that the total disturbances are independent. The IIA property can be overcome by applying models that allow for a non-zero covariance matrix or use a nested choice structure. An example of the prior are Mixed Logit models, an example of the latter Nested Logit and Cross-Nested Logit model.

Comparison & interpretation

Model estimation results can be interpreted and evaluated in several ways (Louviere et al, 2000; Train, 2003). Parameter estimates of 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. Also the parameter estimate $\hat{\beta}_k$ of attribute $k$ in expression $V_i$ of alternative $i$ can be interpreted as the weight of the attribute in the utility expression by multiplying $\hat{\beta}_k$ by the mean or median value of the attribute $X_i$.

Discrete choice models can be used to derive estimates of the willingness-to-pay (WTP) or willingness to accept (WTA) of an individual to obtain a benefit or avoid a cost. In a linear model, where each attribute is associated with a single parameter, the ratio of two parameters
is the WTP or WTA holding all other constant. If one of the attributes is measured in monetary units, the ratio can be interpreted as a valuation.

Finally, models can be evaluated by means of the responsiveness of market shares to changes in each attribute.

Application

The type of model applied depends on the choice dimensions considered (i.e. car ownership, residential mobility, residential location). The majority of studies solely considering residential location choice employ the MNL model (e.g. Lee and Waddell, 2010; Guo and Bhat, 2002; Vyvere et al 1998). Other studies apply Mixed Logit models to account for heterogeneity amongst decision-makers (e.g. Eluru, 2009; Habib and Miller, 2009; Zhou and Kockelman 2008). Nested Logit models are applied when multiple dimension are considered, such as residential relocation and location choice (Lee and Waddell, 2010; Andrew and Meen 2006; Eluru et al., 2009), residential location and activity pattern (Eliasson, 2010; Ben-Akiva and Bowman, 1998) and residential location and auto ownership (Weisbrod et al., 1980). Zondag and Pieters (2005) consider the move or stay decision, followed by a nest representing the decision to remain the current region or move to a different region; they find that moving is a local process.

2.2. Utility function and choice set formulation

Utility function

As discussed in the previous section, the utility function of each alternative contains a number of attributes. The relative weight of the parameter estimates for these attributes give insights in the trade-offs decision-makers carry out, such as the trade-off between alternative specific attributes (location) and socio-demographics (income, age, household composition).

Most studies use cross-sectional data to model residential location; the previous location of the household and changes in household demographics are not taken into account. However, evidence can be found that the previous location plays a role in choosing a new location – households who preferred shorter commute distances at their previous location do so at their new location (Chen et al. 2008). Households also prefer a gain in bedrooms and gains in open area (Habib and Miller 2009), which may be led by a change in household composition.

Cascetta et al. (2011) propose a methodology to identify dominance attributes which may be defined in different ways, in accordance with the specific choice context, and in which way they can be introduced as perception attributes in random utility models. Their estimation results show a generally high significance of all these attributes and a considerable improvement in the model’s goodness-of-fit statistics.

Choice set formation

Every choice is made from the set of the alternatives. These alternatives need be mutually exclusive from the decision maker’s perspective. Choosing one of the alternatives implies not choosing any of the other alternatives. Also, the choice set must be exhaustive, in a way that
all possible alternatives are included. Finally, the number of alternatives must be finite (Train, 2003).

Several approaches are discussed in literature to determine the choice set which contains the alternatives that were available to the decision maker as this universal set of alternatives is unknown to the analyst in a revealed preference environment. Swait (2001) proposes to formulate several choice sets (a set of choice sets) and estimate the probability of a choice set being the true choice set. On the other hand an attempt can be made by following heuristics considered by the decision-maker and thus acknowledging that choice set formation is a dynamic search process. Another way to avoid the burden of working with extremely large choice sets is to estimate parameters from a subset from alternatives in the MNL model where the independence of irrelevant alternatives (IIA) is assumed, as demonstrated by McFadden (1978).

Most studies consider the universal choice set or sample from the universal choice set (e.g. Lee and Waddell, 2010b,a; Habib and Miller, 2009; Chen et al., 2008; Guo and Bhat, 2007; de Palma et al., 2007). De Palma et al (2007) compare different sample sizes. Weighting of the sampling is done according to the number of units per zone. Their findings confirm that variances decrease proportionally to sample size, and that significant estimated parameters remain constant. Insignificant estimated parameters, such as number of stations, travel time to highway, however vary in sign and size. The second approach, considering heuristics followed by the decision-maker, has been considered by Zolfaghari et al. (2012). They apply a hazard based model and set thresholds on acceptable property price and commuting times. However, they find that their choice set formation is outperformed by a choice set formed by random sampling. The previously discussed study by Cascetta et al. (2011) also uses dominance variables, incorporate in the sampling technique. In particular, dominance variables have been used as sampling approach. Their results show that the weighted sampling gives parameters’ estimates “closer” to those obtained with full choice set.
3. Measurements of residential location choice

The purpose of this paper is to identify the influence and trade-offs between different groups of variables in the residential location choice in order to understand which type of variables are of relevance for integrated land use transportation simulation. Within this chapter we explore the dependant variables that have been reported in previous studies on residential location choice models and summarize their findings.

Commonly a distinction is made between location attributes, household attributes and attributes of the residence. For our study we will differentiate further groups of location choice models based on the type of initial data and the calculation method.

As described above, estimates of different models do not allow for direct comparison if not using standardized variables or elasticities. Also within a single model the estimates cannot be compared due to different initial data scales. The comparison of elasticities, e.g. to a price variable would allow for comparison, but unfortunately is not being done in most studies. Within this paper we thus can only report the level of significance\(^1\) and the positive or negative influence on the resulting utility of a location, as well as summarize the main-findings of the authors.

The models in the reviewed literature are very divers and use a wide range of variables. Each of the variables might capture some correlation effects to unused variables, which again will be different between the reported models. If these correlated variable are used in another model, it will result in different estimation results. This creates difficulties when trying to discuss on models that have been reported. Nonetheless we need to explore and group the variety of variables that have been used in previous literature to search for common observations, that captures common human behaviour.

3.1. Location attributes

Location attributes depend on the resolution of the choice set alternative (e.g. zone, neighbourhood, unit) and the disaggregation of the initial data. For instance, zonal information will not allow the exploration of local neighbourhood effects. Guo and Bath (2007) explore the perception of the neighbourhood using different spatial extends and calculation methods. They differentiate a variety of scales on fixed neighbourhood boundaries, based on census tracts and blocks, against sliding neighbourhood extends, using Euclidean distance and network distances, and show that the calculation process has an essential influence on the estimation result. Their findings also prove that the scale of a neighbourhood differ for structural and socioeconomic variables, being smaller for the latter.

