International Journal of Urban Sciences

ISSN: 1226-5934 (Print) 2161-6779 (Online) Journal homepage: https://www.tandfonline.com/loi/rjus20

Understanding residential location choices: an application of the UrbanSim residential location model on Suwon, Korea

Jangik Jin & Hee-Yeon Lee

To cite this article: Jangik Jin & Hee-Yeon Lee (2018) Understanding residential location choices: an application of the UrbanSim residential location model on Suwon, Korea, International Journal of Urban Sciences, 22:2, 216-235, DOI: 10.1080/12265934.2017.1336469

To link to this article: https://doi.org/10.1080/12265934.2017.1336469

Published online: 05 Jun 2017.

INTERNATIONAL JOURNAL OF URBAN SCIENCES
2018, VOL. 22, NO. 2, 216–235
https://doi.org/10.1080/12265934.2017.1336469

Understanding residential location choices: an application of the UrbanSim residential location model on Suwon, Korea

Jangik Jin[a] and Hee-Yeon Lee[b]

[a]Department of Civil and Environmental Engineering, University of Wisconsin-Madison, Madison, WI, USA; [b]Graduate School of Environmental Studies, Seoul National University, Seoul, Korea

ABSTRACT

The residential location choice model is an effective tool to analyze the actual household demand for housing and better living environments, and many researchers have developed various residential location choice models. In this study, a residential location choice model using a discrete choice modelling framework within UrbanSim is applied to Suwon, Korea with the following aims: (1) to investigate factors affecting residential location choice in Suwon, (2) to forecast changes in household residential locations, and (3) to derive policy implications for the local housing market. An extensive database of parcels, households, jobs, land prices, and transportation networks is geocoded on the basis of grid cells that measure 150×150 metres. The estimation results show that access to employment opportunities, the ratio of housing cost to income, mixed land use, and the year that housing was built are important factors in determining household residential locations in Suwon. In addition, different age and income groups have different residential location preferences. UrbanSim, a highly disaggregated microsimulation model, is employed to forecast changes in household residential locations using the estimation results of the residential location choice model. These suggest that different income groups show different migration patterns.

ARTICLE HISTORY
Received 22 August 2016
Accepted 24 May 2017

KEYWORDS
Residential location choice; land use and transportation; discrete choice model; UrbanSim; Suwon

1. Introduction

Many researchers have developed various residential location choice models that are effective tools to analyze the actual household demand for housing, transportation, and living environments (Guo & Bhat, 2007; Waddell, 2006). According to urban economic theory, household residential location choices are a function of a wide range of housing and location attributes that are differentiated by a variety of household characteristics (Rosen, 1974; Sermons & Koppelman, 1998). This differentiation induces or reflects the relative importance of numerous attributes to different types of households. Among the various factors that influence household residential location choices, transportation accessibility has long been recognized as one of the most important factors in explaining residential location decisions. Particularly, several scholars have argued that understanding of

CONTACT Jangik Jin jjin8@wisc.edu Department of Civil and Environmental Engineering, University of Wisconsin-Madison, 1415 Engineering Drive, Madison, WI 53706, USA

the relationship between land use and transportation is essential to investigate urban growth patterns as well as household residential location decisions (Handy, 2005; Waddell, 2002). Unfortunately, the impact has rarely been quantified in an integrated perspective, and there also remains the challenge to determine the factors and measure their impact in a local context.

It is conventional wisdom that land use and transportation are strongly connected (Hanson, 1995). Several previous studies have attempted to build a model considering the relationship between land use and transportation (Hunt, Kriger, & Miller, 2005). The integrated land use and transport models allow researchers to anticipate system responses to new policies, preference functions, economic conditions, and other strategies. However, although there are strong interdependencies between land use and transportation, the planning of both have traditionally been compartmentalized and separated into different agencies (Voigt, Troy, Miles, & Reiss, 2009). Since the 1980s, urban scholars have started to argue that these interdependencies and the plans for them should be considered in an integrated fashion (Cervero, 2003; Giuliano, 1989), and efforts for developing integrated land use and transportation models have increased.

UrbanSim – a highly disaggregated and integrated land use and transportation model – has increasingly gained popularity among various dynamic urban growth models. Through the spatial allocation of household and employment locations, UrbanSim provides information on the factors affecting residential location decisions as well as forecasts land use changes for a certain area. Through a disaggregate frame of the analysis at the gridcell level measured 150 by 150 metres, it enables researchers to not only examine the determinants of residential location choices, but also suggest better policy implications for sustainable land development (Joo, Hassan, & Jun, 2011). Since the model was developed, it has been applied to many American and European growing cities – for example, Eugene-Springfield (Waddell, 2002), San Francisco (Waddell, Wang, Charlton, & Olsen, 2010), Austin (Kakaraparthi & Kockelman, 2011), Zurich (Lochl & Axhausen, 2010), Brussels (Patterson, Kryvobokov, Marchal, & Bierlaire, 2010), and Lyon (Kryvobokov, Kercier, Bonnfous, & Bouf, 2013; Patterson et al., 2010) – in order to suggest policy implications for spatial planning by providing simulated results of the growth patterns of population and employment as well as real estate values.

Like in the West, many cities in Korea also have experienced rapid urbanization in recent years. As a result, numerous urban problems such as housing shortages, declines in downtown areas, traffic congestion, and losses of urban open spaces have increased. Although policy makers and planners have established a variety of policies to solve these problems, especially the housing-related issues, most of these have failed to meet the demand of housing consumers (Kim & Ji, 2007). One of the reasons for this is that supply-oriented policies do not consider the preferences of actual subjects with housing demands and even disregard the fact that most of the housing demanders tend to choose their living environments based on their social and economic conditions. As urban scholars have stressed the importance of interaction between land use and transportation, an integrated approach considering the relationship between land use and transportation should be necessary to understand the needs of housing consumers as well as suggest more appropriate policies for housing and sustainable urban growth (Handy, 2005). Particularly, investigation of the determinants of household residential locations through a microsimulation model is helpful to obtain a better understanding of household preferences in residential location decisions.

Taking this perspective, the objective of this study is to make an extensive analysis for assessing the extent to which transportation and other factors affect residential location decisions. Particularly, we focus on the city of Suwon, which is one of the fastest-growing cities in Korea. We apply the residential location choice model within UrbanSim to quantify the effects of the several components affecting household residential location decisions as well as forecast their changes.

The next section is an overview of UrbanSim, and then, we review the residential location choice model for establishing our empirical model. Section 4 describes our study area and data development. Section 5 provides the estimation results of the residential location choice model with differentiated household characteristics. We also discuss the utility of residential locations and the simulation results of the residential location changes. In Section 6, we summarize and discuss our findings and then suggest policy implications.

2. Overview of UrbanSim

For the last several decades, researchers have developed integrated land use and transportation models that can predict the changes of land use and transportation system as a means to evaluate policies using mathematical, statistical, and logical methods. Specifically, almost twenty models have been developed and reviewed by many researchers (Hunt et al., 2005; Iacono, Levinson, & El-Geneidy, 2008; Miller, Kriger, Hunt, & Badoe, 1998; Wegener, 1995). As described by Lemp, Zhou, Kockelman, and Parmenter (2008), these can be broadly categorized into four types: the gravity allocation model, the cellular automata model, the spatial input-output model, and the discrete simulation model.

In the gravity allocation model, transportation accessibility is an essential factor in spatial distribution of households and employment opportunities. Although this model considers conditions about jobs, households, and land use at the zone level, it overlooks some of influential factors such as price adjustments, geographical conditions, and zoning restrictions (Hunt et al., 2005). The cellular automata model represents many aspects of the dynamic and complex land use systems. For example, the SLEUTH model reflects the dynamic system of land use including slope, land use type, urban extent, transportation and hill shade (Clarke, Hoppen, & Gaydos, 1997). However, it does not provide the process of changes in individual behaviours such as employment and household movement patterns.

