

RESIDENTIAL LOCATION CHOICE MODELLING: A MICRO-SIMULATION APPROACH

JAKUB VOREL

Czech Technical University in Prague, Faculty of Architecture, Department of Spatial Planning, Czech Republic

ABSTRACT

Micro-simulation models have been increasingly used for studying various urban and regional processes. Here, two experimental micro-simulation models are applied to the study of residential location choices of inhabitants of the Tábor micro-region. A wide range of environmental and socio-economic characteristics are analysed for their potential impact on individual residential location choices. The micro-simulation approach proves to be useful for analysing not only housing, neighbourhood, and accessibility characteristics, but also of the interactions between the characteristics of the present and potential new residential locations of individual inhabitants and the role of their personal characteristics in their choice of a new residential location. The ability of the micro-simulation models to replicate the observed residential choices is evaluated by several quantitative indicators with special attention given to the stochasticity of the model behaviour, which is a typical feature of micro-simulation models. The limited availability of sufficiently disaggregated data describing the demographics of households, their socio-economic characteristics, and real estate market dynamics needs to be resolved in order to exploit the full potential of micro-simulation modelling in the future.

Keywords: residential location choice, micro-simulation modelling, discrete choice models, multi-nominal logit, micro-data

1. Introduction

This paper presents an experimental application of a residential location choice micro-simulation model for the Czech Republic. The two main goals of the experimental application were: a) to provide new insights into residential mobility, which is an essential urban process with a strong impact on changes in land use, b) to test the conditions for the applicability of micro-simulation models in the specific context of the Czech Republic, where there is no tradition of micro-simulation modelling and there is limited availability of suitable data.

Micro-simulation modelling is an alternative to the first generation of urban simulation models. The first generation models were considered to be too aggregated to represent the local variability of social and environmental characteristics, and consequently not able to properly represent the human-environment interaction. They were also considered too mechanical, as they ignored the complexity of human decision making, mainly the diversity of factors and constraints influencing the behaviour of individuals (Lee 1975).

Unlike the first generation models, micro-simulation models and related agent-based models are highly disaggregated. The decision making of individual agents – human actors, households and companies – is explicitly represented, and this makes it possible to explain the influences of a broad range of personal characteristics of agents, and also a broad range of characteristics of the environment related to individual agents. This ability makes micro-simulation and agent-based modelling an exceptionally suitable tool for studying the human-environment interaction on an individual level.

Several comprehensive micro-simulation models have been implemented so far: the UrbanSim model in Oregon,

USA (Waddell, Wang, Charlton, & Olsen 2010; Waddell 2002), the San Francisco Bay area model, California, USA (Waddell 2013a), the Île-de-France model, France (IAURIF, THEMA 2004, 2005, 2007), the SimDELTA model in the United Kingdom (Simmonds & Feldmann 2007; Simmonds, Christodoulou, Feldman, & McDonald 2011; Simmonds 2010), the ILLUMASS model in Dortmund, Germany (Strauch et al. 2005; Wegener & Spiekermann 2011), and the ILUTE model in Toronto, Canada (Salvini & Miller 2005, 2005). These models are comprehensive enough to capture the interdependence of essential urban processes, especially population demographics, residential mobility, the evolution of individual companies and their mobility, transportation, the real estate and job markets, and the development of the urban structure and infrastructure. A typical comprehensive model consists of several autonomous sub-models, each addressing particular urban processes in a specific way.

The experimental micro-simulation models described here focus on residential mobility, and specifically on residential location choice. They are intended to supplement an already existing land use change model by the demand side of residential land use changes (Vorel & Grill 2013).

The experimental residential location choice models replicate the residential moves of individual inhabitants in the Tábor micro-region in the southern part of the Czech Republic. The Tábor micro-region consists of 79 municipalities with a total population of 80,641 and a population density of 80.5 inhabitants/km² in year 2011. The Tábor micro-region has an area of 1002 km² and approximates the catchment area (Local Labour System Area) of the town of Tábor, which is the main employment and administrative centre for the micro-region.

The Tábor micro-region features a relatively large proportion of small municipalities: 44 out of 79 municipalities have less than 200 inhabitants in total and contain only 6.35% of the population of the micro-region. Only 10 municipalities have a population greater than 1000 inhabitants. These 10 more populous municipalities contain 78.93% of the population of the micro-region. The highest percentage (43.52%) of the population is concentrated in Tábor, which is the biggest municipality (34,430 inhabitants). Tábor is the only municipality in the micro-region with a population greater than 10,000 (2011 Population and Housing Census 2013).

The average age of the population of the micro-region is higher than the national and regional average. In 2011, the age index (the number of inhabitants older than 64 years per one inhabitant younger than 15 years) was 1.2 for the Tábor micro-region, as against 1.1 for the South Bohemia region and for the whole Czech Republic (2011 Population and Housing Census 2013).

The micro-region shows considerable differences in socio-economic characteristics between the highly-urbanized municipalities that form the central Tábor agglomeration (Tábor, Sezimovo Ústí and Planá nad Lužnicí) and the rural and less populated municipalities on the periphery, especially in terms of age structure, education status and employment structure. The age index is below 0.8 in the Tábor agglomeration municipalities, while it exceeds a value of 2 in the peripheral municipalities. The university-educated population is concentrated in Tábor and neighbouring municipalities, representing more than 8% of the population, while it represents only about 1% of the population in the peripheral municipalities. While the municipalities in the Tábor agglomeration have the biggest share of employment in industry and services, the peripheral municipalities have a share of up to 25% of employment in the primary sector (2011 Population and Housing Census 2013).

The spatial distribution of economic activities in the micro-region is also uneven. Most of the workplaces are concentrated in the Tábor agglomeration. The number of workplaces exceeds the number of economically-active inhabitants in only 10 municipalities (2011 Population and Housing Census 2013).

As will be documented in following text, the heterogeneity of the municipality characteristics, especially the population size, is a challenge for residential location choice modelling. Micro-simulation models are usually implemented as discrete choice models. The sections that follow will first present the theoretical background of discrete choice models and their most frequent operationalization in the form of multi-nominal logit models. Then the concept of residential mobility and an analysis of residential mobility factors will be presented. Two residential location choice models were assembled and evaluated for this purpose. Their usability, validity and the limits to their application in the context of the Czech Republic will be presented.

2. The concept of discrete choice micro-simulation models

2.1 Formal definition of discrete choice models

Discrete choice models operationalize the decision-making of individuals so that each individual makes choices over a finite number of choice alternatives. For example, an inhabitant planning to relocate chooses a place of residence among the municipalities in a micro-region. The choice of the individual is influenced by the characteristics of the choice alternatives, and also by her or his own personal characteristics. Discrete choice models quantify the effects of the characteristics on the choice process, and then use this knowledge to replicate the choices that an individual would make in various hypothetical situations.

Formally, discrete choice models are implemented as generalized linear models, most often as multi-nominal logit (MNL) models.¹ Logit models link the linear combination of $k \in K$ independent variables χ_k and their associated coefficients β_k to the dependent categorical variable J . Variable J represents the set of choice alternatives j , here represented by the individual municipalities. The independent variables χ_k represent the characteristics $k \in K$ of the individual choice alternatives as well as the personal characteristics of the individual making the choice. The coefficients β_k related to independent variables χ_k represent the effect of the characteristics on the choice of alternative $j \in J$.

Unlike in linear regression models, the independent variables χ_k in MNL models are related to the dependent variable J only indirectly via the link function called logit. Logit is defined as the log odds of the choice probabilities of the examined alternative j and the choice probabilities of one of alternatives selected to be the reference alternative j_r :

$$\text{logit}(j) = \ln \left(\frac{P(\text{choice} = j)}{P(\text{choice} = j_r)} \right) = \sum_{k=1}^K \beta_k \chi_k \quad (1)$$

The model parameters can be interpreted more easily by exponentiating both sides of the equation:

$$\text{odds}(j) = \ln \left(\frac{P(\text{choice} = j)}{P(\text{choice} = j_r)} \right) = \prod_{k=1}^K e^{\beta_k \chi_k} = e^{\sum_{k=1}^K \beta_k \chi_k} \quad (2)$$

After transformation the individual e^{β_k} (odds ratios) are directly related to the odds of the choice probabilities $\text{odds}(j)$ of alternative j . A unit change of independent variable χ_k causes an e^{β_k} change of the choice probability of alternative choice j relative to the probability of reference alternative choice j_r .

¹ Extensive information on the use of logit models for discrete choice modelling can be found in (Ben-Akiva & Lerman 1985; Liao 1994; Train 2009)

As the odds ratios e^{β_k} indicate the change in the odds of the probabilities, and not the probability itself, another formal expression of a logit model must be used for directly predicting the probability of the choices:

$$P(j) = P(\text{choice} = j) = \frac{e^{\sum_{k=1}^K \beta_k \chi_k}}{\sum_{j=1}^J e^{\sum_{k=1}^K \beta_k \chi_k}} \quad (3)$$

As demonstrated, the logits, odds and probabilities are convertible to each other. The use of one or another form of logit model depends on the context of the use: the odds ratios $odds(j)$ are the most suitable for interpreting the factor influencing the choice, while the probabilities $P(j)$ are more often used for predicting the choices, utilizing Monte Carlo methods.