Beside the level of aggregation also the spatial representation of data varies. On the one side (disaggregated) census data representing socio-economic information is located in space

\(^1\) For our review we will define the commonly used significance-level of 0.05 as statistically significant.
through various individuums or an abstract grid cell and is represented as spatial points. On the other side cadastral information as buildings, parcels and blocks possess a geometric or volumetric shape which occupy spaces and will interact to each other. A specific form of cadastral information are the locations of public functions which ignore the spatial representation and summarise attributes to a spatial point, e.g. the location of a hospital or of a school. Last form of spatial data will be the network of roads and public transport, which is represented through abstraction in form of spatial lines and connects locations in a city².

Software for Urban Simulation as summarized by Wegener (2004) and Zoellig (2011) generally cannot handle all forms of spatial data and mainly uses the socioeconomic characteristics for the simulation process. For our project SustainCity we use the current version of UrbanSim (www.urbansim.org; Waddell et al, 2003). This version does not support geometric features, but represents objects in form of python objects, that are linked to each other by unique ids. Only the objects of parcels will have a defined location as coordinates.

![Data structure of the parcel based version of UrbanSim](image)

Data structure of the parcel based version of UrbanSim

Overview of UrbanSim and the Open Platform for Urban Simulation, presentation, UrbanSim Tutorial, Zurich

This structure reduces the options of spatial evaluations within the simulation. Calculation of build densities on a parcel is possible but the ability to reconfigure the spatial configuration is limited. For example the fraction of open space after creation of new buildings cannot be evaluated, neither does network-based accessibility. This limitation has been recognized by Vanegas et al (2009) who proposes the interaction of geometrical models and behavioural models.

As the purpose of this paper is to explore the types of variables used in residential location choice in order to understand their impact for urban simulation processes we will address this current limitation and differentiate location attributes into attributes representing the built-space (shapes), networks, points of interest and socioeconomic distribution.

---

² The initial form of the road as the width or sideways are seen as part of the geometric objects as parcels in this classification.
Location attributes are characterized through their spatial distribution and require a variety of geographic functions for the data preparation process. Their use with the simulation routine of integrated land use simulation might be limited as they depend on the data model and the methods implemented. As not all of those methods are implemented in urban simulation software, we will further distinguish these processing techniques and group the reviewed variables based on:

- buffer
- distance (Euclidean distance, network distance, travel time, travel cost)
- density (kernel density, number of object in spatial reference)
- spatial ratio (share, fraction of area in spatial reference, fraction of population)
- accessibilities (weighted distance function)

### 3.1.1. Built Space

Built space variables include all characteristics that are derived from the geometries or volumes of spatial object. They include the buildings, parcels, blocks, extend of urban zones and the connecting networks which only can be changed by construction activity. Meinel et al (2009) use these objects to differentiate spatial typologies which they link to socio-demographic distributions. With this approach they are able to impute socio-demographic distributions in a city based on urban shapes. This invokes a dependency of urban shapes and social distribution.. Also Gil et al (2009) and Dillenberger (2010) explore the shapes of objects and find characteristic configurations representing typological features. Sevtsuk (2010) includes variables on the urban form in his location choice models for retail services and can proof that including the build space significantly enhances the model results. As most location choice models are based on census data the use of built space variables derived of cadastral information is not very common though. Within the reviewed literature we can differentiate variables on land-use mix, open-space, structural densities, built densities, network buffers and settlement areas.

**Built densities**

Waddell (2006) is the only one to include dwelling density in his residential location model for the Pudget Sound Region. He finds this to have a negative impact on the residential utility for all family households. The density of dwelling units as logarithmic expression however has a significant and positive influence for all household types tested. As for population density, which will be described later in detail, the residential density has a positive sign for young households. This is not surprising as it is expected to be correlated with population density. This allows for the assumption that build space information can be used as alternative to certain socio-demographic variables when not given.

**Structural densities**

Pinjari et al (2008) attempt to implement structural variables based on the shape and urban form and introduce the length of networks and number of blocks per square mile. The number of bike lanes show a positive impact for all households while the number of blocks has a
negative impact for households with a high income. This is the only study to reports such kind of variables. Nonetheless we understand that these type of measurements can be an interesting approach to represent geometric information and urban characteristics and demand for further research.

Network and noise
Another type of variables representing the built space are network-related attributes. Bürgle (2006) reports that the proximity to major roads or railways has a negative effect on residential utility in the Zürich area. This can be interpreted as an indicator to noise which also is observed by Vyvere et al (1998) within their stated preference survey in Louvain-la-Neuve (Belgium). An opposite influence is reported by Waddell (2006) for Seattle and de Palma (2005) in Paris who observe a positive impact of proximity to arterials and highways. As the study of de Palma is on a quite aggregated zonal level and both reports do not include accessibility calculations, the distance to highway can be expected to capture these effects instead of noise effects. A buffer to noisy networks would thus capture noise and can be expected of having a negative impact on residential location utility.

Fraction of open space
The use of fraction of open space and green area as variables in residential location choice has been reported by several authors, although it is often undefined whether the authors mean recreational areas and or unbuilt area. Habib and Miller (2009) find the percentage of green area in the neighbourhood to be of general positive impact for households in Toronto. Similar results have been found by Chen et al in Seattle (2008) and - for open space - also by Zondag and Pieters in the Netherlands (2005), with both reports testing the variable on various household types. The latter additionally finds a positive impact of percentage of water surface on the residential utility for workers, which reflects the architectural attraction toward waterfronts on the residential housing market in the Netherlands. Guo and Bhat (2007) report a negative impact of open space variables for couple-only households in San Francisco. As these results are specific only to one household-type, we expect the open space to be of common positive impact for all lifecycles and household-types.