Through the discrete choice theory, spatial input-output models such as TRANUS (De la Barra, Perez, & Vera, 1984), PECAS (Hunt & Abraham, 2005), and RUBMRIO (Kockelman, Jin, Zhao, & Ruiz-Juri, 2005) enable researchers to predict spatial and economic interactions of employment and household sectors across zones. The discrete choice approach is based on a microsimulation framework at the highly-disaggregated level. UrbanSim is designed to simulate residential and employment location choices while forecasting changes in patterns of future urban land use. UrbanSim was first developed by Waddell (2002), and then has been reviewed as a very useful model not only because it can integrate numerous aspects affecting land use changes, but also because it can perform a scenario analysis to address long-term planning issues (Voigt et al., 2009). UrbanSim tracks the cell locations of individual households and employments using a logit model to simulate the relocation decisions of existing households and firms, place households and jobs in grid cells, and anticipate the changes in their locations (Waddell, 2002).

UrbanSim has several sub-models, such as the economic and demographic transition models, the household and employment mobility models, the accessibility models, the household and employment location choice models, the real estate development models, and the land price models. The locations in the model are based on a grid of 150 by 150 metres as well as parcel and census block, which allows analyses of the spatial distribution of employment and jobs at the highly-disaggregated level. This disaggregated agent-based model enables researchers to simulate future urban growth patterns that suggest a big picture of the future urban spatial structure for policy makers and planners. Particularly, the residential location choice model presents preferred spatial locations of households and suggests policy implications for the future housing development.

3. Residential location choice model

For the past few decades, numerous studies have developed residential location choice models. Alonso (1964) was the first one to attempt to explain personal residential choice behaviours on the basis of the concept of utility maximization. He argued that household utility can be measured by the expenditure in goods, the distance from the CBD, and the size of the land lots. Later, Muth (1969), Mills (1972), Evans (1973), and Wheaton (1974) extended the model, which is referred to as the classical urban land market model. However, many empirical studies have proved that the assumption of monocentricity, single-worker households, and exogenous workplaces in the classical urban models is not appropriate in the real world (Waddell, 1996).

Another stream of research on modelling residential location choices is based on the discrete choice framework, which was developed by McFadden (1978). One of the most important advantages of this approach is that it enables researchers to consider the physical and social characteristics of the surrounding environments as well as housing attributes based on the random utility maximization theory. Through this, researchers can understand how household tradeoff among various choice sets plays in deciding their residential locations. Moreover, this approach provides a way of understanding how the interaction between socio-demographic attributes and spatial characteristics affects household residential choices through interaction terms (Guo & Bhat, 2007; Thill & Wheeler, 2000).

The residential location choice model within UrbanSim adopts this discrete choice framework based on the assumption that individual households choose their residential location to maximize their own utilities from a variety of alternative locations. Particularly, this model assumes that each alternative residential location i is associated with its utility (U_i) that consists of a systematic part (u_i), which is measurable, and a random part (ε_i), which is unobservable.

$$U_i = u_i + \varepsilon_i \tag{1}$$

where $u_i = \beta \cdot x_i$, β is a vector of i coefficients, x_i is a vector of independent alternative variables that may interact with the characteristics of the household in the residential location choice model, and ε_i is an unobserved random term. Assuming that the unobserved part (ε_i) is distributed with a Gumbel distribution makes it possible that the residential location choice model can be represented as the multinomial logit model (McFadden, 1978):

$$P_i = \frac{e^{u_i}}{\sum_j e^{u_i}} \tag{2}$$

where j is an index over all possible residential locations and the coefficient β is estimated by the maximum likelihood method.

Following this framework, we build our empirical residential location choice model. In this study, our dependent variable is a grid cell that is chosen by an individual household, which will be explained in detail in section 5. The residential location choice model is specified as a multinomial model with a systematic part, which describes the utility of the residential location choice. It can be established using the linear combination of variables following the form:

$$u_i = \beta_A x_A + \beta_E x_E + \beta_N x_N + \beta_H x_H + \varepsilon_i \tag{3}$$

where A indicates accessibility to employment opportunities, E indicates economic factors such as housing price to income ratio and land price, and N reflects neighbourhood effects, which consist of average building age, land use mix, and residential density, and H indicates household characteristics, which are categorized by the age of the household head, income, and the number of children.

The selected variables used in this study are drawn not only from previous research on Korean residential location choice models, but also many residential location choice studies. Particularly, previous studies have argued that accessibility (Ben-Akiva & Bowman, 1998; Sermons & Koppelman, 1998), housing price to income ratio (De Palma, Motamedi, Picard, & Waddell, 2005; Zhou & Kockelman, 2009), land price (Kim & Ji, 2007), average building age (Lochl & Axhausen, 2010), land use mix (Guo & Bhat, 2007; Waddell, 2006), residential density (Ben-Akiva & Bowman, 1998; Sermons & Koppelman, 1998), and household characteristics (Ben-Akiva & Bowman, 1998 Kim, Pagliara, & Preston, 2005;) are important determinants of residential location choices. Also, the school district is one of the most important variables in residential location decisions in Korea (Kim & Kim, 2013). However, it is considered only by households with children under the age of 20 (Park, 2014), and its effect can be capitalized into housing prices or land prices. Hence, we include a land price variable that can be representative of living environment factors. According to classic urban economic theory, residential location choices are determined by a trade-off between land cost and transportation (Waddell, 2000). Therefore, we include accessibility to employment as an additional important variable in our residential location choice model.

In this study, accessibility for a given neighbourhood is measured by the distribution of opportunities weighted by the composite utility of all modes (private cars and public transit including buses, subways, and taxis) of travel to destinations. The accessibility measure is depicted as follows:

$$A_i = \sum_{j=1}^{J} D_j e^{f(k)} \tag{4}$$

$$f(x) = ln \sum_{m \in C} e^{U_{ijm}}$$

$$U_{ijm} = \alpha + \beta TC_{ijm} + \gamma TT_{ijm}$$

where D_j is the quantity of activity in location (the number of employment) j, $f(k)$ is the function of composite utility for households with a vehicle ownership level k from location i to location j, m is trip mode (private car, subway, bus, or taxi), TC is travel cost,[1] and TT is travel time. The advantage of this gravity-based accessibility model is that it provides a simple and accurate single parameter measurement of actual commuting patterns (Cervero, Rood, & Appleyard, 1999; Waddell et al., 2003).

4. Context of the study area and data

The setting for the analysis in our study is the city of Suwon, located in the south of Seoul and one of the most populous satellite cities in the Seoul Metropolitan Area. Recently, the urban structure of Suwon has dynamically changed. Over the last 20 years, Suwon has gained substantial population growth with residential land development and housing construction. Suwon has developed from a small town to a major industrial city where huge companies such as Samsung Electronics and SK Corporation exist. Also, the urban structure has changed, particularly because of the huge amount of housing construction in the southern parts of Suwon. Thus the old CBD declined, and the new town has experienced growth.