2.2 Estimating the parameter values

Each combination of parameter values β_k , $k \in K$ leads to a specific likelihood value, which is equal to the probability of the dependent variable being precisely predicted given the parameter values β_k . The goal is to set the values of parameters β_k to maximize the likelihood of the model. Because the likelihood is usually too small for computational purposes, the log likelihood (LL) is used instead. The value of LL is in the range from negative infinity to zero; the closer to zero, the better the fit of the model to the observed data (Ben-Akiva & Lerman 1985; Liao 1994; Train 2009).

The parameters β_k are themselves random variables, and the probability that they are equal to zero (null hypotheses) should be tested. The log likelihood ratio test, which is the ratio of the log likelihoods of two models – one with the tested variable and the other without the tested variable – is used as the test statistics:

$$-2(LL_{\text{base model}} - LL_{\text{estimated model}}) \quad (4)$$

This statistics follows the chi-square distribution with the degrees of freedom equal to the difference between the numbers of parameters used in the models.

Alternatively, we can use the Wald statistic, which tests the significance of individual parameters of the model:

$$\text{Wald} = \beta_i / \text{standard error of } \beta_i \quad (5)$$

3. Experimental micro-simulation models of residential location choice

Residential mobility is the output of two distinct decisions made by individual households: the decision to relocate, and the choice of a new residence (Coulombel 2011; Pacione 2009). Households decide to relocate when they reach a certain level of stress due to discordance between their housing needs, aspirations and expectations, on the one hand, and their actual living conditions, on the other.

Only the residential location choice is addressed by the experimental models presented here. The models are limited to a single catchment area, therefore replicating only short-distance mobility inside a micro-region. Long-distance mobility, which involves relocation from one housing and labour market to another, is exogenous to the simulation models. The available data on relocations is aggregated to the municipalities, which predetermines them to be choice alternatives.

Residential location choice can be influenced by a number of residence and neighbourhood characteristics, and also by the characteristics of individual households making the choice. Micro-simulation discrete choice models are suitable for studying the interdependences between the choices made by individuals and their personal characteristics. The explicit representation of the choice process enables one to experiment with choice constraints in various phases of the decision process (Ben-Akiva & Lerman 1985; Train 2009; Waddell 2002).

Residential location choice is the outcome of collective decision making by the members of a household. This is an extremely complex matter, as the interdependence of the activities of individual household members and their different interests leads to conflicts that have to be resolved during the decision making process (Axhausen 2005). To cope with the complexity, most of the reviewed residential location choice models assume that households, rather than individual persons, are the decision making entities. The decision-making processes are therefore usually modelled on the basis of the characteristics of households rather than on the characteristics of individuals. Demographic changes on the level of households, changes in economic status and in the working place of economically active household members, and the number of cars used by households, are characteristics that usually enter the decision process.

Unfortunately, no data on residential mobility of households is available at the moment in the Czech Republic. There is only data on the mobility of individual actors.² It was not possible to aggregate the individuals to households on the basis of their temporal and spatial coincidence of relocation, and by matching the personal characteristics of individuals, because significant numbers of individuals relocate in order to join each other in new households, and household formation would therefore need to be controlled on an individual level. Instead, the age of the individual was used to indicate her or his role in the collective decision making of the households. Young individuals were assumed to follow the decisions of their parents, and therefore to have a similar propensity to relocate and the same choice preferences as their parents.

Two approaches were adopted for an examination of the impact of personal characteristics on the decision

² Data on the movement of individual households will be available for research purposes in the new population census, when it is made available.

making process: a) a comparison of several models, each representing the decision making of individual population strata, b) measuring the interaction effects between personal and choice alternative characteristics in a single model that includes the whole population. The first approach leads to a stratified model, while the second approach leads to a general population model.

3.1 The stratified model

The stratified model stratifies the population into five age groups. Three sub-models bring together age groups of individuals that are expected to be joined by common households and therefore by similar residential choice behaviour.

- sub-model 1 for young-age group: 0–9 year-old and 25–34 year-old individuals,

- sub-model 2 for middle-age group: 10–24 year-old and 35–54 year-old individuals,
- sub-model 3 for old-age group: 55 year-old and older individuals.

With one exception, which is discussed below, all three sub-models use the same set of characteristics. The aim is to compare how their effects on the decision making differ between particular age groups. This approach was adopted although the effects of several characteristics were not significant in all three sub-models.

3.2 The general population model

The general population model does not stratify the population on the basis of personal characteristics, but it includes observed relocations of all members of the population in a single discrete choice model. To measure

Tab. 1 The odds ratios e^{β} of the stratified model with their statistical significance levels indicated: * 0.05 (t-value 1.95), **0.001 (t-value 3.29).

Personal characteristics and characteristics of a municipality	Name of variable	Odds ratios e^{β} of sub-models		
		1	2	3
The proportion of young age population (0–19 years old)	YOUNG	12.29718**	9.956643**	1.38227
Log of number of jobs	JOBS	1.40641**	1.31367	1.08789
Average distance from municipality to railway station less than 3500 m	$I_{DR}(1)$	1.417592**	1.68804**	1.35744**
The proportion of apartment houses in municipalities having a population less than 1600	$BD^*I_{PC}(1)$	0.33425*	0.56185	0.04773**
The proportion of apartment houses in municipalities having a population larger than 1600	$BD^*I_{PC}(0)$	1.75439**	2.35679**	1.47815**
The proportion of forest area	FOREST	3.21694**	4.11031**	2.92046*
The percentage of municipality area designated as developable land	DEVLAND	1.00398	1.02373*	1.00042
Log of number of public services	PUBS	1.27027**	1.27227**	1.63406**
Minimum one basic school	$I_{BS}(1)$	1.54169**	1.36665**	3.29649**
Bechyňsko from outside	$I_{BE}^e(0) \wedge I_{BE}^a(1)$	0.22375**	0.24837**	
Bechyňsko from inside	$I_{BE}^e(1) \wedge I_{BE}^a(1)$	5.85455**	8.26806**	
Choustecko from outside	$I_{CO}^e(0) \wedge I_{CO}^a(1)$	0.38052**	0.42309**	
Choustecko from inside	$I_{CO}^e(1) \wedge I_{CO}^a(1)$	4.282984**	4.186688**	
Chýnovsko from outside	$I_{CY}^e(0) \wedge I_{CY}^a(1)$	0.92207	1.05846	
Chýnovsko from inside	$I_{CY}^e(1) \wedge I_{CY}^a(1)$	2.45532**	2.20266**	
Jistebnicko from outside	$I_{JL}^e(0) \wedge I_{JL}^a(1)$	0.73666**	1.10580	
Jistebnicko from inside	$I_{JL}^e(1) \wedge I_{JL}^a(1)$	0.83305	1.75303	
Malšicko from outside	$I_{MA}^e(0) \wedge I_{MA}^a(1)$	0.88445	1.07926	
Malšicko from inside	$I_{MA}^e(1) \wedge I_{MA}^a(1)$	1.01569	3.18381**	
Vožicko, from outside	$I_{VO}^e(0) \wedge I_{VO}^a(1)$	0.52209**	0.59904**	
Vožicko from inside	$I_{VO}^e(1) \wedge I_{VO}^a(1)$	5.53449**	11.21913**	
Bechyňsko	$I_{BE}^e(1)$			0.65437*
Choustecko	$I_{CO}^e(1)$			0.35921*
Chýnovsko	$I_{CY}^e(1)$			2.41013**
Jistebnicko	$I_{JL}^e(1)$			0.77554
Malšicko	$I_{MA}^e(1)$			0.70229*
Vožicko	$I_{VO}^e(1)$			0.92920

how much the personal characteristics influence decision making, they must be linked by the interaction term with the characteristics of the choice alternatives. The interaction term indicates how much the personal characteristic modifies the effect that the municipality characteristic has on the choice probability of the municipality. The number of municipality characteristics was tested for their potential interactions with the age, education and family status of individual actors.

3.3 Interpreting the model parameters

Due to the non-linear relationship between the dependent and independent variables, the model parameters are not straightforward to interpret. The interpretation of three types of independent variables: continuous variables, categorical variables and variables entering into interaction with other variables will be demonstrated.

In the case of continuous variables, the β_k parameters represent the change of the $\text{logit}(j)$ caused by a unit change of a continuous variable k . For example, a unit change of the *DEVLAND* variable that represents the percentage of the area of the municipality designated for development will have the following effect on $\text{logit}(j)$ in a single population model:

$$\text{logit}(j) = \beta_{DEVLAND} \times DEVLAND = 0.0368722 \times DEVLAND$$

To interpret the effect of the variable meaningfully, the parameter $\beta_{DEVLAND}$ has to be transformed to the odds ratio $e^{\beta_{DEVLAND}}$. The odds ratio indicates the

increase of the choice probability of the municipality by $e^{0.0368722} = 1.03756$ times when *DEVLAND* increases by one per cent.