Fraction of land use
Also land use mix has been explored in residential location models, but often lack a further specification. A common observation is the negative impact of industrial land use in the neighbourhood (Habib and Miller 2009, Weisbrod et al. 1980), so that we expect mixed land to only differentiate between residential and commercial (office and retail) rather than including industrial land-use. Waddell (2006) includes mixed use environment in his study on Seattle. His models shows a positive impact for young households of all sizes, while Guo and Bhat (2007) also find a positive estimate for persons without cars. The latter also report on a negative influence of the fraction of residential land use in close neighbourhood for all households types, which supports the general preference for mixed land use areas. Meanwhile Pinjari et al (2009) report an opposite effect for the same study area, observing that homogenous regions are favoured when using commercial fraction and land use mix as variables in their residential location choice. They do not differentiate between household types in their study, so that this leads us to expect that mixed land development is valued by
households that favour urban areas, generally young households and households not owning a car, whereas other household types seem to prefer a more homogenous neighbourhoods. A differentiation of household with kids is still outstanding but will be envisioned, as these also dislike urban areas.

**Urban character area**

Several paper include explicit variables on urban characteristics into their location choice models. Andrew and Meen (2006) find a relation to lifecycles in their study on London and show that households tend to move toward the city core when they are young, and later move away from the city. De Palma et al (2005, 2007) show a significant negative value for a dummy variable of the settlement area of Paris. Also Kim et al (2005) implement a variable for city settlement in their stated preference survey on Oxfordshire. Their residential model shows a clear tendency to move out of the city as well. Belart (2011) and Bürgle (2006) define the Central Business District in Zurich as a spatial reference point and report a trend of all households to move away from this one spatial point. A more differentiated approach has been performed by Zondag and Pieters (2005). They distinguish four types of urban characteristics (urban centres, urban neighbourhoods, local village centres and local village green neighbourhoods) and explore their impact on the residential location choice for different household types. Although several of their models report a significant effect of urban characteristics, it is not possible to distinguish general tendencies, which might rely on the classifications of households. The other discussed studies lead us to the assumption that urban characteristics have an effect on residential location, closely related to lifecycles of households. Households tend to move away from the city core during that one. However all these models have the problem of missing reproducibility of the variables representing urban characteristics. It is not clearly defined what it is the extend of the core, or where the centre is placed. The definition of a point or zone as urban core is less convincing as they are not reproducible nor defined. Instead models should investigate to capture these characteristics by other spatial variables, e.g. built density, service density and public transport density.

### 3.1.2. Points of Interest

Points of interests (POIs) represent the location of public use, which can be placed in buildings or outside and have an abstract representation as spatial point. Examples are educational facilities as schools or recreational facilities as sport fields. Certain types of POIs as administrative functions are retail are indicators of centre structures as well. Since they are also not part of the data model used in UrbanSim the placement of new POIs and the distance calculations cannot be performed within the simulation.

POIs that have been reported in residential location choice literature can be grouped into the categories: education, transportation, leisure, retail and service, urban centres. Variables on POIs appear in two forms in our reviewed literature: on the one side as distance from a location in form of network distances, absolute distance or travel costs, on the other side as density of POIs within a certain neighbourhood.
**Education**
A common observation for various study areas is that residents like the proximity of educational facilities. Pinjari et al (2009) mention a positive impact of density of schools in a zone, although not being of high significance. Axhausen (2004) and Vyvere et al (1998) report observe a similar effect, as they report a negative sign for distance to schools. No report has mentioned the use of other educational facilities or the differentiation of scales.

**Service and retail**
Service functions have been explored by Zondag and Pieters (2005) who find their density in a zone to enhance the residential quality for all household types tested. Only non-single household with workers do not show a significant effect on these but react on accessibility of all travel-purposes, which can be assumed to be partly correlated. Lee and Waddell (2010) use the logarithm of jobs in the neighbourhood as variable in their model and prove this to have a positive effect, while Guo and Bhat (2007) find that lower income households and single households are more likely to reside near employment centres when using employment accessibility. We therefore can expect the service density and retail density have a common positive influence on residential choice.

The influence of proximity to retail facilities has only been explored within stated preference surveys for the literature we reviewed. All studies find them to have a positive effect on the residential utility. Vyviere et al (1998) show that the distance to grocery shops is valued when less than 500m and the distance to a shopping centre is appreciated when being under 5 km. Kim et al (2005) find the travel costs (which we can expect to be correlated to distance) to enlarge the probability of moving out of a location and to reduce the utility of a residential location.

**Recreation and sport**
Pinjari et al (2008, 2011) report on the use of variables representing the density of sport and recreational facilities for residential location choice. They find the number of physically active recreation centres to significantly enhance the utility of a location. The number of natural recreational centres has only a minor influence for households with bicycles. No other study we reviewed has used measurements on the sport and recreational facilities, but based on this report and own observations we expect the proximity to sport and recreation facilities to have a positive effect as long as they are not noisy.

**Transportation**
The last group of POIs we found to be mentioned on location choice models are related to transportation. De Palma et al (2005) report that the proximity to subways in Paris is valued, while the proximity to railway stations is disliked by households. They state that this is related to the noisy environment of railways stations and the multiple retail services that group around subway stations. Vyvere et al (1998) report a positive effect of proximity to bus-stations in their stated preference survey in Louvain-la-Neuve. Habib and Miller (2009) report on individual car traffic and find the proximity to highway exits to be of positive effect. We thus expect that proximity to local public transport stops (bus, tram, subway) is of relevance for residents, while long distance transport POIs as railway stations and highway exits might
only be valued by certain groups of persons depending on them and otherwise are avoided due to their noisy environment.

Urban centre

Finally an abstract form of POIs are mentioned by several authors, by defining the location of a central business district (CBD) or center and including it in choice modelling. Axhausen et al (2004) include the distance to Mittelzentrum and Oberzentrum. Their model show the distance to a mayor centre (Oberzentrum) to be valued by residing households. A similar observation is reported by de Palma et al (2005,2007) and Kim et al. (2005) for the urban cores of Paris and Oxfordshire. Also Bürgle (2006) and Belart (2011) report that households tend to move away from the city centre in the region of Zurich. Meanwhile the proximity to smaller centre structures (Mittelzentrum) is reported to have a positive impact on the residential utility by Axhausen et al (2004), leading to the expectation that dense urban areas are generally disliked while local centre structures are valued within residential location choices. A further differentiation of household types should be envisioned as young households tend to favour urban areas. As discussed previously it should also be envisioned to derive the location of centres from the data instead of manually predefine it in unreproducible way.

3.1.3. Socio-economic attributes

Socio-economic measurements are the group of location attributes being reported most for residential location choice. While social aspects are mainly related to population distribution, the economic aspect handles a variety of variables on employment, housing, taxes and the perceived quality of locations. This group of variables thus represent the non-fixed configuration of the urban landscape, which is flexible and constantly changing. Land-use simulation models as UrbanSim have their focus on simulating these dynamics and use data models that can capture those measurements. In their analyse of neighbourhood perception Guo and Bhat (2007) find out that social-economic and demographic composition values have smaller spatial extend of influence than land-use variables.