To analyze the determinants of the residential location choice for an individual household in Suwon, our study area was divided into 5650 grid cells[2] of 150 × 150 metres. Each grid cell that can be chosen by individual households is our dependent variable. An extensive database of households, jobs, housing characteristics, buildings, transportation networks, and parcel data were constructed to develop Suwon's residential location choice model. Household data were derived from the 2005 Korean Census and classified into five types, namely, income, household size, age of the household head, the number of children, and the number of vehicles. Annual income[3] of each household was estimated using a 2% sample of the 2000 Household Income and Expenditure Survey (HIES) data with

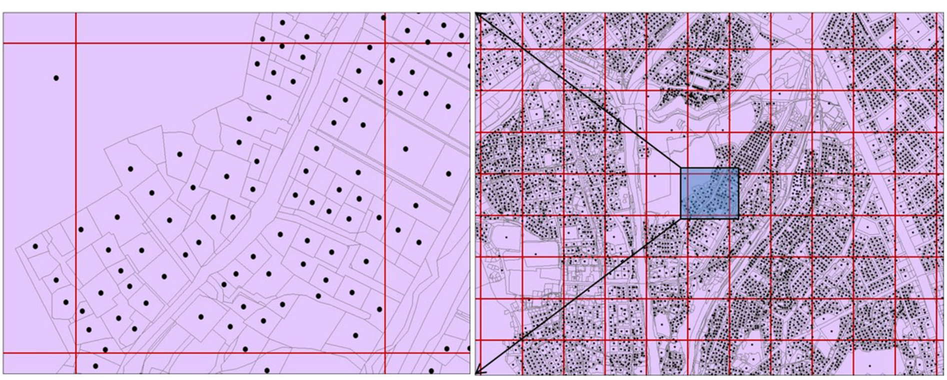

Figure 1. Sample data work.

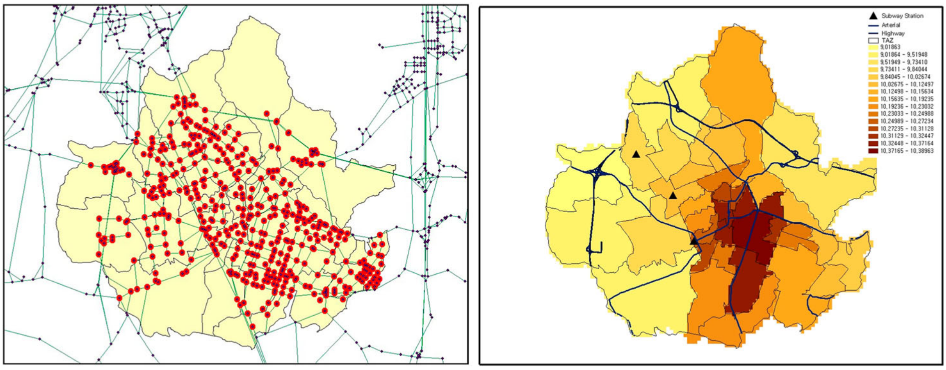

Figure 2. 2005 Korean transport database (a) / Job Accessibility by TAZ (b).

consideration of socioeconomic characteristics (e.g. housing type, housing size, housing ownership, and education status) because the Korean Census does not provide information on household income. Furthermore, we developed detailed employment data obtained from Korean Census so as to calculate households' accessibility to jobs of a particular grid cell. Employment data were established at the micro level of individual businesses with several employment sectors. Such household and employment data are based on the micro census ('jipgegu[4]'), which have been released to the public recently. As shown in Figure 1, we first geocoded households, jobs, and buildings to each parcel; these parcel data were then transformed into a grid cell, which is the spatial unit of our analysis. In addition, the year of building construction, land use classification, land development type, the building's square footage, land area, and land price were geocoded in every grid cell based on the proportion of the area. For the data, we used Architectural

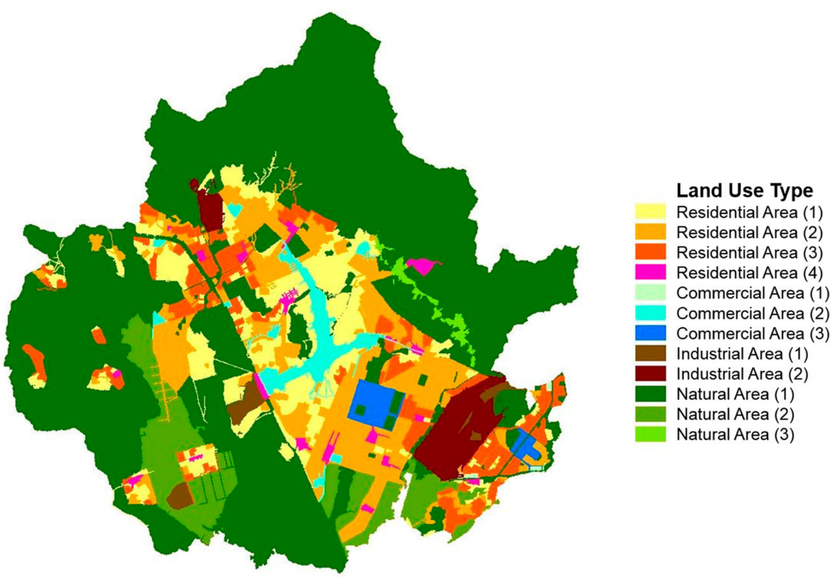

Figure 3. Land use type in Suwon.

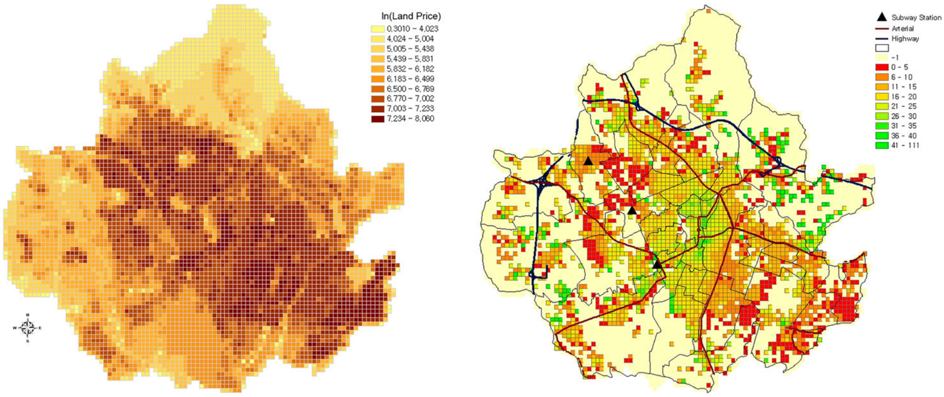

Figure 4. Land price by grid (a) / Year built by grid (b).

Information System (AIS) data, Land Management Information System (LMIS) data, and a cadastral map, obtained from various institutes including the city of Suwon.

As mentioned above, accessibility is a critical driver in determining household residential locations. Travel time and travel cost are more important factors than physical distance to measure the accessibility. Therefore, a travel demand model (TDM) was developed using EMME2 with the 2005 Korean Transport Database (KTDB) to compute composite utility, namely, the logsum accessibility indices (see Figure 2). In particular, the travel times and costs of public transport (subway, bus, and taxi) and private cars between traffic analysis zones (TAZ) were calculated using a nested mode choice model and then used for the calculation of the logsum. Like other data, these measured accessibility indices were also converted into every grid cell as shown in Figure 1. In order to calculate the mixed land use index, we used an entropy index[5] with the map of land use classification (See Figure 3). Figure 4 presents two examples (land price and building age) of constructed data by the grid cells.

5. Empirical analysis

5.1. Estimation results of the residential location choice model

Before deciding on our final residential location choice model, we tested how the preferences of residential locations are different among ages and income groups. Stratifying the sample of households by age and income is helpful to identify the heterogeneity in housing preferences among different market segments as well as to investigate whether differences in location preferences exist between different groups. All the household location choice models performed reasonably well, as shown in Tables 1–3.

The empirical results show that different age groups have different residential location preferences. As shown in Table 1, the access to employment is statistically significant and positive in Groups 1 (age 30 or less) and 3 (age over 51), but insignificant in Group 2 (age 31–50). This result implies that job opportunity is an important factor in household residential location choice decisions, especially for the young and old household groups. On the other hand, the households in Group 2 do not tend to prefer locations where job

Table 1. Estimation results categorized by the age of the household head.