To evaluate the effects of categorical independent variables, their discrete values have first to be transformed to indicator variables, here referred to as dummy variables. An individual dummy variable is created for each category of original variable, with the exception of one implicit reference category. For example, the dichotomous variable *BS* is represented by single indicator variable $I_{BS}(1)$ indicates municipalities with at least one basic school, while $I_{BS}(0)$ indicates municipalities without a basic school. The value of parameter β_{BS} of this dummy variable indicates the change in $\text{logit}(j)$ caused by the presence of a basic school in the municipality, in this case for the choices made by the young population:

$$\text{logit}(j) = \beta_{BS} I_{BS}(1) = 0.432887 \times 1 = 0.432887$$

The probability that an individual member of the young population will choose a municipality with at least one basic school is $e^{0.432887} = 1.541702$ times higher than the probability that she/he will choose a municipality without a basic school.

Some continuous variables did not prove to be significant unless they were dichotomized. This is the case for several continuous variables: average distance from municipality to railway station shorter/greater than 3500 m, the municipality having more/less than 0.8 jobs per one economically-active inhabitant, and municipalities in which the housing stock expanded by more/less

Tab. 2 The odds ratios e^{β} of the general population model with their statistical significance levels indicated: * 0.05 (t-value 1.95), **0.001 (t-value 3.29).

The personal characteristics and characteristics of a municipality	Name of variable	Odds ratios e^{β}
Average distance from municipality to railway station less than 3500m	<i>D_RAIL</i>	1.17509**
Log of average road distance from the municipality to Tábor municipality (m)	<i>D_CENTRE</i>	0.99994**
Job centre (more than 0.8 jobs per one economically active inhabitant) for old population	$I_{JC}(1) * I_{OA}(1)$	0.75261**
Job centre for young population	$I_{JC}(1) * I_{YA}(1)$	1.21213**
Job centre for middle-age population	$I_{JC}(1) * I_{MA}(1)$	1.1857**
Number of public services	<i>PUBS</i>	1.01671**
Impact of the population size of the municipality on the effect of additional public services	<i>PUBS * PSIZE</i>	0.9999995**
Presence of at least one basic school, if the relative increase in the number of flats between 1999 and 2006 is less than 4%	$I_{BS}(1) * I_{FG}(0)$	1.287709**
Presence of at least one basic school, if the relative increase in the number of flats between 1999 and 2006 is more than 4%	$I_{BS}(1) * I_{FG}(1)$	1.017397**
The proportion of flats in apartment houses in municipalities having a population less than 1600	$BD * I_{PC}(0)$	0.92676
The proportion of flats in apartment houses in municipalities having a population higher than 1600	$BD * I_{PC}(1)$	2.63611*
The proportion of the population older than 65 years	<i>OLDPOP</i>	0.97772**
The proportion of forest area	<i>FOREST</i>	2.85211**
The percentage of the municipality area that is designated for development	<i>DEVLAND</i>	1.03756**
Log of municipality area (m ²)	<i>ASIZE</i>	1.69561**
Log of number of flats built between 1999 and 2006	<i>NFLATS</i>	1.27702**
The proportion of vacant family houses	<i>VACANCY</i>	0.41312**

than 4% between 1999 and 2006. The thresholds used to dichotomize the variables were empirically established in such a way that the statistical significance of the dichotomized variables was maximized. The municipality population characteristics were dichotomized on the basis of the municipality having the status of a town (the seven biggest municipalities with a population size more than 1600: Tábor, Sezimovo Ústí, Bechyně, Planá nad Lužnicí, Mladá Vožice, Chýnov and Jistebnice).

The parameters related to interaction terms are the most difficult to interpret. The interaction term indicates how one independent variable influences the effect that another independent variable has on the choice probability. Two types of interactions can be tested. The first type is an interaction between two or more characteristics of choice alternatives, here municipalities. For example, residential choices are influenced by the proportion of apartment houses in the municipality. The direction and the magnitude of the influence of this characteristic is moderated by the municipality population size.

The second type is the interaction between personal characteristics and the characteristics of the choice alternatives. The following example demonstrates how the job concentration in municipalities is evaluated differently by each age group. The age is represented by categorical variables transformed to dummy variables:

$I_{YA}(1)$ indicates the age of an individual 0–9 or 25–34, otherwise $I_{YA}(0)$;

$I_{MA}(1)$ indicates the age of an individual 10–24 or 35–54 years, otherwise $I_{MA}(0)$.

The dummy variable I_{MA} for the age group 55 and higher is a reference category and is therefore not expressed explicitly in the model.

Job concentration is represented by dummy variable I_{JC} : a municipality is considered as a job centre $I_{JC}(1)$ if the number of jobs located in the municipality covers at least 80% of its economically active inhabitants (16 out of 79 municipalities), otherwise $I_{JC}(0)$.

Having both categorical variables transformed to dummy variables, the interaction term can be expressed as:

$$\text{logit}(j) = \beta_{OA}I_{JC} + \beta_{YA}I_{JC}I_{YA} + \beta_{MA}I_{JC}I_{MA} = -0.284205I_{JC} + 0.476586I_{JC}I_{YA} + 0.454538I_{JC}I_{MA},$$

where β_{OA} indicates how the choices of the old age groups are influenced by the municipality being a job centre, β_{YA} indicates how this influence differs when the choice is made by a member of the young age group, and β_{MA} indicates how this influence differs when the choice is made by a member of the middle age group.

The resulting effects on the residential choice of a municipality being a job centre is:

$$\text{logit}(j) = -0.284205 + 0.476586 \times 1 + 0.454538 \times 0 = 0.192381$$

for an individual belonging to the young age group,

$$\text{logit}(j) = -0.284205 + 0.476586 \times 0 + 0.454538 \times 1 = 0.170333$$

for an individual belonging to the middle age group, and

$$\text{logit}(j) = -0.284205 + 0.476586 \times 0 + 0.454538 \times 0 = -0.284205$$

for an individual belonging to the old age group.

The β coefficients are related to the $\text{logit}(j)$. However, they can be easily transformed to e^β to represent the odds of choice probabilities. The municipality being a job centre changes the probability that it will be chosen by an individual belonging to the young age group $e^{0.192381} = 1.212132$ times and $e^{-0.284205} = 0.7526123$ times if the individual belongs to the old age group. This result indicates a significant impact of a personal characteristic on the choices.

4. Residential location choices analysed

The reviewed applications of residential location choice models provided an initial list of characteristics related to choice alternatives and individuals making the choice (IAURIF, THEMA 2004, 2005, 2007; Patterson, Kryvobokov, Marchal, & Bierlaire 2010; Vorel & Franke 2012, 2012; Waddell & Borning 2008). The listed characteristics supported by suitable data were analysed for their potential impact on the observed choices of individuals by means of discrete choice models.

The analyses were performed by experimental residential location choice models that were coded in Python programming language. The Open Platform for Urban Simulation (OPUS) open source libraries, together with the Biogeme open source libraries for estimating discrete choice models were used for the estimation and simulation runs of the experimental models ("Biogeme" 2013; Waddell 2013b). The data was stored in the PostGIS database and visualized by QGIS (PostGIS 2013, QGIS 2013).

4.1 The interaction between characteristics of present and future residential locations

The characteristics of the present residence, its neighbourhood and its proximity to the potential new residence are assumed to be significant for residential choices (Coulombel 2011; IAURIF, THEMA 2007). As the combination of characteristics is unique for each individual making a decision, their effect on decision making must be evaluated on an individual level. The interaction terms combine the characteristics of present and potential new residential locations. Out of all tested interactions between the characteristics of present and future residential location, only their proximity proved to be significant.

To operationalize the proximity term, the municipalities were sorted into seven sub-regions: Bechyňsko, Malšicko-Opařansko, Choustnicko, Mladovožicko, Chýnovsko, Tábořsko, Jistebnicko. The seven sub-regions were delimited by the Planning Analytical Documents



Fig. 1 Division of the Tábor microregion into seven sub-regions (MÚ Tábor, 2012).

2010 of ORP Tábor on the basis of micro-catchment areas and similarities of the social and natural characteristics of the municipalities (MÚ Tábor 2012). Delimitation of the sub-regions was aimed at aggregating the neighbouring municipalities into groups with distinctive characteristics. Municipalities that are located in the same sub-region are considered to be proximal municipalities, in the sense that their characteristics are more similar than the characteristics of non-proximal municipalities. It is assumed that the adjacency as well as the similarities of the municipalities in the same sub-region leads to social and emotional attachment of their inhabitants and to the

higher proportion of relocations that take place inside sub-regions (Coulombel 2011; Pacione 2009).