Many of the socio-economic attributes are defined as interaction terms with characteristics of the households and thus could also be seen as household attributes. For our classification we report on them in this chapter to allow for comparison as we also did for the other location attributes.

Population density

An attribute which is used in most residential choice models is population density. De Palma et al (2005, 2007) use it in three different formulations in their model, when implementing the absolute density, the log of density and the change of population density. The results are the

---

3 These are classifications in the German planning system (Raumordnung) that define the importance of a city and clarify transport connections and administrative institutions, with Oberzentrum being the higher level of both.
same in the two reports, showing the absolute population density to be dominating and being
negative. Also other studies find the population density to have a negative effect for
residential choice (Kim et al., 2005; Lee et al., 2010; Weisbrod et al., 1980). Zondag and
Pieters (2005) differentiate between household types in the Netherlands. They report that
household generally dislike population density, meanwhile single-households with workers
feel attracted by areas of dense population. Guo and Bhat (2007) find a negative impact on
residential utility for small families and high income households but a positive impact for all
the others, including young households and single-person households and thus supports the
previous observations. Bürgle (2006) shows that especially young households are attracted by
areas of high populations density in Zurich. As mentioned previously this type of households
has a general preference for urban areas and proofed to have a tendency for mixed use areas
and a high density of dwelling units as well. Pinjari (2008, 2011) finds a general positive
influence using a logarithmic formulation of population density, but a negative impact for
senior households, and households with children in the untransformed formulation. Only
Zolfagafghi et al (2011) report a positive influence of population density for all households.
Concluding, it can be expected that residential density is generally disliked by households,
especially by families but that certain types of households, namely young households and
single person households value the population density due to their preference for urban areas.

*Household type*

Several authors have used the share or density of same household types as the relocating
household as attributes in their location choice models. De Palma (2005, 2007) shows that
most households in Paris tend to search the proximity to households of same size. Only two
person households dislike this variable but on a low level of significance. Also Guo an Bhat
(2007) find a positive influence on the residential utility. In addition they use the difference of
the household size to the average household size in the zone as variable that proof to have a
negative impact, underlining the previous results. This latter approach also has been
mentioned in other studies (Pinjari et al., 2008, 2011; Weisbrod et al., 1980; Zolfagahari et al.,
2011), all finding a negative impact when household sizes differ. Waddell (2006) uses the total
number of households with a similar size for different household types of age and size
instead. He is the only author who reports a significant effect for elder families (age higher
than 40). It is expected the general antipathy for dense populations is leading to this contrary
observation, as he does not including population density in his model nor the share of same
households (but the total number of households).

Lee and Waddell (Lee et al., 2010; Lee and Waddell, 2010) demonstrate that families like the
proximity to other households with children using the percentage of families as an explaining
variable in their location choice models. Also Bürgle (2006) observes this behaviour using
density of children in the neighbourhood as variable.

Two studies further report that young households tend to cluster as well for the study areas of
Seattle and Paris (Lee and Waddell, 2010; de Palma et al 2005, 2007) The latter study
additionally shows a negative influence of a high density of young households for all others,
which might represent correlation effects to the preference of young households for urban
as explaining variables on residential choice of senior households and find a positive influence as well. Waddell (2006) uses the density of households with same age instead of the fraction. His models only shows a significant positive influence for young single-person households. Similar to his observations on household size, we expect that the integration of a density instead of a percentage, captures effects of population density that has a negative influence.

De Palma et al (2005, 2007) classify the households based on the number of workers and assumes that households also segregate based on this characteristic. In a first study in 2005 they find a significant positive impact, an effect which is not observed in a later study again.

These reports lead us to the assumption that household prefer to locate around households of same household type concerning age, size and the presence of children. It thus is an essential variable for modelling residential location choice and needs to be integrated as interaction terms. The share of households with a same type tends to be valued positive and the difference of the average size to the own size can be expected to be negative.

A question that has not been explored within the reviewed literature and demands further research is to what point socio-economic variables are causal explanatory variables at all. Is it that households of same kind tend to group or that they have the same preference for space and thus segregate.

**Household origin and race**

Various studies have observed segregation effects when defining ethnic groups or origin of households as a variable in residential location choice models. De Palma et al (2005, 2007) differentiate between households with foreign head and French headed households and demonstrates that households with a foreign head tend to group while households with a French head perceive the vicinity of foreign households negative. Waddell (2006) finds the very same effect for minority household and white households. While minorities favour the neighbourhood of minorities, white households will avoid the neighbourhood of minorities. Pinjari et al (2008,2011) and Guo et al (2007) differentiate various ethnic groups within their models and observe that all households tend to locate near households of the same group, so that we expect that the household like to group according to their ethnic group and origin and that this variable is an essential one to residential choice models.

**Household income**

Several studies used income groups as explaining variable in their residential location choice models and proved this to have a significant effect. Weisbrod et al (1980) observes a preference of high income households to locate around other high income households, while De Palma et al (2005, 2007) observe a grouping of low income households in his models. Also Zondag and Pieters (2005) show that middle or high income households like to reside in the neighbourhood of the same income group, but find a preference for low income households to reside near middle income households. Their models also differentiates between household types and show that retired seniors prefer middle income neighbourhoods and dislike high income neighbourhoods in the Netherlands. Waddell (2006) includes
interaction terms for low, middle and high income distribution with the according household types that also differ in age and size. His models show a positive influence for all income groups, but being of significance only for some of the low and high income household types. Further reports (Pinjari et al, 2008, 2011; Guo and Bhat, 2007) use the difference of the individual household income to the average income in the zone in their model and finds this attribute to have a significant negative value of high significance. All these reports show that households have a tendency to relocate around household groups of same or similar income.

**Employment**

A common observation described later in detail, is that households seek to be located close to their job. Different approaches have been reported implementing the distribution of employment into residential location choice models. Two studies integrate the unemployment rate in their models and show a negative effect on the utility of a residential location for London and Toronto (Andrew and Meen, 2006; Habib and Miller, 2009). Other studies use the density of jobs as an explaining variable. Zondag and Pieters (2005) demonstrate this to have a significant positive effect but only for non-single households with workers of middle or high income, while Srour et al (2002) observe a general attraction for locations with a high job density in Dallas. Pinjari (2008, 2011) test this variable for different household types, but do not find any significant effects. Also de Palma et al (2005, 2007) do not find any significant and constant behaviour, as this variable shows opposite signs in their two studies. This leads to the expectation that job distribution does not have any constant effect on residential location choice, but that households avoid locations with a high unemployment rate. The distribution of jobs is expected to rely on factors, that are partly already integrated as variables in the models, which are commonly used in employment location choice models. The correlation of those variables might explain the low significance when all variables are used in the same model.