Coefficient	Group1 Age 30 or Less Estimate	Group2 Age 31–50 Estimate	Group3 Age over 51 Estimate
Access to employment	1.0801**	−0.0579	0.0894*
Average building age	−0.0368**	−0.0204**	0.0288**
Land price	−0.1292**	0.0851**	−0.0134
Cost to income ratio	−0.0142**	−0.0178**	−0.0045**
High residential density	−0.5136**	−0.3479**	−0.3295**
Mixed land use	0.5649**	−0.3109**	−0.1359**
Number of observations	10,000	10,000	10,000
Prob. > Chi2 (*p*-value)	0.00	0.00	0.00
Log-likelihood	−31,824.96	−33,612.21	−33,086.92

*significant at the 0.05 level and **significant at the 0.01 level.

opportunities are relatively good. Rather, land prices are positive and significant in Group 2, which means they are likely to choose locations where land prices are high. This is probably because they value good living environments, such as those with good school districts, safe neighbourhoods, and plentiful amenities, for their family members (e.g. children or parents), despite the high land costs. The effects of price on residential location choices can be captured by the ratio of housing cost to income (Waddell, 2006). A negative value for the variable implies that households prefer spending a small fraction of their income on housing. The result shows that the cost to income ratio is negative in all groups, which is consistent with our expectation.

Mixed land use commonly refers to diverse types of land use, such as residential, industrial, and commercial, close together (Litman, 2011). The diversity in land use within a certain region might be expected to influence residential location choices and their travel behaviours because mixed land use could offer residents to live, work, shop, and enjoy recreational activities all within one place (Dieleman, Kijst, & Burghouwt, 2002). As shown in Table 1, mixed land use has a positive effect on the residential location choices only in Group 1 but negative effects on the other groups, indicating that young people prefer locations where the diversity index of land is high. This is because these locations provide more diverse opportunities for jobs, recreation, and shops to young people. High residential density has a significant and negative effect on residential location choices in all groups, and the average building age also has a significant and negative effect

Table 2. Estimation results categorized by the household income.

Coefficient	High income Estimate	Low income Estimate
Access to employment	−0.0120	0.1578**
Average building age	−0.3543**	−0.0033*
Land price	0.0709**	−0.0296
Cost to income ratio	−0.0380**	−0.0094**
High residential density	−0.2913**	−0.4572**
% high income households within walking distance	0.0386**	
% low income households within walking distance		0.0264**
Mixed land use	−0.5236**	0.0671**
Number of observations	10,000	10,000
Prob. > Chi2 (*p*-value)	0.00	0.00
Log-likelihood	−31,945.48	−35,297.41

Note: High-income households are defined to have monthly income above 5,000,000 won (approximately 4400 US Dollar), while low-income households have monthly income below 1,500,000 won (approximately 1300 US Dollar).
*significant at the 0.05 level and **significant at the 0.01 level.

Table 3. Estimation results of the residential location choice model.

	Coefficient	Estimate
ATE	Access To Employment	0.0846*
ABA	Average Building Age	−0.0102**
LP	Land Price	0.0360
CTIR	Cost To Income Ratio	−0.0060**
RUWHHC	Residential Units When Household Has Children	0.2041**
HDR	High Residential Density	−0.4779**
HIHW	% High Income Households Within Walking Distance If High Income	0.0447**
LIHW	% Low Income Households Within Walking Distance If Low Income	0.0411**
MLU	Mixed Land Use	−0.2210**
YHIMLU	Young Household In Mixed Land Use	0.6343**
	Number of observations	10,000
	Prob. > Chi2 (*p*-value)	0.00
	Log-likelihood	−32,714.872

*significant at the 0.05 level and **significant at the 0.01 level.

on household residential location decisions in Groups 1 and 2 but not in group 3, which implies that younger people prefer newer buildings. In other words, while older people tend to live continuously in their older housing units, young people prefer newly constructed housings.

As shown in Table 2, different income groups also have different preferences in choosing their residential locations. The estimation results show that access to employment has a positive effect on the residential location choices of the low-income group, but this is not statistically significant in the high-income group. This implies that access to employment opportunities can be regarded as a more significant factor for low-income households than for high-income households. While the estimation result of the land prices is positive in the high-income household group, it is negative in the low-income households. Such estimated results may indicate spatial segregation by income level; the high-income household group tends to place themselves in neighbourhoods with better living environments and amenities despite their high land prices; in contrast, the low-income group prefers residential locations in regions where the land prices are low. However, the cost to income ratio has a negative influence on the residential location choices in both income groups. The average building age has a negative effect on the residential location choices of high-income households, whereas the effect is not statistically significant in the low-income group.

Finally, we estimated the household residential location choice model using the total household sample in Suwon (see Table 3). All the estimated coefficients included in this model are statistically significant, except for land price. Particularly, with consideration of the above results differentiated by age groups, two interaction terms are used in this model, and their estimation results are statistically significant and positive: the first one is the interaction between the residential units and the households with children, and the second one is the interaction between the young households and the mixed land use. Households with children prefer locations where the number of residential units is large, indicating that parents who have children prefer more resident-friendly areas. This can be explained by the fact that such areas ensure greater safety for children. The second interaction term appears to be consistent with the above results in Table 1 (i.e. mixed land use has a positive effect on the residential location choices of the young age group). However, mixed land use has a negative effect on the residential location decisions of all the households in total. Namely, most households are not likely to choose mixed land use areas as their residential locations, whereas younger households prefer mixed land use areas for their residential locations.

5.2. Utility and probability of residential location choices

This study visualized the utility of residential locations in a 150×150 metre grid cell by using the utility function. Based on the estimation results in our final model, the utility function can be described as follows:

$$\begin{aligned} U_i = {} & 0.0846{*}\mathrm{ATE} - 0.0102{*}\mathrm{ABA} - 0.0060{*}\mathrm{CTIR} + 0.2041{*}\mathrm{RUWHHC} \\ & - 0.4779{*}\mathrm{HDR} + 0.0447{*}\mathrm{HIHW} + 0.0411{*}\mathrm{LIHW} - 0.2210{*}\mathrm{MLU} \\ & + 0.6343{*}\mathrm{YHIMLU} \end{aligned} \tag{5}$$

where U is utility function of the household residential location choices and i is the number of the grid cell. Based on the results of our final model estimation, we obtained the utility model to explain decisions on household residential location choice. The above equation illustrates the model with nine variables, not including land prices (land price was eliminated due to its insignificant results). As shown in Figure 5, the high-utility areas are distributed in the centre of Suwon or following the arterial roads.