Two indicator variables I_s^e and I_s^a were created for each sub-region, where index s is replaced by a concrete sub-region identification (see table 1), where:

I_s^e is equal to one $I_s^e(1)$ when the present residential location of the individual is inside the sub-region, otherwise $I_s^e(0)$ and

I_s^a is equal to one $I_s^a(1)$ when the potential alternative residential location of the individual is inside the sub-region, otherwise $I_s^a(0)$.

If both indicator variables $I_s^e(1)$ and $I_s^a(1)$ are equal to one, then the residential move is realized within sub-region s , if $I_s^e(0)$ and $I_s^a(1)$, then the residential move to sub-region s is realized from another sub-region.

The interaction term $logit(j) = \beta_s I_s^e + \beta_s^* I_s^e I_s^a$ was then tested, where:

β_s represents the change of logit if sub-region S is selected by an individual;

β_s^* indicates the change of logit if the individual is living in the same sub-region.

The logit coefficients can be easily transformed to e^{β_s} and $e^{\beta_s^*}$ to represent the odds of choice probabilities.

Proximity is not evaluated in the general population model, because the use of higher level interaction terms, including age, sub-region and the characteristics of the choice alternatives could not be evaluated with the limited number of available observations.

In the stratified model, only the young age group and the middle age group sub-models evaluated the impact of proximity. The number of 962 observed residential choices made by individuals of the old age group (individuals older than 54 years) did not allow the interaction between the present and the potential new residential location $\beta_s^* I_s^e I_s^a$ to be evaluated, and only the main effect

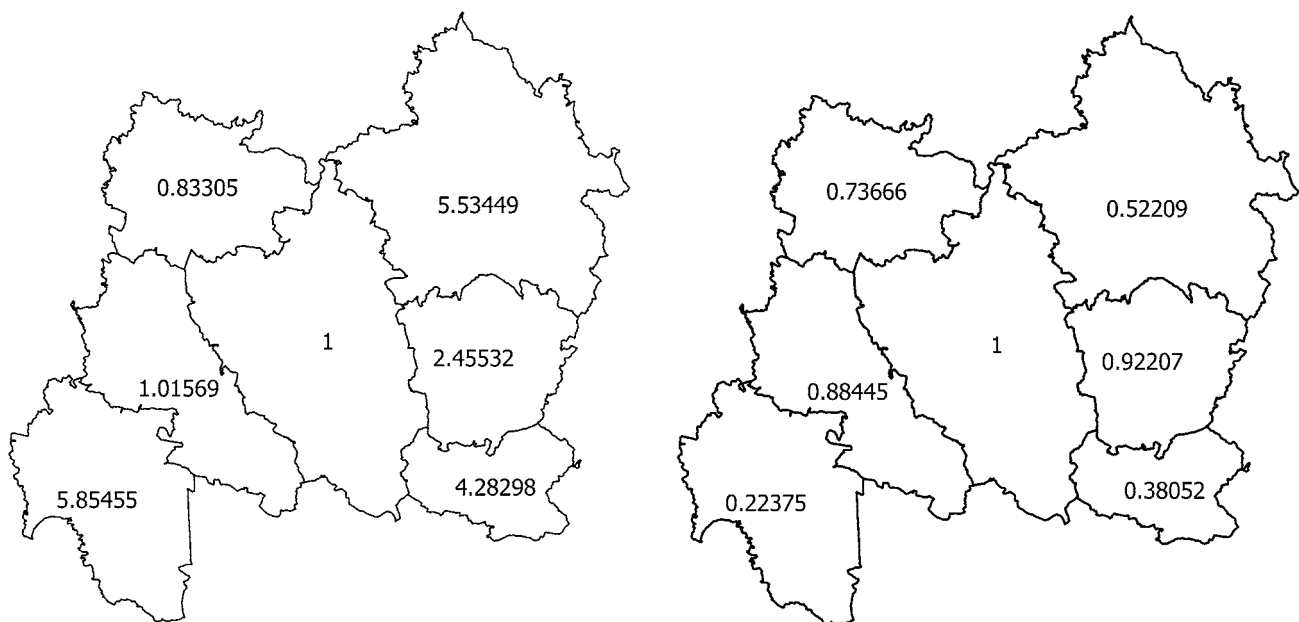


Fig. 2 The relative probability of a sub-region being chosen (Tábořsko sub-region is reference choice) for individuals in the young age groups. On the left are individuals living in the sub-region, on the right individuals living outside the sub-region.

– attractiveness of new residential location $\beta_s I_s^a$ – was evaluated.

The relative probability (chance) of a sub-region being chosen by an individual compared to the reference Tábor-sko sub-region was evaluated. The impact of proximity proved significant for the choices of the Bechyňsko, Choustecko and Vožicko sub-regions. The effect of proximity is strong enough even to reverse the evaluation of the municipalities in those sub-regions: the evaluation changes from negative, when it is made by residents of other sub-regions, to positive, when it is made by residents living in the evaluated sub-regions.

For example, given that an individual is a member of the young population living in the Bechyňsko sub-region, the probability that she or he will relocate inside this sub-region is $\exp(1.767) = 5.855$ higher than the probability that she or he will relocate to the reference Tábor-sko sub-region. For another individual living outside the Bechyňsko sub-region, the choice probability of the Bechyňsko sub-region is only $\exp(-1.497) = 0.224$ of the Tábor-sko sub-region choice probability. The Bechyňsko sub-region therefore has $5.855/0.224 = 26.14$ higher probability of being chosen by an individual already living in this sub-region than by an individual living in another sub-region.

The choices of municipalities in other sub-regions – Malšicko, Chýnovsko, Jistebnicko – are less dependent on the present location of the individual. This indicates that their self-containment is lower than the self-containment in the other three sub-regions.

All sub-regions have the highest relative probability of being chosen by their own inhabitants, with the exception of Jistebnicko for the young population. Attachment to sub-regions is weaker in the case of the young age groups. The reviewed literature suggests that the weaker attachment of young age groups could be due to their search for

a new job, usually a first job, and due to the formation of a new household. This usually leads to more distant relocation (migration) than with other age groups. However, the relocation of middle age population is usually caused by changing housing needs only, and there are much less strong incentives for more distant relocation (Coulombel 2011; Pacione 2009).

For example, the probability that a middle age individual living in the Bechyňsko sub-region will relocate inside the sub-region is $\exp(2.1124) = 8.26806$ times higher than the probability that the individual will relocate to the reference Tábor-sko sub-region. However, the probability is only 5.855 in the case of the young age group. An individual from the middle age group therefore has a $8.268/5.855 = 1.41$ higher probability of relocating inside the Bechyňsko sub-region than an individual from the young age group.

For the young groups, the Tábor-sko sub-region, which contains the biggest urban municipalities – Tábor, Sezimovo Ústí and Planá nad Lužnicí – is the second most attractive choice, while for the middle age groups the sub-regions adjacent to the Tábor-sko sub-region – the Malšicko, Chýnovsko, Jistebnicko sub-regions – are the second best choice. The relatively high employment and socializing opportunities in the Tábor agglomeration could explain the attractiveness of the Tábor-sko sub-region for the young age groups. Different factors, namely disposable land for development in municipalities adjacent to the Tábor sub-region, influence the residential choice of the middle age population.

4.2 Housing characteristics

The proportion of flats in apartment houses *BD* significantly influences the residential location choice in both

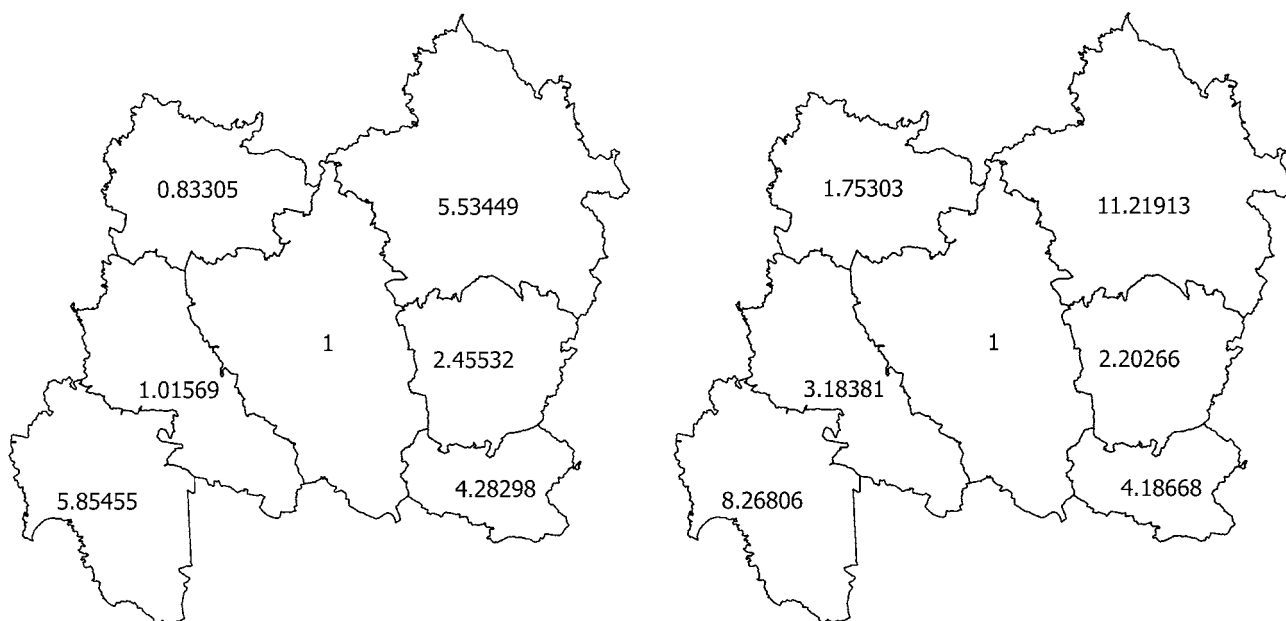


Fig. 3 The relative choice probability of a sub-region, in which the individual is already living (Tábor-sko sub-region is reference choice). On the left is young age group, on the right the middle age group.