**Housing costs**

Due to missing individual house prices several studies use zonal average costs of an accommodation within their residential choice models. All of those studies observe that these housing costs have a significant and negative influence on the utility of a location (Axhausen et al., 2004; Pinjari et al., 2008, 2011; Zondag and Pieters, 2005). Guo and Bhat (2007) use the ratio of income to the average housing price in a zone within their model instead of the absolute formulation, which is has a negative sign. As those studies use a zone as choice-alternative instead of an individual dwelling unit, this is not surprising and is comparable to the individual cost of an accommodation which is described later in this paper. It is not expected that beyond this effect, the average price of a zone will have any explanatory argumentation within the disaggregated residential choice models.

**School quality**

Various models of residential location choice have been reported using school quality as an explaining variable. Kim et al (2005) and Zhou et al (2008) use school quality as a non-interacted variable and find it to be a significant positive influence for all households. Chen
et al (2008) use it as interaction variable for households with children and without children and find it to be slightly stronger for households with children, but having a positive and significant sign for both. However the attribute “school quality” is not clearly specified and hard to argue on. Andrew and Meen (2006) use the GCSE-levels and A-level scores (proportion of children obtaining 5+ GCSEs grade; proportion of children obtaining 3+A/AS levels;) in their model they find a positive impact of the first variable but also report of a negative impact for the latter. They conclude that households rather send their children to other schools as reaction to poor local neighbourhood schools instead of moving, but also expect that another reason might be the correlation to deprivation index. Weisbrod (1980) uses the ratio of teachers to scholars as a quality indicator, which is simple to argue. He finds this one to have a positive influence on the residential utility of a location for households with children.

We would assume that school quality is of relevant for households that have children or are planning to have children, even if some studies find this variable to be valued positive be all households. Further research will have to explore the expected correlation to socio-demographic distributions and precise how school quality is measured best.

Divers variables
Some additional variables are only used in a few studies so that a general expectation on their influence and their use within residential choice models is not possible. Weisbrod et al (1980) uses the property tax rate per household as a variable in their residential location model which does show up to be of negative impact with a low significance. Bürgle (2006) also finds a negative impact for the study area of the Greater Zurich Region, though with a high significance, which reflects the local divergence of tax rates and the highly competitive behaviour of municipalities in Switzerland.

Two studies mention the use of rental vacancy within residential choice models. While Zondag and Pieters (2005) report of the positive influence of vacant housing in the neighbourhood for the Netherlands on all household types tested, Bürgle (2006) finds a negative impact instead using the municipal rental vacancy.

A common observation of Andrew and Meen (2006) and Weisbrod et al (1980) is that the crime rate has a significant negative impact on residential location choice.

Guo and Bhat (2007) are the only one to use the share of owner-occupied housing for residential location choice models and find this variable to be of positive effect for the location utility of owners.

3.1.3. Accessibilities

Accessibility is a measurement of the spatial distribution of activities about a point, adjusted for the ability and the desire of people or firms to overcome this spatial separation or ‘the potential of opportunities of interaction’ (Hansen, 1959). Accessibility-calculations are commonly applied to represent local and regional differences within the urban landscape. Five
main types of accessibility measures have emerged in literature (Bhat et al. 2002): spatial separation, cumulative opportunities, gravity measures, utility measures and time / space measures. Waddell (2006) summarizes a methodological overview.

Srour et al (2002) measures accessibility by means of different methods. They find that the cumulative opportunity accessibility index provides best results for residential location choice, based on statistical significance and behavioural interpretation. Ben-Akiva and Bowman (1998) consider a logsum measure based on the activity schedule of an individual and find this measure not to improve their model.

The activities that are used for accessibility calculations vary across the studies and can be of individual or generic character. We differentiate socio-demographic attributes and POIs as initial data.

**Socio-demographic attributes**

The accessibility of employments is reported to be of significance for several residential location choice models, but not showing a constant behaviour. On the one side Zolfaghari et al (2011) as well as Guo and Bhat (2007) observe a negative impact on the residential utility for San Francisco and London, The latter study also proves out that single households are more likely to reside near employment centres, a behaviour which already was reported within the population density variables and that is expected to reflect preferences for urban areas. On the other side Srour et al (2002) report of a positive influence for Dallas County.

Belart (2011) differentiates accessibilities based on public transport networks and road network in Zurich. The jobs accessible through public transport network prove to have a significant and positive effect for the location choice of households having no car, meanwhile the accessibility based on road network shows a negative influence on residential utility for car owners. Bürgle (2006) also observes a preference households not owning a car to live in places of high accessibility using the population distribution.

Zondag and Pieters (2005) calculate a logsum measure for all travel and find accessibility to be a significant variable in the move–stay choice. Accessibility of a specific location is only of significance for working non-single households and retired single households, being disliked by the earlier and favoured by the latter.

The impact of accessibility for socio-demographic attributes reported in literature does not allow to expect a common behaviour. The differentiation into socio-demographic groups (car owners, single person households) seems to capture observations that have already been done before: some types of households tend to have a preference for urban areas. Eventually the differentiation of these can lead to common observations, but the reviewed literature does not allow on a clear statement.

**Points of interest**

The implementation of accessibilities based on POIs have been leading to more regular results in the reported literature on residential location choice. All report that show significant effect
find a positive influence of the accessibility to shops (Zolfaghari et al, 2011; Zondag and Pieters, 2005), which fits to the observation discussed in chapter 3.1.2.

Lee and Waddell (2010) apply a time space prism measure to calculate shop accessibility and a highly disaggregated work travel time measure. These two measures are highly significant and have a relatively large influence. They argue that these factors are being considered by decision-makers and should be included in residential location models.

The accessibility to recreational facilities is rather found to be disliked by residing households, though (Zolfaghari et al, 2011; Srour et al, 2002).

Although we observe a constant behaviour on the accessibilities based on POIs, the number of reviewed paper does not allow for a statement here as well. Zondag and Pieters (2005) state that the effect of accessibility is marginal in comparison to the attributes of the location and the residential unit, as local differences in accessibility are small.