The residential location choice model predicts the individual household location choices to a grid cell rather than to specific dwelling units. In other words, this model

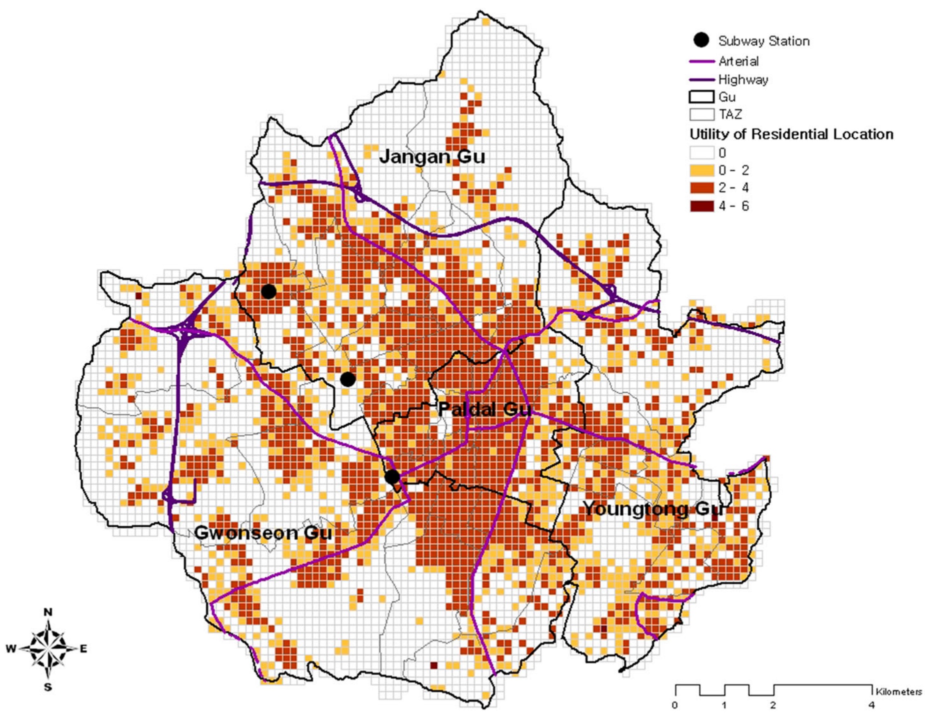

Figure 5. Utility of residential location Choices in Suwon, 2005.

predicts the probability that a household will select a location in a specific grid cell of 150×150 metres. Generally, the higher the utility of a residential location is, the higher the probability of choice by individual households. To calculate the probability, we considered a few conditions, including housing units. For example, if a grid cell has no vacant housing units, although its utility of residential location might be extremely high, its probability as a residential location choice would be very low. Thus, it is necessary to calculate how many vacant housing units are in each grid cell.

As mentioned above, the number of housing units and the number of households were geocoded based on a grid cell using the 2005 Korean census. During this process, it is assumed that the vacancy rate of housing units in Suwon is 5%, which is based on the housing statistics from the Korean Census. If the number of households is greater than the number of housing units, it is impossible to calculate the probability of a residential location choice. In this regard, grid cells that have vacant housing units were targets for calculating the probability of residential location choices, and they are calculated by using equation 2 (see Figure 6).

Comparing between the spatial patterns of utility and probability in household residential location choices, the two patterns are quite different; the centre of Suwon has high utility but a very low probability of residential location. One of the possible reasons for this is that though this area provides a good opportunity for employment and good accessibility, it does not provide enough available housings for households. Therefore, households who want to live in this location cannot find affordable housing in the centre of Suwon. In this study, we do not consider a supply-side model because we only focus on the factors affecting residential location choices and attempt to simulate a one-year change of household residential location choices. Although it is difficult to tell how many housing units could be constructed in the inner-city, the different spatial patterns of utility and probability suggest an important housing policy implication that this area can be a good place for infill development or urban regeneration. In the future, further investigations are needed to examine its economic effect, possible negative externalities, and land use regulations on the supply side.

5.3. Simulation of residential location choices

To forecast the change of individual residential locations, we used the choice process within the UrbanSim location choice model (see Figure 7) based on the utility equation suggested in section 5.2. This model calculates the utilities of each residential location and the probability of residential location choices. In this process, households can only select grid cells with vacant housing. However, the number of alternatives is the total number of available housing units, which creates a very large choice set. We set by selecting a random sample of nine alternatives from the variety of vacant housing in implementing this model. Thus, the estimation process uses a random sample of 9 non-chosen alternatives (Waddell et al., 2003). This random sampling of alternative techniques is commonly used to estimate multinomial logit models with large numbers of alternatives.

We consider two types of households for this simulation: the first type is households that immigrated from other areas into Suwon, and the second one is households that changed their residential locations within Suwon. In the case of Suwon, the total population has increased in the past few decades. The number of increased households is

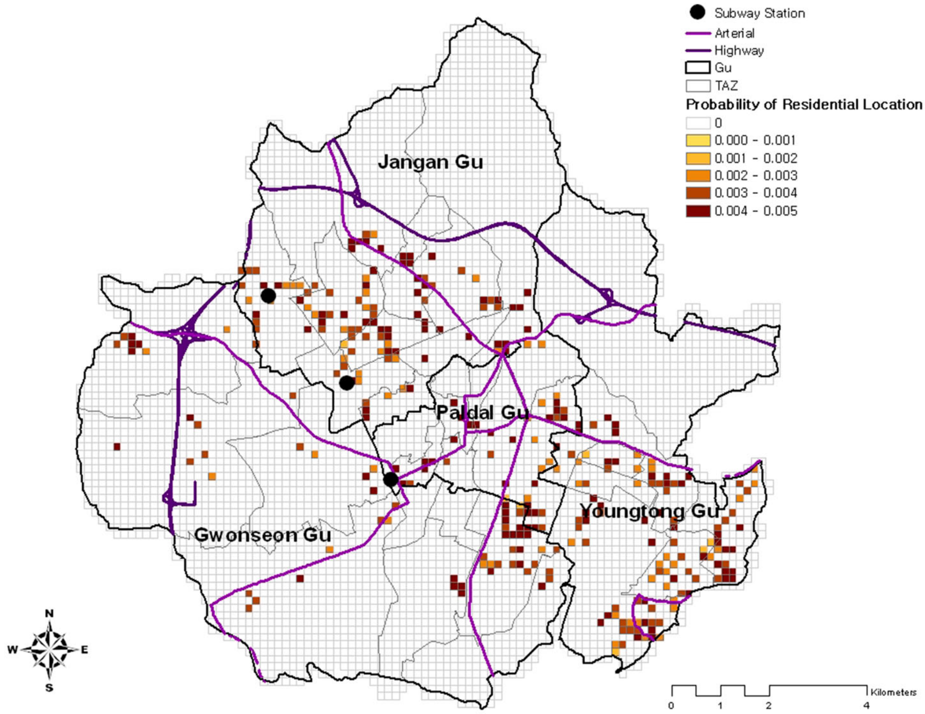

Figure 6. Probability of residential location choices in Suwon, 2005.

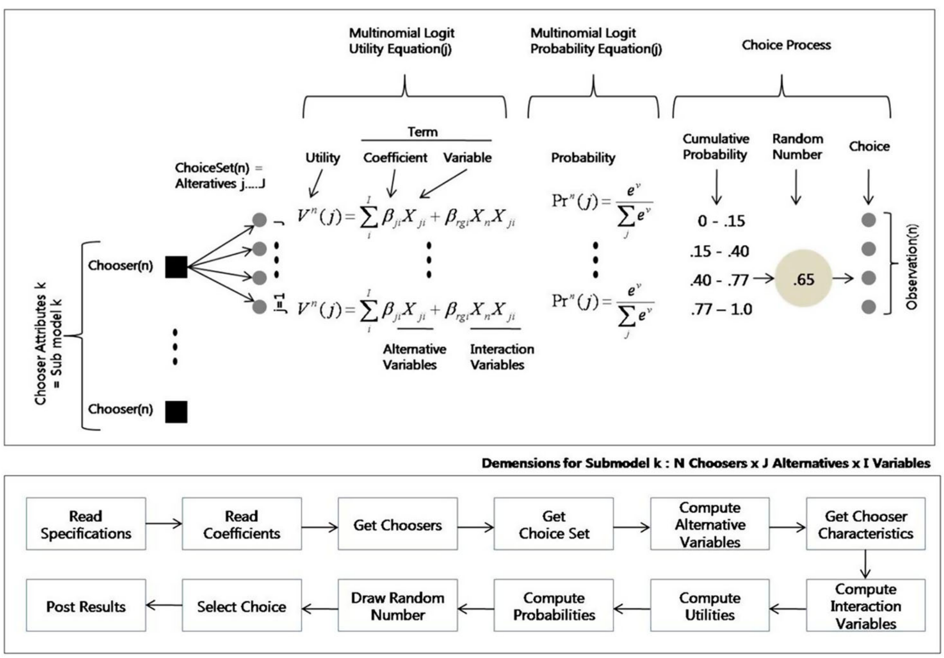

Figure 7. Choice process in UrbanSim location choice models. Source: p.14; Waddell, 2006.

exogenous, so we added the number of new migrants (about 4000 in 2005) based on the population statistics in Suwon (i.e. we added approximately 1800 households as a new entry). In particular, we used the demographic transition model within UrbanSim that is designed to add the number of increased residents, and also used the household mobility model that is designed to determine the mobility rate of households (this is determined by the vacancy rates of the housing stock) (see Waddell et al., 2003). These types of households were also classified according to age, income, household size, and the number of children. After that, we simulated the residential location choices of individual households for one year, from 2005 to 2006 and compared the results with the actual population change from the same period. The reason we simulated for such a short period of time is that we had limited information and data to forecast the change of the housing units. Nevertheless, this study can be a starting point for developing a residential location model that forecasts household residential location changes at a micro-level in Suwon's local housing market.