models. In addition, the population size of a municipality significantly modifies the effect of this characteristic, so that the proportion of flats in apartment houses increases the choice probability of municipalities that are population centres $I_{PC}(1)$ and reduces the choice probability of municipalities that are not population centres $I_{PC}(0)$. Population centres are municipalities with the status of a town. In the Tábora micro-region, the population centres have a minimum population of 1600. Increasing the proportion of flats in apartment houses BD by 10% increases the probability of population centre choice 1.101 times, but decreases the choice probability 0.992 times if the municipality is not a population centre. This conclusion corresponds to the observed higher vacancy rate and lower price of flats in apartment houses in small, rural municipalities than in population centres.

The single population model revealed the significance of two additional housing characteristics. The number of flats built between 1999 and 2006 $NFLATS$ has a positive influence on the choice of a municipality. 10 new flats built between 1999 and 2006 increased the probability of municipality choice 1.756 times, but 100 new flats increased the probability only 3.083 times. This demonstrates that the effect of a unit increase of an independent variable is not necessarily always constant for all values of the variable.

A high proportion of vacant family houses $VACANCY$ has a negative influence on the residential location choice. An increase in the vacancy rate by 10% decreases the probability of municipality choice 0.915 times.

The quality of the flats, indicated by the proportion of the highest quality class (first class) did not prove to be significant in either of the two experimental models. The age structure of the housing stock³ was only partially significant, as only the periods of construction between 1920 and 1945, and between 1981 and 1990 proved to have a statistically significant, though rather weak, effect on the attractiveness of a municipality. The absence of the price of flats as a trade-off characteristic compensating for differences in the quality of flats could be a reason for the non-significance of the quality and age of flats. The prices of flats were not included due to unavailability, and consequently neither quality characteristics nor age characteristics were included in the experimental models.

In reality, it is an individual house or flat – and not a municipality – that is being chosen by the inhabitants. Aggregating housing characteristics on the level of a municipality leads to a loss of information about the local variability of the housing stock characteristics. The aggregated housing characteristics can then potentially correlate with the neighbourhood characteristics, causing multicollinearity and statistical insignificance of some housing characteristics. In order to model

residential choices on the level of individual houses and flats, it would be necessary to observe the characteristics and also the choices on the level of individual dwellings, as was demonstrated in applying the UrbanSim micro-simulation model in the San Francisco Bay area (Waddell 2013a). In the Czech Republic, there is no data available on choices of individual houses and flats, so it is not at present possible to model the choices on the level of individual houses and flats.

4.3 Neighbourhood characteristics

The evaluation of the natural, social and economic characteristics of a neighbourhood, as well as the public amenities, the land use and the percentage of the area of the municipality designated for development was made in terms of their effects on individual choices.

With regard to natural characteristics, the proportion of the area of the municipality covered by forest had a significant positive effect on choice probability. A one per cent increase in forest cover increases the choice probability of the municipality 1.011 times for the young population and 1.014 times for the middle age population. Other natural characteristics: proximity to water flows, average slope of the terrain, proportion of arable land, and proportion of areas of nature protection were not significant. This conclusion does not correspond to our expectation, or to the evidence presented in the reviewed literature (IAURIF, THEMA 2004, 2005, 2007; Patterson, Kryvobokov, Marchal, & Bierlaire 2010; Vorel & Franke 2012, 2012; Waddell & Borning 2008). Averaging the characteristics on the level of municipalities with average size 12.6 km² makes the model ignore the important part played by intra-municipality variation in natural characteristics.

With regard to social and economic characteristics: ethnic composition, income and household size were significant in most of the reviewed residential location choice models that have been applied in metropolitan regions (Coulombel 2011; IAURIF 2007; IAURIF, THEMA 2005; Waddell & Borning 2008). Unfortunately, data on income and ethnicity is not available with a sufficient level of detail in the Czech Republic. Data on household sizes is regularly provided by the general population census, but only in ten-year intervals. The 2011 census data was not available for the experimental models presented in this paper. Because of lack of data, the effect of income, ethnicity and household size characteristics on residential location choice could not be evaluated.

The level of economic activity is indicated by the number of jobs located in a municipality. In correspondence with the reviewed location choice models, both models indicate that higher economic activity in a municipality makes the municipality more attractive for residential use (IAURIF, THEMA 2004; Waddell & Borning 2008; Wegener 2011). An increase in the number of jobs in a municipality from 100 to 200 increases the choice probability

³ The age structure of the housing stock was indicated by the percentage of flats built in the periods: before 1919, 1920–1945, 1946–1960, 1961–1970, 1971–1980, 1981–1990 and 1991–2001.

1.266 times for the young population and 1.06 times for the old population. An increase from 900 to 1000 jobs increases the choice probability only 1.037 times for the young population and 1.009 times for the old population. This shows that the effect of a marginal change in the number of jobs is not constant.

The interaction between the economic activity in a municipality and the age of the individual making the choice proved to be significant in the single population model. A municipality that is a job centre $I_{JC}(1)$ (a municipality having more than 0.8 jobs per one economically active resident, i.e. 20% of all municipalities in the micro-region) has a 1.212 times higher probability of being chosen by the young population, 1.1857 times higher for the middle age population and 0.756 times lower for the old population than a municipality that is not a job centre $I_{JC}(0)$.

The natural logarithm of the number of all public services (nursery schools, basic and secondary schools, health-care facilities, cultural facilities, social facilities) in municipalities *PUBS* proved to have a statistically significant effect on the choice of a residential location. The stratified model indicates that the presence of public services in a municipality has a positive influence on its attractiveness for all age groups, especially for the old population. For example, an increase from 10 to 11 in the number of public services in a municipality increases the choice probability 1.023 times for the young population and 1.048 times for the old population. However, an increase in public services from 100 to 101 increases the choice probability only 1.002 times for the young population and 1.004 times for the old population. The effect of a marginal change in the number of public services is not constant, but is generally expected to follow the law of diminishing marginal return.

The $PUBS \times PSIZE$ interaction term of the general population model indicates that the positive effect of public services is slightly reduced with increasing population size of the municipality causing an increased number of inhabitants per public service.

Out of specific public services, only the presence of a basic school proved to have a significant effect. The presence of at least one basic school $I_{BS}(1)$ significantly increases the attractiveness of a municipality. The interaction term $I_{BS} \times I_{FG}$ in the single population model indicates that the importance of the characteristic for the choice paradoxically decreases in the case of fast-growing municipalities $I_{FG}(1)$ with a bigger than 4% growth in the housing stock between 1999–2006. The presence of a basic school in these municipalities increases the choice probability only 1.017 times, while the choice probability increases 1.288 times for a municipality with a slowly growing housing stock $I_{FG}(0)$. The inhabitants choosing the fast growing municipalities probably anticipate the location of a basic school in the future, or make use of the relatively good accessibility to basic schools in the nearby Tábör municipality.

Concerning access to public transportation, spatial proximity to a railway station I_{DR} has a positive effect on all age groups, the highest positive effect being on the middle-age population. For the young age group, a municipality with an average distance to a railway station less than 3.5 km $I_{DR}(1)$ has a 1.42 times higher probability of being chosen than a municipality with a longer average distance to a railway station $I_{DR}(0)$. The proximity of a railway station increases the probability of the choice 1.69 times for the middle-age population, and only 1.36 times for the old age population in stratified model. In the general population model, the choice probability is 1.18 times higher for all age groups together.

Unlike proximity to a railway station, the number of bus stops in a municipality did not prove to be significant. The most probable reason is the relatively even spatial distribution of bus stops and the resulting low variance across municipalities.