### 3.2. Residential unit

The choice alternative in the reported residential choice models have different scales, mainly depending on the initial data that was available. On less disaggregated models the agents might choose between different zones, city districts or grid cells (Andrew and Meen, 2006; Axhausen et al., 2004; Chen et al., 2008; de Palma et al., 2005, 2007; Guo and Bhat, 2007; Pinjari et al., 2008, 2011; Waddell, 2006; Weisbrod et al., 1980). Only few studies have been published using a building or the residential unit as alternatives, i.e. the dwelling, and single-family houses (Belart, 2011; Bürgle, 2006; Habib and Miller, 2009; Lee et al., 2010; Lee and Waddell, 2010; Vyvere et al., 1998; Zhou and Kockelman, 2008). Due to the different scales, also the variables that can be used to describe alternatives are divers. Zonal attributes are mainly captured by the location attributes described above, so we will concentrate on disaggregated characteristics of the residential unit in form of parcels, buildings and dwellings in this chapter. Lee and Waddell (2010) observes that these dwelling characteristics tend to dominate over accessibility indicators and recommends their use in models of residential location choice.

**Size**

An aspect which is often implemented in residential location choice models is the size of the alternative. While Zhou and Kockelman (2008) use this attribute in absolute values, Bürgle (2006) and Belart (2011) create a ratio using the household size to capture the space per person. Axhausen et al (2004) subtract the observed mean space per person from the individual space per person to capture regional difference. However, all of those approaches get to the same result, showing that households prefer to have more space per person when relocating and that this is always a relevant attribute in location choice models.

Beside the size of a dwelling also the number of (bed)rooms has been used in some location choice formulations. Habib and Miller (2009) observe a positive impact for a gain in number of rooms in their reference dependant model. Eliasson and Pagliara (2010) present an
approach similar to a two-step regression to measure the number of rooms in reference to the floor space. This formulation allows implementing both variables and avoids correlation problems of floor space and number of bedrooms. Their model show that single-person households prefer a lower number of rooms with more floor space per rooms, while all non-single households favour additional rooms instead. The integration of number of rooms into the residential choice model can thus be expected to enhance the explanatory power of the model, but eventually demands for correction of correlation with the floor space. Similar to the floor space variable, the number of rooms might then be integrated in relation to household size, although no study reported this approach.

Costs, price and value
Another attribute of the residential unit which is often used in reported studies is the price or rental costs. It is expected that this value captures various location characteristics as it can be modelled through regression models (Löchl and Axhausen, 2010), but nonetheless all models that used the price report on its significant negative impact. While some studies implement price as untransformed value (Andrew and Meen, 2006; Kim et al., 2005; Vyvere et al., 1998; Zolfaghari et al., 2011) others use the logarithmic expression of price (de Palma et al., 2005, 2007; Lee and Waddell, 2010; Lee et al., 2010; Habib and Miller, 2009). Some of those studies additionally include a ratio of price with the households income (Habib and Miller, 2009; Zolfaghari et al, 2011), which also has shown to be of significance when used as only price related variable (Waddell, 2006; Weisbrod et al, 1980; Belart, 2011; Bürgle, 2006; Lee and Waddell, 2010; Zhou and Kockelmann, 2008). Walker and Li (2007) observe that the pricesensitivity reduces with rising income, so that the integration of the ratio end eventually the logarithmic expression of this one seems reasonable. Alternatively the interaction to income groups could be useful. As Axhausen et al. (2004) state that owners and shared accommodation have a higher willingness to pay for their location compared to other users, a further differentiation of those groups is recommended.

Srour et al (2002) and Waddell (2006) also explore the effect of the improvement value, on the residential utility of a residence. Both report that this one is valued positive by all households types tested, which is not surprising as a high value of a building, representing high quality can be expected to be favoured by households. Nonetheless it should be mentioned that Weisbrod et al. (1980) did not find any significant impact on this variable back in 1980.

Srour et al (2002) also use the average lot value in a zone (price of the soil) as explaining variable. It is found to be of negative impact on the residential utility of a location but should be captured in the price of a residential unit for disaggregated models.

House type
Various studies have observed the preference of specific house types. Due to differences in the historic heritage of cities and the diverge ownership and rent-behaviour within different nations these are expected to be off different behaviour for diverse study areas. Srour et al

---

4 According to the population census 2000 and the Euroconstruct (2008), homeownership rates are at 24.8% for the Canton of Zurich, compared to Germany (43%), France (58%), Austria (56%) and Italy (73%).
(1998) report of a general preference for houses in Belgium and Habib et al (2009) find a negative utility for attached houses in Toronto. Lee et al (Lee and Waddell, 2010; Lee et al., 2010) reports that single person households and renters a preference for multifamily houses. Their model also show that households with children favour single family buildings. Axhausen et al (2004) observes that households tend to stay with the same type of housing and proposes to include the previous location type as variable. Eventually this could capture some self-selection effects as a preference for certain house types by different lifestyle groups. On the other side this would ignore changes in the household composition that might be the reason to relocate. In conclusion of the reviewed literature we expect that housing types are of relevance for the residential location choice and differ within household-types, with families favouring single family (detached) houses and single person households being more attracted by multifamily houses.

Features of the dwelling

Only few studies have reported on the use of specific residential unit features and characteristics, mainly due to missing data on those variables. Vyvere (1998) finds the number of garages to be of importance to households when relocating, although we can expect that this is only of importance to households with a cars. He also shows that households dislike buildings being built before 1960. The value of historic buildings has not been explored explicitly, although Srour et al (2002) find the average age of buildings in a zone to be valued positive, but we would expect a positive influence.

3.3. Household attributes

The implementation of attributes specific to a household is implemented using interaction terms if not being systematically different for the alternatives, as the estimation process of the choice models will not converge otherwise. Many household attributes have thus already been described in the previous chapters to express behavioural differentiations. In this chapter we will not summarize these main findings but concentrate on the attributes that are systematically different.