As shown in Figure 8, quite a few households that lived in Paldal-gu, which is located in the centre of Suwon, moved to other areas, whereas many households moved near the arterial roads in Paldal-gu. Moreover, many households moved near the subway station in Jangan-gu, which has many newly constructed apartments (see Figure 4(b)). This pattern describes how many households tend to consider good accessibility when they choose their residential locations. In addition, many households moved to Youngtong-gu, which is a new neighbourhood in Suwon because it provides a relatively good living environment with lots of new housing.

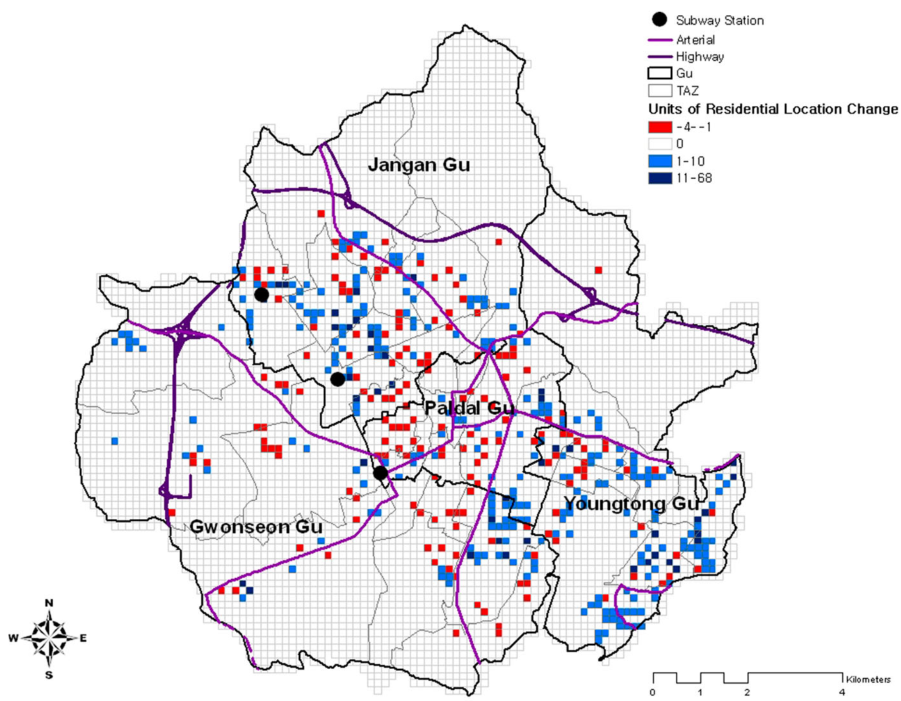

Figure 8. Simulation of residential location choices in Suwon, 2005–2006.

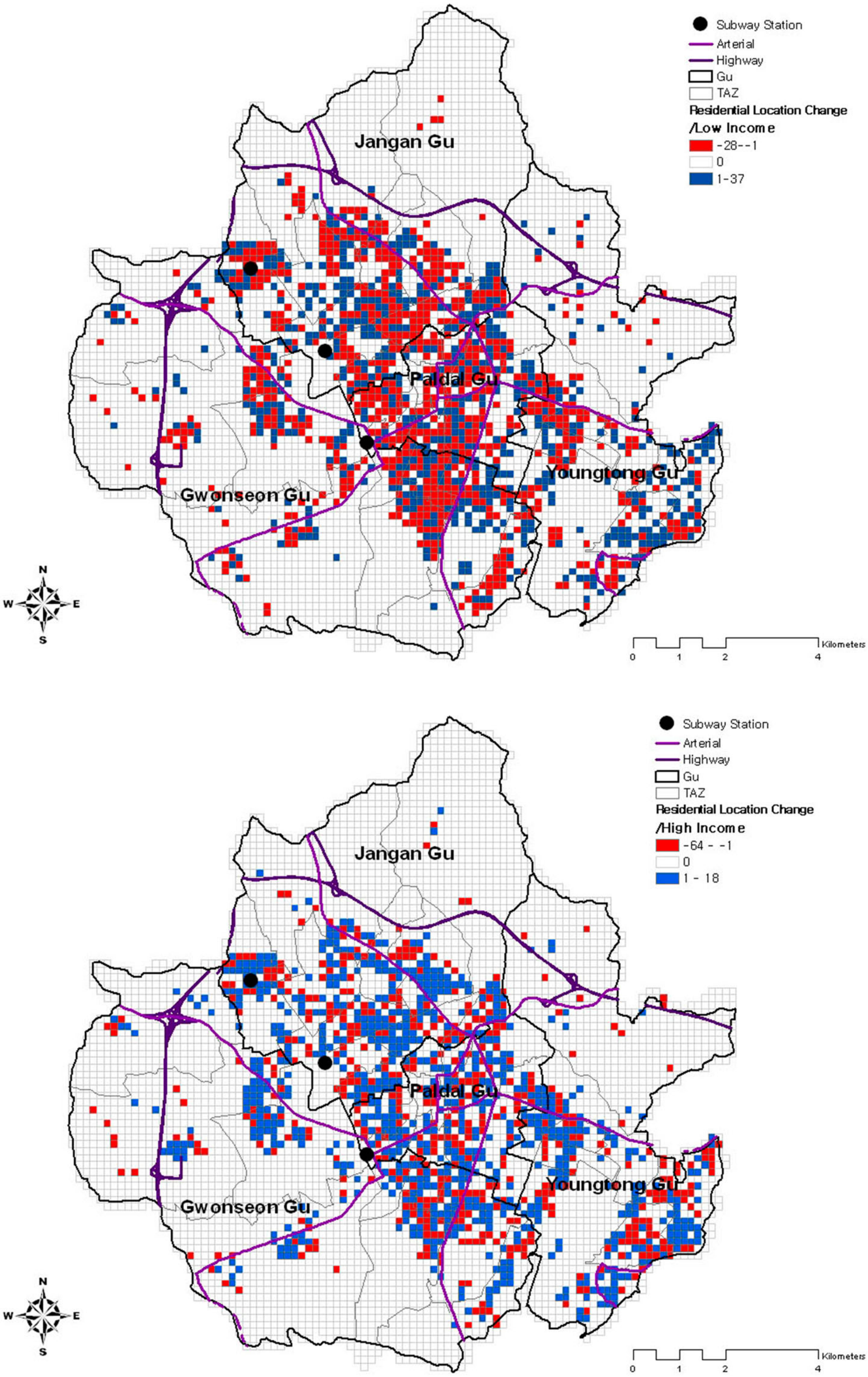

Figure 9. Simulation of residential location choices by income level in Suwon, 2005–2006.