The supply of land designated by the land use plan for urban development *DEVLAND* was confirmed by both models to be a significant factor for residential choice. A one per cent increase in the area of the municipality designated for development increases the choice probability of the municipality 1.004 times for the young age group, 1.024 times for the middle age group, but only 1.00042 times for the old population in stratified model. These effects are statistically insignificant for the young population and for the old population. The land use mix, although generally considered to have a significant impact on residential choice, did not prove to be significant at municipality level. The variability of the land use composition probably needs to be captured on a much finer scale than on the level of municipalities with an average area of 12.6 km².

4.4 Overall accessibility

Access to regular work, education and shopping facilities are generally considered to have a significant influence on residential location choice. In an ideal case, the access should be evaluated by a transportation model in terms of transportation time or transportation cost. As no such transportation model has been implemented in the Tábör micro-region, the accessibilities were measured only in terms of road distance.

In reality, each individual has a unique action space, resulting from the location of her or his individual activities in the territory. The decision making of each individual should therefore be analysed with regard to her or his individual action space. Because data on the co-location of activities on an individual level is not usually available, general accessibility to activities is used instead. The ILUTE and ILLUMASS micro-simulation models are the only exceptions, being based on purpose-made surveys (activity logs) of the daily activities of the residents (Salvini & Miller 2005; Wegener 2005). In the case of the experimental models presented here,

single general accessibility to work and services activities was tested.

To cope with the high collinearity between work, services and housing characteristics, services were selected as the type of activity that best represents the importance of a municipality for the inhabitants in the Tábora micro-region. The municipalities were ranked according to the number of public services: the town of Tábora as the primary urban centre in the micro-region had 285 public services, while Bechyně, Opařany, Chotoviny, Mladá Vožice, Chýnov, Sezimovo Ústí and Planá nad Lužnicí, as secondary urban centres, had more than 13 but fewer than 49 public services. The effect of the distance by road to both the primary urban centre and the secondary urban centres was tested. Only the average road distance to the town of Tábora D_{CENTRE} in the general population model had a significant effect on residential choice. An increase of ten kilometres in road distance from Tábora decreases the probability of municipality choice 0.51 times, everything else being equal. The fact that this characteristic is not significant in the stratified model implies that accessibility measured by road distance is somewhat weaker predictor for residential choice than other characteristics, such as the distance of the residential move and neighbourhood characteristics.

Distance thresholds to important roads (highways, motorways and first-class roads) as another accessibility measure did not prove to be significant for most of the tested models, and when they were significant they had only a small impact on the choice. These characteristics were therefore not included in the final models.

The weak effects of global accessibility factors correspond to the generally accepted thesis that when access to regular activities in the territory is reasonably good, accessibility as a location factor is not decisive for residential choice.

5. An evaluation of the experimental models

Various indicators can be applied to judge the quality of simulation models. Here, four indicators: McFadden R^2 , mean absolute percentage errors (MAPE), relative

errors (RE) and the individual choice success rate (ICSR) were applied.

5.1 McFadden R^2

McFadden R^2 represents the proportion of variance explained by the model. It is equivalent to the coefficient of determination used in linear regression. McFadden R^2 is defined as the ratio between the log likelihood (LL) of the estimated model and the LL of the base model (Ben-Akiva & Lerman 1985; Liao 1994; Train 2009):

$$R^2 = 1 - (LL_{estimated\ model} / LL_{base\ model}) \quad (6)$$

The base model assumes that no characteristics of the alternatives have an impact on the choice process, and that all alternatives therefore have an equal probability of being chosen.

The values for McFadden R^2 in the general population model and in all three sub-models of the stratified model are in the range from 0.30 to 0.33 (see table 3), and they are comparable to values from 0.20 to 0.32 of the reviewed location choice models (IAURIF, THEMA 2004 2005, 2007; Patterson et al. 2010; Vorel & Franke 2012, 2012; Waddell & Borning 2008).

5.2 Mean absolute percentage error (MAPE)

The other three indicators of model quality focus on differences between the simulated choices and the observed choices. The simulated choices are realized by Monte Carlo simulation based on the probabilities calculated by the MNL models. The number of simulated choices of each municipality is therefore stochastic variable. To analyse its variance, it is necessary to run the simulation a number of times. Here, 100 simulation runs were performed and the standard deviations of the relative errors between the number of observed and simulated choices of individual municipalities were evaluated. The average number of choices was used to evaluate the MAPE and RE indicators.

Tab. 3 Indicators of the quality of the experimental simulation models.

Characteristics of municipality	The sub-models of stratified model			General population model
	1	2	3	
Number of observations	4586	4555	962	10103
Log-likelihood	-13364	-13487	-2837	-30587
Null Log-likelihood	-20038	-19902	-4203	-44144
McFadden R^2	0.33198	0.32127	0.32131	0.30672300000
Suggested t-value	2.90357	2.90240	2.62088	3.03654206103
Akaike's Information Criterion (AIC)	26771	27017	5705	61208
ACE (sub-models)	0.55200	0.56400	0.90300	0.10200000000
ACE	0.16600			

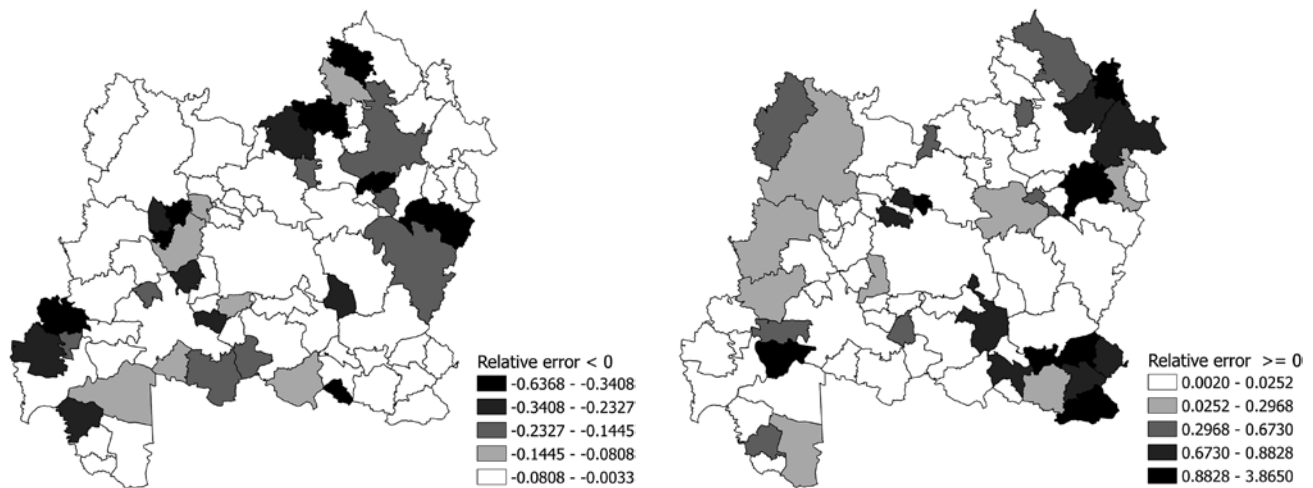


Figure 4. The relative error RE of the simulated choices for each municipality in the Tábor micro-region: on the left are negative errors indicating overestimated municipalities, on the right are positive errors indicating underestimated municipalities.

The mean absolute percentage error ($MAPE$) averages the relative differences between the number of observed O_n and simulated S_n choices of individual municipalities n over all municipalities in the choice set $n \in N$ (IAURIF, THEMA 2004):

$$MAPE = \frac{1}{N} \sum_{n=0}^N \frac{|O_n - S_n|}{O_n} \quad (7)$$

The $MAPE$ of the general population and stratified model is equal to 0.102 for the general population model and 0.166 for the stratified model. Both values are similar to $MAPE$ 0.12 of the UrbanSim household location choice model applied in the Puget Sound region in Washington, USA⁴ (Waddell & Borning 2008).

5.3 Relative errors (RE)

Unlike $MAPE$, RE represents the relative difference between the number of simulated S_n and observed O_n choices of individual municipalities $n \in N$:

$$RE = (O_n - S_n) / O_n \quad (8)$$

In 67% of the municipalities, the maximum relative error ranged from -29.9% to 74.4% for the general population model and from -25.9% to 68.8% for the stratified model.

The RE of both models can be compared to the RE of the reviewed residential location choice models: the RE in the Puget Sound application was in the range from -22% to 124% , with 9 alternative choices. The RE in the Lyon application, with 777 Transportation Analysis Zones as choice alternatives, was in the range from -2% to $+2\%$

⁴ The comparison of the models has to take into consideration differences in the number of alternatives in the choice set: 9 alternatives for the Puget Sound application, and 79 alternatives for the experimental models presented here.

in 11% of the choices, from -5% to $+5\%$ in 29% of the choices, and from -10% to $+10\%$ in 67% of the choices. In the Paris application, the absolute RE was smaller than 15% in 67% of the choices (IAURIF 2007).