Commuting time

The commute time is often used in residential location choice models, when the workplace of the choosing person is given and the commuting time is available for all alternative. It generally obtained from a transport model and needs to be calculated for every alternative of the choice set. The commute time to work from a location tends to be of negative influence to the residential utility a location (Guo and Bhat, 2007; Habib and Miller, 2009; Zhou and Kockelmann, 2008; Lee and Waddell, 2010; Axhausen, 2004; Zolfaghari et al, 2011). If the commuting time is not given some studies use the commuting distance as network based distance or Euclidean distance instead (Belart, 2011, Bürgle, 2006; Chen et al, 2008; Srour et al, 2002), which always has been reported to have the same negative sign, but has a lower significance. Few studies differentiate between commuting time by car, public transport and eventually commuting costs (Pinjari, 2008, 2011; Kim et al, 2005), which all are reported to have a negative impact on the residential utility when being of significance. Due to correlation not all
variables are of significance when being implemented in a model (de Palma 2007). Belart (2011) and Pinjari et al (2008) include the average commuting time respectively commuting distance of all workers in a household which proofs to further enhance the model quality. A weighted approach based on percentage of work has not proofed to enhance the model though (Belart 2011). To capture the reduced effect of long distance variations, Bürgle (2006) uses a formulation within an exponential function which is also implemented by Belart (2011).

Distance to previous location and social networks
Only few studies explore the influence of distance to previous location on the utility of a new residential location. Axhausen et al (2004) and de Palma et al (2005, 2007) report that households tend to stay close to their previous location or to stay in the same district. Zondag and Pieters (2005) combine the Euclidean distance with its logarithmic transformation and find it to be the most dominant variable in their model for various household types tested, all being negative except for households over 65. Gordan (1992) reports of a desire to maintain social networks, so it can be expected that this variable is of relevance for residential relocation. Vvvere et al (1998) introduce the distance to social contacts in their model, and show a negative impact on residential utility. Belart (2006) further explores this impact and observes a preference for proximity to social contacts when using distance to social contacts weighted with the number of meetings per month. Our own later studies with the same dataset showed that this value is correlated to the distance to previous location, though. In addition it is unclear whether location is influences by social contacts or whether the social network is formed by the location. We thus suggest to use social network distance or distance to previous location in a model with a preference with the easier to observe latter one.

Lifecycles and lifestyles
As distance to previous location is disliked by residing households we would assume that households tend to stay on the same location, which is contradictory to the observation of lifecycle relocations. Various studies proved that the relocation probability depends on different lifecycle events as change of marital status job change, number of children, size of household, retirement and age (Andrew and Meen, 2006; Eluru et al., 2009; Kim et al., 2005; Lee et al., 2010). When moving the relocation tends to be toward the city centre in the early part of the lifecycle and tends to move away from the city in later part (Andrew and Meen, 2006), an observation which is also captured by the different behaviour for location attributes on urban characteristics: As described previously, young individuals tend to like mixed use areas with a higher density of persons, while others household types dislike these.

Beside the differentiation of households based on their lifecycles, some studies will use the concept of lifestyles instead. Lifestyles are an approach of the social sciences that have been explored since the 1980s and have proven to be of explanatory value for transportation models. They identify groups of persons based on their common behaviour in daily life and their cultural, social and leisure behaviour (Müller, 1992). Belart (2006) explores the self-selection-effects as alterative to the lifecycle approach, based on the classification of Otte (2005) and an own classification. He differentiates four groups of lifestyles that will slightly
enhance the model fit when being interacted with population density, open space, travel time and percentage of same household in neighbourhood.

Belart (2006) also gives a general overview on the lifestyle concept and links to other studies that have used lifestyles as explaining variables in their residential choice models. Walker and Li (2007) differentiate three types of lifestyles that show different behaviours for urban density, retail and service density as well as lot sizes, with car oriented households having a preference for lower densities and bigger lot sizes. Krizek und Waddell (2002) differentiate nine types of lifestyles that show differences in location choice behaviour and mobility behaviour as well by interaction with urban density and travel distances.

At present there has been no systematic comparison of the lifestyle approach to lifecycle classes, which are easier to argue. The variation of the lifestyle clusters used in literature does not allow a comparison of the model results, so that the evaluation of the lifestyle-concept, versus the lifecycle concept will demand some further research.

3.4. Summary

The reviewed literature shows common behaviour for some attributes on different study areas that will be summarized here.

While the built space variables have not been used in many models up to now, especially the density measures and land use mix seem to be a promising way to capture the extend of urban areas. This variable has often been implemented by predefining the extend and showed to be closely related to lifecycles in various studies. A differentiation of households using age should thus be envisioned. Beside the density the network of road and railways is an indicator for noise that can be implemented as a buffer and showed a negative impact for all household types. Guo and Bhat (2007) have stated that the perception of build space in terms of land use is wider then for socio economic variables, an aspect which will demand for further research.

The points of interest form a group of variable that has appeared in several studies with a constant behaviour as well. Facilities of education, service, retail and local transport are being valued by all household types, i.e. distance is disliked and density of these facilities is appreciated. For stations and highway exits the preferences are divers, depending on the mobility of the household and the noise level. The location of the Central Business District reflect the same behaviour as for urban settlements: young households tend to favour the proximity while all others prefer to locate away from it.

Socio-economic variables are the most used group of variables for residential location choice, often used in form of interaction terms. The population density seem to reflects the location of urban centres and eventually might be correlated to it. It is often preferred by single households and young persons. All household types tend to dislike the unemployment rate in a location and the crime rate. And almost all models implementing the share of household types found a significant grouping tendency of households types based on size, age, children,
income and the origin or ethnic group. Accessibilities being based on job distribution, though having been addressed in several reports often with a high level of significance didn't show any constant behaviour. We expect that these reflect different preferences of mobility and would suggest to interact these with transport modes as done by Belart (2011). It is expected that these accessibilities correlate with variables on urban structure as dwelling density or population density. This would explain the sometimes reported preference of single households for highly accessible locations. The accessibility of shops has shown to be valued by all household types, but is partly already captured by the density of shops.

On the characteristics of the residential unit we observe a less household type specific behaviour. Price shows to be disliked in all models, but with a different attitude level for owners, renters and shared accommodations. Floor space is valued by all households, but single person households tend to prefer less and bigger rooms in comparison to other household types. The differentiation of housing types into single family (detached) and multifamily are mainly of relevance for families with children.

For the individual networks of the household, we can observe that commuting distance variable is reported to be of significant negative impact in most models, as well as the distance to the previous location, which is rarely used, but highly recommended.