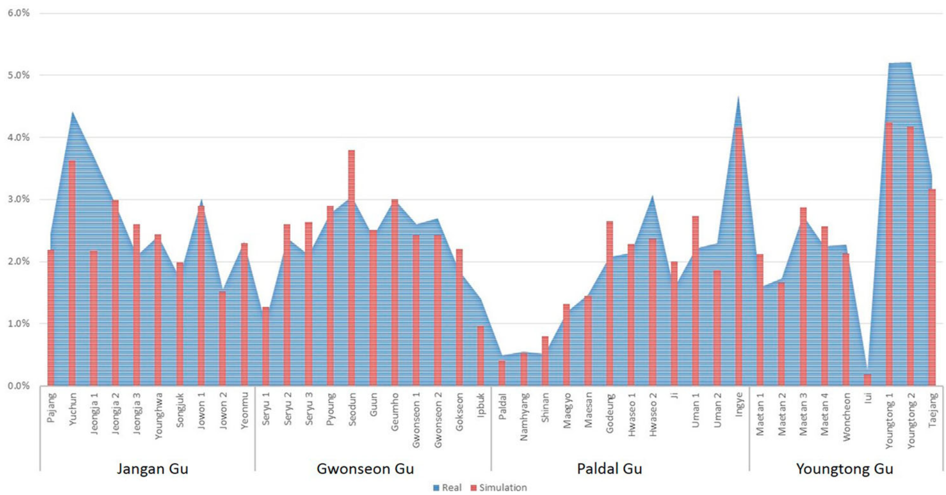

Figure 10. Comparison of simulation results and real household changes in Suwon, 2005–2006.

When comparing the migration patterns between high-income households and low-income households, both income groups tend to move to locations that have a high probability of residential location choices. However, the two income groups show different migration patterns: the high-income households are distributed across a wide range of the Suwon area, whereas the low-income households are located in a relatively small portion of the Suwon area (see Figure 9). In other words, the grid cells with increased low-income households occupy a relatively small portion of the entire region as compared to the grid cells with increased high-income households. As shown in our estimation results in Table 2, this may be explained by the fact that they have different preferences in residential location decisions. High-income households do not consider accessibility to employment, but prefer higher land prices (i.e. better living environments) when making their residential decisions, whereas low-income households tend to choose areas with high accessibility to employment and mixed land use. These estimation results are consistent with previous studies (Park, 2014). In addition, the simulation results show that high-income households have more options to from which to choose their residential locations, and low-income households are relatively much less free to choose their residential locations, which suggests that strategic housing development should be necessary for low-income households.

To verify the model, we compared the results of the simulation with the actual population changes between 2005 and 2006. Due to a lack of information on the 2006 population at the micro level (there is no available data at the micro level (jipgegu) in 2006), all the modelled outputs are aggregated to an administrative district (the Dong-level). As shown in Figure 10, the simulation results are very similar to the actual population changes, even though some changes in the estimated population in some areas are slightly different from the real changes. One possible explanation for the slight differences is the variety of housing development in Suwon. There has actually been plenty of housing and new town development in Suwon. In particular, housing development or apartment development was conducted in many areas of Youngtong-gu and

Jangan-gu. Therefore, more people were able to move to these regions as compared to the results in our simulation. But as mentioned above, our simulation could not account for the supply-side, such as the new housing units, due to a lack of this information. For this reason, the simulation results are a little different from the actual household migration between 2005 and 2006. However, these results still provide us with a reasonable level of confidence in the model's forecasts. Further research needs to develop the residential location choice model not only considering supply-side models, but also reflecting housing development policies.

6. Conclusions and discussion

In this study, we examine the determinants of household residential location decisions in Suwon using the residential location choice model within UrbanSim and then analyze their simulated residential location choices. Focussing on the city of Suwon, we estimate the residential location choice model that includes various attributes, such as household characteristics, neighbourhood characteristics, and accessibility at the micro level. Our findings reveal that households have different preferences with regard to some residential attributes, especially by age and income level.

Our major findings are as follows. First, we found that access to employment opportunities, housing cost to income ratio, mixed land use, and year when housing was built are all important factors in determining household residential locations in Suwon. However, different groups by income and age have different preferences on these factors in their residential location decisions. Second, residential segregation by income exists. High-income households are likely to choose high quality of neighbourhood environments, whereas low income households prefer to choose locations based on good accessibility to job opportunities and mixed land use. Finally, the probability of residential location choices does not correspond to the utility of residential locations. This may be because the old CBD located in the centre of Suwon provides good accessibility but does not provide enough affordable housings. Therefore, households could not move to the old CBD in the simulation, despite its high utility.

These results suggest some important policy implications for housing development in Suwon. One implication is that policy makers should consider the preferences of different household groups, especially by age and income level when they establish housing development plans because their patterns of residential location choice are not the same. For example, providing affordable housing in locations with higher mixed land use would be helpful for young and low-income households. Also, as shown in Figure 9, low-income households have more limited options when moving than high-income households, in terms of their residential locations. Strategic housing development policies that improve the social mix as well as reduce residential segregation are necessary for sustainable community development. Another implication is that policies for affordable housing should reflect the actual housing demands of individual households. In the case of Suwon, the infill development strategy or urban regeneration plans can be more efficient because a high utility of household residential locations still exist in the inner city. In other words, providing affordable housing in inner-city neighbourhoods could invigorate declining downtowns and help low-income households live in the inner-city neighboruhoods that provide better job accessibility and mixed land use. However, more investigation with consideration of the preferences of the residential locations of different household groups should be conducted.

Our research contributes to a better understanding of household residential location choice behaviours, especially in Suwon, Korea, at the micro-level by applying the residential location choice model within UrbanSim. Nevertheless, our research has limitations. First, these results should not be treated as definitive because many factors, such as economic conditions and social and physical environments, influence changes in human behaviour over time, especially regarding residential location decisions. Second, UrbanSim consists of not only the residential location choice model but also several other models, such as the employment location choice model, the real estate development model, and the land price model. However, the simulation model used in this study is not fully developed due to a lack of data. Additional longitudinal datasets are necessary for future modelling work, and a longer-term forecast is much more helpful for planners and policy makers to make housing development plans for many cities in Korea. In addition, collecting more data to add other factors that affect household residential location choices would improve the predictability of the model. Applying parcel-version applications with spatially disaggregated data would be also helpful to improve the accuracy of forecasting (see Lee & Waddell, 2010). Finally, we suggest that future work should simulate urban growth patterns using various housing development scenarios to suggest better policy implications.

Notes

1. Travel cost by each mode was calculated with consideration of fuel cost (1600 won), base fares of subways and buses plus extra charges per extra distance, and base fares of taxi (1900 won) plus extra charges per extra distance (100 won per 144 m).
2. The parcel-version is much more detailed and sophisticated to forecast urban land use changes than the zone- and grid-version. However, allocation of the population and jobs is not easy because our 'jipgegu' data are spatially larger than parcel data. Hence, we apply a grid-version of UrbanSim.
3. HIES and the Korean Census have the same categories of socioeconomic characteristics: housing type, household size, housing ownership, and education status. Therefore, we used these four categories to estimate household income, and then the estimated household income was inflation-adjusted to reflect calendar year 2005. More detailed information is available upon request.
4. Jipgegu is a level of micro census in Korea, and its median area is approximately 0.02 km^2. Each jipgegu has the numbers of households, employments, housing buildings, and firms (https://sgis.kostat.go.kr).
5. Entropy $= -\frac{\sum_j^J L_j ln(L_j)}{ln(N)}$, where N is the number of land use types under consideration and L_j is the fraction of the neighborhood that is of land use type j.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Alonso, W. (1964). *Location and land use*. Cambridge: Harvard University Press.