The relatively high RE of the experimental models presented here is due to a) the overall small number of observed choices compared to the reviewed applications typically applied on the scale of metropolitan regions, b) the uneven distribution of the number of choices across the municipalities. For example, the residential location choice model in Paris was based on 5.893 million observed relocations of individuals between 1991 and 2001 (IAURIF, THEMA 2007), in comparison with 10,103 observed residential choices made between 2001 and 2011 in the Tábor micro-region. In addition, the number of choices is unevenly distributed among the municipalities in the Tábor micro-region: 11 out of 79 municipalities were selected less than 10 times during the 10-year observation period. As is indicated in following scatterplot in figure 5, municipalities with a number of observed choices lower than 100 are associated with high RE values. This will lead to the aggregation of least frequent choice alternatives in future versions of the models.

5.4 Individual choice success rate ($ICSR$)

The age characteristic entering the choices in the general population model and the stratified model is expected to improve the prediction of choices on the level of individual actors. To test the improvement, the proportion of individuals having the simulated choice identical with the observed choice is measured by the individual choice success rate ($ICSR$). The two models were compared with the random model, which does not include the age characteristic. The $ICSR$ for the single population model was 0.112, while for the stratified model the value was 0.118, but it was only 0.0126 for the random model. This simple demonstration shows that the inclusion of

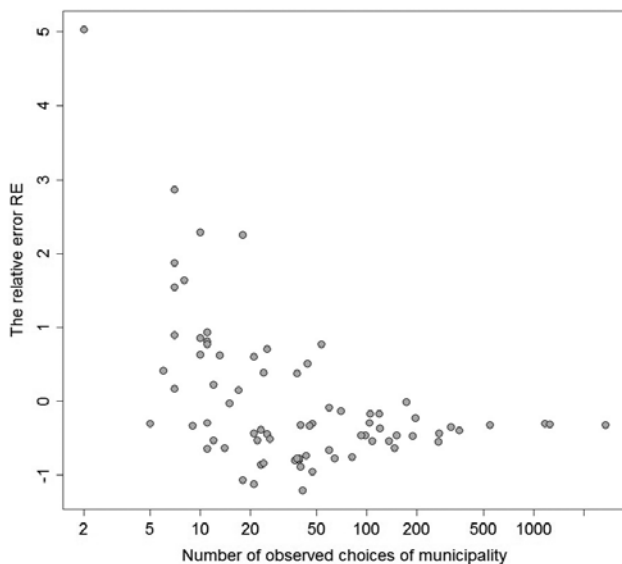


Fig. 5 The relation between the number of observed choices of a municipality and the relative errors *RE* of simulated choices of the municipality.

just a single personal characteristic in the micro-simulation model leads to a significant improvement in the predictions of individual choices.

5.5 Stochasticity of the model

The micro-simulation models presented in this paper are principally stochastic. Stochastic variation is measured by the standard deviation of relative errors *RE* on a sample of 100 simulation runs. The scatterplot below in figure 6 indicates that the stochastic variation is indirectly proportional to the number of simulated choices.

The stochastic variation depends not only on the number of simulated choices, but also on the number of alternatives in the choice set. The greater the number of alternatives, the greater the stochastic variation, everything else being equal (Wegener & Spiekermann 2011). The solution to the problem of stochastic variation is presented in the following section.

6. Conclusion

The experimental models presented here demonstrate the applicability of the micro-simulation approach where limited data is available. Two major application areas demonstrated in the paper are in analysing the factors influencing residential choice, and in simulating residential choices.

The models proved to be useful for analysing the factors that influence observed residential location choice. To study the residential choices, first a list of factors that are assumed to influence residential location choice was compiled on the basis of a review of applications of the UrbanSim model, and then available data on the factors was searched for. Extensive statistical testing proved the

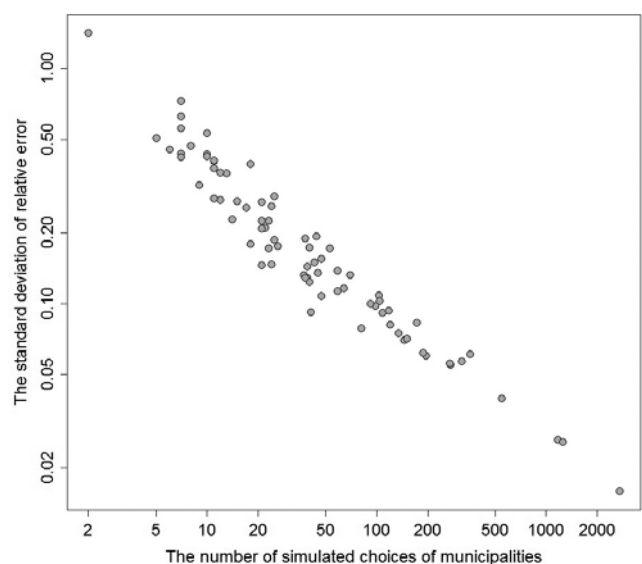


Fig. 6 The relation between stochastic variation and number of simulated choices.

significant effect of several factors. Apart from the main effects of the factors, more complex interdependences were also identified, namely: an interaction between the characteristics of present and potential new residence, an interaction between the characteristics of the municipality and the impact of personal characteristics on the decision making.

The experimental applications of both stratified and general population models did not provide a clear judgement on the superiority of either of the models. The models offer two comparable approaches for analysing how personal characteristics influence decision making: a) a comparison of the sub-models that represent the choices made by selected population strata or b) by building a single model encompassing all the population and then testing the interdependence between the characteristics of particular choices and the personal characteristics of the individuals making the choices.

While the micro-simulation approach proved to be useful in analysing residential location choice factors, the influences have to be interpreted with care as several important factors were not applied due to a lack of suitable data.

The socio-economic and demographic factors influencing decision making are not well covered by suitable disaggregated data. Households are considered to be the decision-making entity in most reviewed residential choice models. In the Czech Republic, however, there is a lack of data on household characteristics that is disaggregated to the level of individual households. There is a lack of data on the household level for simulating residential mobility, and also a lack of data on household relocations, income, mobility, car ownership, as well as household demographic characteristics and their transitions (marriage, divorce, birth of children, children leaving the household). In addition, there is a lack of statistics in a

form that would enable synthetic populations of households to be created. Therefore, only relocation of individuals can be simulated at present.

The price and the availability of houses and flats on the housing market are other important factors influencing residential choice that are not covered by disaggregated data. In addition, data on occupancy of the housing stock and household tenure are not available at the moment.

Data on residential choices is aggregated to municipalities. This might raise concern about the proper representation of residence and neighbourhood characteristics. Aggregation of individual residence characteristics to municipality level may hide large inter-locality variances, and may make them interfere with the neighbourhood characteristics in multi-nominal logit models. This could be one of the reasons for the weak significance of some of residence characteristics. Unfortunately, no data on residential choices related to individual houses and flats is currently available in the Czech Republic.

The second area of applying models that has been demonstrated is the simulation of residential location choices. Predictions of decision making on the level of individual actors can in principle be only probabilistic, which causes high stochastic variation in the model results. The choices are simulated using the Monte Carlo techniques on the basis of choice probabilities predicted by multi-nominal models. The stochastic variation could easily be mitigated by making the choices in direct proportion to these probabilities instead of employing probabilistic choice process, but then the unknown factors that influence the decision making of individuals would remain hidden. However, if stochasticity is admitted, the reliability of the model results can be assessed.

A pragmatic and theoretically sound approach to stochastic variation is to scale it according to the purposes for which the model was built. This approach follows the trade-off between the stochastic variation of the model outcomes, on the one hand, and the number of choice alternatives and the number of simulated choices, on the other. This trade-off implies that stochastic variation can be decreased to an acceptable level by spatial aggregation of alternative choices, or by increasing the number of simulated choices.

Based on these conclusions, several recommendations for further research can be made:

- households and not individuals should be represented as decision making entities; for this purpose, synthetic populations of households could be derived from the general population census data, from existing surveys (EU-SILC), and from ad-hoc household surveys;
- market prices and data on the occupancy of buildings should be collected to increase the validity of residential choice models;
- the number of observations should be increased by expanding the area of analysis, or by selecting areas with a dynamic residential mobility pattern.

As most data on individuals is not made public due to privacy issues, these objectives can only be achieved with the involvement of institutions that provide data, i.e. the Czech Statistical Office, the Czech Office for Surveying, Mapping and Cadastre, the tax offices, and the Czech Social Security Administration.

Acknowledgements

Funding from the Grant Agency of the Czech Republic within the framework of grant P104/12/1948 is gratefully acknowledged.