A final overview on the variables found to be relevant within the reviewed literature is given in Table 1.
Table 1: Conclusion on common behaviour and expected influences

<table>
<thead>
<tr>
<th>group</th>
<th>variable</th>
<th>function</th>
<th>radius</th>
<th>expected sign</th>
<th>expected correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>built space</td>
<td>density</td>
<td>density</td>
<td>600+</td>
<td>+</td>
<td>density of population</td>
</tr>
<tr>
<td></td>
<td>density of dwelling units * nonsingle hh with children</td>
<td>density</td>
<td>600-</td>
<td>-</td>
<td>density of population</td>
</tr>
<tr>
<td></td>
<td>density of dwelling units * young hh</td>
<td>density</td>
<td>600+</td>
<td>+</td>
<td>density of population</td>
</tr>
<tr>
<td>open space</td>
<td>open space or unbuild space</td>
<td>spatial</td>
<td>ratio</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td>percentage of water</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>land use</td>
<td>mixed land use * young household</td>
<td>spatial</td>
<td>ratio</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mixed land use * no car</td>
<td>spatial</td>
<td>ratio</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mixed land use * children</td>
<td>spatial</td>
<td>ratio</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>share of industrial land</td>
<td>spatial</td>
<td>ratio</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>share of residential land</td>
<td>spatial</td>
<td>ratio</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td>share of commercial land-use</td>
<td>spatial</td>
<td>ratio</td>
<td></td>
<td></td>
</tr>
<tr>
<td>urban character</td>
<td>dummy for city, urban core, urban fringe</td>
<td>distance</td>
<td>+</td>
<td></td>
<td>accessibility of jobs</td>
</tr>
<tr>
<td></td>
<td>distance to center</td>
<td>distance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>network/noise</td>
<td>buffer to arteries and railways (noise)</td>
<td>buffer</td>
<td></td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>points of interest</td>
<td>distance to school</td>
<td>distance</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>density of schools</td>
<td>density</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>distance to university</td>
<td>distance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>distance to kindergarten</td>
<td>distance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>service and retail</td>
<td>distance to retail</td>
<td>distance</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>density of retail</td>
<td>density</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>density of service</td>
<td>density</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sport and recreation</td>
<td>density of sport activity centers</td>
<td>density</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>density of natural recreation centers</td>
<td>density</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>transport</td>
<td>distance to local transport</td>
<td>distance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>distance to station * no car owner</td>
<td>distance</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>distance to highway exit * car owner</td>
<td>distance</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>density of local transport</td>
<td>density</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>urban center</td>
<td>distance to urban center (CBD) * young household</td>
<td>distance</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>distance to urban center (CBD) * young household</td>
<td>distance</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>distance to local center</td>
<td>distance</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>socioeconomic</td>
<td>population density</td>
<td>density</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>populations density</td>
<td>density</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>share of hh with same size</td>
<td>spatial</td>
<td>ratio</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>share of hh with same age</td>
<td>spatial</td>
<td>ratio</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td>share of hh with children * hh has children</td>
<td>spatial</td>
<td>ratio</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td>share of hh with same income cat</td>
<td>spatial</td>
<td>ratio</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>share of hh with same ethnic group/origin</td>
<td>spatial</td>
<td>ratio</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>unemployment rate</td>
<td>value</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>crime rate</td>
<td>value</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>accessibility</td>
<td>accessibility (cumulative opportunities) of jobs * no car or young or single</td>
<td>acc</td>
<td>+</td>
<td></td>
<td>distance to centre</td>
</tr>
<tr>
<td></td>
<td>accessibility (cumulative opportunities) of jobs * car</td>
<td>acc</td>
<td>-</td>
<td></td>
<td>distance to centre</td>
</tr>
<tr>
<td></td>
<td>accessibility (cumulative opportunities) of shops</td>
<td>acc</td>
<td>+</td>
<td></td>
<td>density of retail</td>
</tr>
<tr>
<td>residential unit</td>
<td>floorspace per person</td>
<td>value</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>mom density * non single hh</td>
<td>value</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>mom density * single hh</td>
<td>value</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>log (price/income) * hh is owner</td>
<td>value</td>
<td></td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>log (price/income) * hh is owner</td>
<td>value</td>
<td></td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>log (price/income) * share accommodation</td>
<td>value</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td>improvement value or building quality</td>
<td>value</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>historic building (build before 1920)</td>
<td>value</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>household type</td>
<td>distant single family * hh with children</td>
<td>value</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>multfamily housing * single person hh</td>
<td>value</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>building age</td>
<td>value</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>commuting time/distance/costs by car * available</td>
<td>distance</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>commuting time/distance/costs by pt * no car available</td>
<td>distance</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>distance to previous location</td>
<td>distance</td>
<td>-</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4. Summary and Outlook

This paper presented a comparison and review the findings of recent publications on disaggregated residential location choice models, in order to use their findings and inform an implementation of a land use simulations models for the Kanton Zurich, Switzerland.

Many of these papers differ in methodological approaches, and as shown in Figure 1, not only the methodology but also the estimated variables and their formulation varies extremely. The comparison of the variables' influence can thus only be performed in a qualitative way, including some prior expectations to sign. Nonetheless this comparison of measurements for residential location utility, allowed us to observe common behaviour of residing household across a variety of study areas.

Figure 1: Amount of variables reported for residential location choice

Based on this review we summarized a basic set of variables with a general expectation on their sign and influence (Table 1). We expect these variables to allow a “default” setup for the residential location modelling, i.e. to be a common set of variables with a strong explanatory power for the location utility in western cities. We believe this reduced default set of variables is a first step to also allow for comparison of future residential location choice studies within
different study areas. They are thus of important use in disaggregated land use simulations and should be implemented within the data structure of land use simulation software.

In order to allow for this integration, variables have been structured into three categories according to their initial data sets. We differentiate between geometric objects, points of interests, which are commonly planned objects, and socioeconomic variables that represent the spatial occupation of the urban landscape through individuals. Not all of these are currently implemented in the data model of our SustainCity project.

However, based on the variance in model configurations, we are aware that this review cannot be seen as proof of concept and demands for further research. On one hand it will be necessary to explore the variables that are expected to be of common relevance using the disaggregated data of the SustainCity project. On the other hand it will be of importance to study the influence of the different variable groups in separate models. Figure 2 shows the total number of variables reported in the reviewed literature and demonstrates the strong dominance of the socio-economic variables. We could already observe different behaviours within those variable groups and Guo and Bhat (2007) report on differences in their spatial perception. Only through understanding these communalities and differences we will be able to argue on their relevance for simulation.

Figure 2: Distribution of variables in the reviewed literature
5. References


Hall, Peter, Ed. Von Thünen's Isolated State (English translation by Carla M. Wartenberg, with an introduction by the editor), Pergamon Press. 1966.