Ben-Akiva, M., & Bowman, J. (1998). Integration of an activity-based model system and a residential location model. *Urban Studies*, *35*(7), 1131–1153.

Cervero, R. (2003). Growing smart by linking transportation and land use: Perspectives from California. *Built Environment*, *29*(1), 66–78.

Cervero, R., Rood, T., & Appleyard, B. (1999). Tracking accessibility: Employment and housing opportunities in the San Francisco Bay Area. *Environment and Planning A*, *31*(7), 1259–1278.

Clarke, K. C., Hoppen, S., & Gaydos, L. (1997). A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area. *Environment and Planning B: Planning and Design*, *24*, 247–261.

De la Barra, T., Perez, B., & Vera, N. (1984). TRANUS-J: Putting large models into small computers. *Environment and Planning*, *11B*, 87–101.

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

Dieleman, F., Kijst, M., & Burghouwt, G. (2002). Urban form and travel behavior. Micro-level household attributes and residential context. *Urban Studies*, *39*(3), 507–527.

Evans, A. W. (1973). *The economics of residential location*. London: Macmillan Press.

Giuliano, G. (1989). New directions for understanding transportation and land use. *Environment and Planning A*, *21*(2), 145–159.

Guo, J., & Bhat, C. (2007). Operationalizing the concept of neighborhood: Application to residential location choice analysis. *Journal of Transport Geography*, *15*(1), 31–45.

Handy, S. (2005). Smart growth and the transportation-land use connection: What does the result tell us? *International Regional Science Review*, *28*(2), 146–167.

Hanson, S. (1995). *The geography of urban transportation* (2nd ed.). New York: Guilford Press.

Hunt, J., & Abraham, J. (2005). Design and implementation of PECAS: A generalized system for allocating economic production, exchange and consumption quantities. In M. Lee-Gosselin, & S. Doherty (Eds.), *Integrated land-use and transportation models: Behavioral foundations* (pp. 253–273). Amsterdam: Elsevier.

Hunt, J., Kriger, D., & Miller, E. (2005). Current operational urban land-use transport modeling frameworks: A review. *Transport Reviews*, *25*(3), 329–376.

Iacono, M., Levinson, D., & El-Geneidy, A. (2008). Models of transportation and land use change: A guide to the territory. *Journal of Planning Literature*, *22*(4), 323–340.

Joo, Y., Hassan, M., & Jun, C. (2011). An application of the UrbanSim land price model in Youngsan-gu, Seoul, Korea. *International Journal of Urban Sciences*, *15*(1), 15–24.

Kakaraparthi, S., & Kockelman, K. (2011). Application of UrbanSim to the Austin, Texas, region: Integrated-model forecasts for the year 2030. *Journal of Urban Planning and Development*, *137*(3), 238–247.

Kim, J., & Ji, K. (2007). The characteristics of low-income families' residential mobility: The case of public rental housing tenants in the Seoul Metropolitan Area. *Journal of Korean Planners Association*, *42*(1), 81–97.

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

Kim, S.-Y., & Kim, Y.-H. (2013). A study on the house price determinants model considering spatial attributes of residential environment: The case of apartment in Seoul. *Journal of the Korean Cartographic Association*, *3*(3), 41–56.

Kockelman, K., Jin, L., Zhao, Y., & Ruiz-Juri, N. (2005). Tracking land use, transport, and industrial production using random-utility-based multizonal input–output models: Applications for texas trade. *Journal of Transport Geography*, *13*(3), 275–286.

Kryvobokov, M., Kercier, A., Bonnfous, A., & Bouf, D. (2013). Simulating housing prices with UrbanSim: Predictive capacity and sensitivity analysis. *Letters in Spatial and Resource Sciences*, *6*, 31–44.

Lee, B., & Waddell, P. (2010). Residential mobility and location choice: A nested logit model with sampling of alternatives. *Transportation*, *37*, 587–601.

Lemp, J., Zhou, B., Kockelman, K., & Parmenter, B. (2008). Visioning versus modeling: Analyzing the land use-transportation futures of urban regions. *Journal of Urban Planning and Development*, *134*(3), 97–109.

Litman, T. (2011). *Measuring transportation: Traffic, mobility and accessibility*. Victoria Transport Policy Institute.

Lochl, M., & Axhausen, K. (2010). Modeling hedonic residential rents for land use and transport simulation while considering spatial effects. *Journal of Transport and Land Use*, *3*(2), 39–63.

McFadden, D. (1978). Modeling the choice of residential location. In A. Krlqvist, L. Lundqvist, F. Snickars, & J.W. Weibull Edition (Eds.), *Spatial interaction theory and planning models* (pp. 75–96). Amsterdam: North Holland.

Miller, E. J., Kriger, D. S., Hunt, J. D., & Badoe, D. A. (1998). *Integrated urban models for simulation of transit and land-use policies*. Retrieved July 23, 2010 from Transportation Research Board website http://www.trb.org/Main/Blurbs/153800.aspx

Mills, E. S. (1972). *Urban economics*. Glenview: Scott Foresman.

Muth, R. F. (1969). *Cities and housing*. Chicago: University of Chicago Press.

Park, W. (2014). Analysis of residential location preference factors by characteristics of households in the case of Daegu-Gyeongbuk households. *Journal of the Economic Geographical Society of Korea*, *17*(4), 702–717.

Patterson, Z., Kryvobokov, M., Marchal, F., & Bierlaire, M. (2010). Disaggregate models with aggregate data: Two UrbanSim applications. *The Journal of Transport and Land Use*, *3*(2), 5–37.

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy*, *82*(1), 34–55.

Sermons, W., & Koppelman, F. (1998). Factor analytic approach to incorporating systematic taste variation into models of residential location choice. *Transportation Research Record*, *1617*, 194–202.

Thill, J.-C., & Wheeler, A. (2000). Tree induction of spatial choice behavior. *Transportation Research Record*, *1719*, 250–258.

Voigt, B., Troy, A., Miles, B., & Reiss, A. (2009). Testing an integrated land use and transportation modeling framework for a small metropolitan area. *Transportation Research Record*, *2133*, 83–91.

Waddell, P. (1996). *Accessibility and residential location: The interaction of workplace, residential mobility, tenure, and location choices*. Paper presented at the Lincoln Land Institute TRED Conference.

Waddell, P. (2000). A behavioral simulation model for metropolitan policy analysis and planning: Residential location and housing market components of UrbanSim. *Environment and Planning B: Planning and Design*, *27*(2), 247–263.

Waddell, P. (2002). Urbansim: Modelling urban development for land use, transportation and environmental planning. *Journal of the American Planning Association*, *68*(3), 297–314.

Waddell, P. (2006). *Reconciling household residential location choices and neighborhood dynamics* (Unpublished manuscript). Retrieved January 6, 2008, from www.urbansim.org/pub/Research/ResearchPapers/smr-urbansim.pdf

Waddell, P., Borning, A., Noth, M., Freier, N., Becke, M., & Ulfarsson, G. (2003). Microsimulation of urban development and location choices: Design and implementation of UrbanSim. *Networks and Spatial Economics*, *3*(1), 43–67.

Waddell, P., Wang, L., Charlton, B., & Olsen, A. (2010). Microsimulation parcel-level land use activity-based travel. *The Journal of Transport and Land Use*, *3*(2), 65–84.

Wegener, M. (1995, February). *Current and future land use models*. Paper presented at the Land Use Model Conference at the Texas Transportation Institute, Dallas.

Wheaton, W. (1974). Linear programming and locational equilibrium: The Herbert-Stevens model revisited. *Journal of Urban Economics*, *1*(3), 278–287.

Zhou, B., & Kockelman, K. (2009). Lessons learned in developing and applying land use model systems. *Transportation Research Record*, *2133*, 75–82.