REFERENCES

- 2011 POPULATION AND HOUSING CENSUS (2013): General population census 2011. Available at <http://www.scitani.cz/> (accessed 18 October 2013).
- AXHAUSEN, K. W. (2005): A Dynamic Understanding of Travel Demand: A Sketch. In M. L. Gosselin (Ed.), *Integrated Land-Use Transportation Models: Behavioral Foundations*. Oxford: Elsevier.
- BEN-AKIVA, M., & LERMAN, S. (1985): *Discrete Choice Analysis: Theory and Application to Travel Demand*. The MIT Press.
- BIOGEME (2013): Available at <http://biogeme.epfl.ch/> (accessed 15 October 2013).
- COULOMBEL, N. (2011): Residential choice and household behaviour: State of the Art. *SustainCity Working Paper, 2.2a*, École normale supérieure de Cachan.
- IAURIF, T. (2007): *Modèle dynamique de simulation de l'interaction Urbanisation-transports en Région Ile-de-France. Application à la Tangentielle nord. Rapport final de la 2ème phase*. IAURIF, Université de Cergy Pontoise, THEMA.
- IAURIF, THEMA (2004): *Modèle dynamique de simulation de l'interaction Urbanisation-transports en Région Ile-de-France. Application à la Tangentielle nord. Rapport final de la 1ère année*. IAURIF, Université de Cergy Pontoise, THEMA.
- IAURIF, THEMA (2005): *Modèle dynamique de simulation de l'interaction Urbanisation-transports en Région Ile-de-France. Application à la Tangentielle nord. Rapport intermédiaire de la deuxième phase année*. IAURIF, Université de Cergy Pontoise, THEMA.
- IAURIF, THEMA (2007): *Modèle dynamique de simulation de l'interaction Urbanisation-transports en Région Ile-de-France. Application à la Tangentielle nord. Rapport final de la 2ème phase*. IAURIF, Université de Cergy Pontoise, THEMA.
- LEE, D. B. (1975): Requiem for large-scale models. *ACM SIGSIM Simulation Digest*, 6(2–3), 16–29. doi:10.1145/1102945.1102950
- LIAO, T. F. (1994): *Interpreting probability models: logit, probit, and other generalized linear models*. Thousand Oaks, Calif: Sage.
- MŮ TÁBOR (2012): 2. úplná aktualizace územně analytických podkladů pro správní území obce s rozšířenou působností Tábor. Městský úřad Tábor – odbor územního rozvoje. Available at <http://www.taborcz.eu/uap-rok-2012/ds-1570/archiv=0>.
- PACIONE, M. (2009): *Urban geography: a global perspective* (3rd ed.). London, New York: Routledge.
- PATTERSON, Z., KRYVOBOKOV, M., MARCHAL, F., & BIERLAIRE, M. (2010): Disaggregate models with aggregate data:

- Two UrbanSim applications. *Journal of Transport and Land Use*, 3.
- POSTGIS (2013): PostGIS — Spatial and Geographic Objects for PostgreSQL. Available at <http://postgis.net/> (accessed 15 October 2013).
- QGIS. (2013): QGIS. Available at <http://qgis.org/en/site/> (accessed 15 October 2013).
- SALVINI, P., & MILLER, E. J. (2005): ILUTE: An Operational Prototype of a Comprehensive Microsimulation Model of Urban Systems. *Networks and Spatial Economics*, 5(2), 217–234. doi:10.1007/s11067-005-2630-5
- SIMMONDS, D. (2010): The DELTA Residential Location Model. In F. Pagliara, J. Preston, & D. Simmonds (Ed.), *Residential Location Choice: Models and Applications* (Advances in Spatial Science). Berlin, Heidelberg: Springer.
- SIMMONDS, D., & FELDMANN, O. (2007): Advances in integrated urban/regional land use/transport modelling using the Delta package. Berkeley, USA: World Conference on Transport Research Society.
- SIMMONDS, D., CHRISTODOULOU, A., FELDMAN, O., & MCDONALD, M. (2011): Latest Developments in the SIM-DELTA Model: Investigation of Stochastic Variation and Development of a Disaggregate Car Ownership Model. Lake Louise, Alberta, USA: 12th International Conference on Computers in Urban Planning and Urban Management (CUPUM).
- STRAUCH, D., MOECKEL, R., WEGENER, M., GRÄFE, J., MÜHLHANS, H., RINDSFÜSER, G., & BECKMANN, K.-J. (2005): Linking Transport and Land Use Planning. In *GeoDynamics* (p. 295–311). CRC Press, Taylor & Francis Group.
- TRAIN, K. E. (2009): *Discrete Choice Methods with Simulation*. Cambridge University Press.
- VOREL, J., & FRANKE, D. (2012): *UrbanSIM: návrh metodiky pro založení, správu a aktualizaci vstupních data pro modelovací systém UrbanSim/OPUS*. Praha: Útvar rozvoje hl. m. města Praha.
- VOREL, J., & GRILL, S. (2013): Land Use Modelling in the Czech Republic: An Experimental Approach with Limited Data. In *PROCEEDINGS OF CUPUM 2013* (Vol. 13). Utrecht, Netherlands: Springer-Verlag.
- WADDELL, P. (2002a): *UrbanSim: Modeling Urban Development for Land Use, Transportation and Environmental Planning*. University of Washington.
- WADDELL, P. (2013a): Draft Technical Documentation: San Francisco Bay Area UrbanSim Application. Institute of Urban And Regional Development University of California, Berkeley.
- WADDELL, P. (2013b): *UrbanSim*. UrbanSim. Available at <http://www.urbansim.org/Main/WebHome> (accessed 15 October 2013).
- WADDELL, P., & BORNING, A. (2008a): Documentation: Puget Sound UrbanSim Application. University of Washington. Available at http://www.psrc.org/assets/2830/CUSPA_UrbanSim_Doc.pdf (accessed 15 October 2013).
- WADDELL, P., WANG, L., CHARLTON, B., & OLSEN, A. (2010): Microsimulating parcel-level land use and activity-based travel: Development of a prototype application in San Francisco. *Journal of Transport and Land Use*, 3.
- WEGENER, M. (2005): *Urban Land-Use Transportation Models*. In GIS, spatial analysis, and modeling. Redlands, Calif.: ESRI Press.
- WEGENER, M. (2011): *The IRPUD Model*. Spiekermann & Wegener Urban and Regional Research (S&W). Available at http://www.spiekermann-wegener.com/index_e.htm.
- WEGENER, M., & SPIEKERMANN, K. (2011): From Macro to Micro – How Much Micro si too Much? *Transport Reviews*, 31(2), 161–177.

RESUMÉ

Modely výběru místa bydliště: mikrosimulační přístup

Mikrosimulační modely popisují rozhodovací procesy na úrovni jedinců a umožňují tak hodnotit vliv celé řady vnějších faktorů a osobních charakteristik jedinců na jejich rozhodování. Tento příspěvek ověřuje použití mikrosimulačních multinomálních logitových modelů na zkoumání faktorů stěhování obyvatel v územním obvodu obce s rozšířenou působností Tábor. Na základě pozorovaných stěhování obyvatel v letech 2001–2011 byl ověřován potenciální vliv velkého množství faktorů a pro každou kombinaci obce a jedince byla odvozena pravděpodobnost výběru nového místa bydliště.

Experimentální použití mikrosimulačních modelů v této podobě přineslo řadu zjištění: a) statisticky významnou se prokázala být interakce věku jedince s některými charakteristikami místa bydliště; vliv řady dalších významných osobních charakteristik jedinců, zejména vzdělání a příjmu, nemohly být testovány z důvodu nedostupnosti vhodných dat; b) nebyla prokázána významná interakce mezi charakteristikami místa současného a nového bydliště s výjimkou jejich vzájemné polohy: obyvatelé výrazně preferují stěhování v rámci sousedních obcí a tyto preference jsou ovlivněny jejich věkem; c) mikrodata o stěhování obyvatel nezachycují příslušnost jedinců k domácnostem a nebylo proto možné testovat vliv charakteristik domácností na výběr lokality bydliště; předpokládaná vzájemná podmíněnost rozhodování členů domácností byla ověřena vytvořením modelů pro specifické věkové skupiny dle jejich předpokládané příslušnosti k domácnostem; d) hodnoty ukazatelů průměrné absolutní procentuální chyby (MAPE) a procentuální chyby za jednotlivé obce jsou srovnatelné s obdobnými aplikacemi mikrosimulačních modelů v zahraničí; e) procentuální odchylky simulovaných a pozorovaných hodnot vztažené k jednotlivým obcím jsou dle očekávání nepřímo úměrné počtu pozorovaných stěhování obyvatel těchto obcí; příliš nízký počet pozorování v některých obcích výrazně snižuje spolehlivost predikce; f) využití osobních charakteristik jedinců v modelu významně zvýšilo míru shody pozorovaných a simulovaných výběrů na úrovni jedinců; g) spolehlivost predikovaného počtu stěhujících se obyvatel na úrovni obcí je ovlivněna vysokou mírou stochastičity, zejména u obcí s populací menší než sto obyvatel; h) úplné využití potenciálu mikrosimulačních modelů je omezené nedostupností podrobných a spolehlivých dat, zejména sociodemografických údajů o domácnostech a údajů o nemovitostech a jejich blízkém okolí.

Jakub Vorel
Czech Technical University in Prague
Faculty of Architecture, Department of Spatial Planning
Thákurova 7
166 29 Prague 6
Czech Republic
E-mail: kubavorel@gmail.com