Rajabi *et al. Computational Urban Science* (2024) 4:2
https://doi.org/10.1007/s43762-024-00115-3

Computational Urban Science

ORIGINAL PAPER

Open Access

Residential location choice: an investigation of transportation, public facilities, and social factors

Hamed Rajabi[1], Hamid Mirzahossein[1*], Seyed Mohsen Hosseinian[1] and Xia Jin[2]

Abstract

Residential location choice is a crucial topic in transportation planning research since land use as well as residential land use can significantly affect a city's attractiveness for development and residence. Understanding the factors that influence households in their residential location choice is essential for policymakers to evaluate the effect of their decisions. In this study, the impact of transportation factors on the attractiveness of residential areas was investigated in Qazvin city, Iran, using the stated preference (SP) method and structural equation modeling (SEM). The results indicated that the type of housing and private house preference were significant factors influencing the residential location choice. Additionally, proximity to health centers, low pollution levels, and access to public transportation and taxi stations were the top priorities for residents when choosing a place to live. Notably, households with children in education had a greater emphasis on air pollution and the proximity to taxi stations, as these factors could affect their children's health and education. Overall, the findings suggested that transportation factors played a critical role in the residential location choice and that policymakers should prioritize public transportation and taxi services, as well as reduce pollution levels, to make residential areas more attractive and livable for Qazvin residents.

Keywords Residential location choice, Stated preferences, Structural equation modeling, Accessibility

1 Introduction

With the increase in urbanization, planners and policymakers face significant social, economic, and environmental challenges caused by the complex and dynamic interactions between the environment and humans (Ballesteros et al., 2023; Khodayari et al., 2019; Ourang, 2022). Choosing a place of residence is a part of the interaction in transportation planning and land use. Therefore, it is essential to provide better insight to policymakers and planners to help them make better decisions in line with sustainable urban development and related facilities (Dehghani et al., 2022, 2023; Le & Le, 2022).

Decisions related to residential location choice have been widely investigated and modeled using choice theory and based on the concept of utility maximization. Residential location choice models are essential for analyzing urban economic and housing policies, transportation policies, and urban social structure (Li et al., 2020; Schirmer et al., 2014). Understanding the households' choice in residential environments in travel demand models can indicate their preferred travel behaviors and access to jobs and household needs (ASLAM et al., 2019). Considering the importance of investigating the effective factors on the residential location choice for the residents

*Correspondence:
Hamid Mirzahossein
mirzahossein@eng.ikiu.ac.ir
[1] Department of Civil-Transportation Planning, Imam Khomeini International University, Qazvin, Iran
[2] Department of Civil and Environmental Engineering, Florida International University, Miami, Fl, USA

© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

of a city and the definition of different urban land uses that ultimately make an area in the city attractive or unattractive, investigating the effective factors for households on the residential location choice and the importance of each one, and finally, estimating the attractiveness of each area according to the existing features, can be very effective in evaluating the effects of policymaking.

Due to the importance of residential location choice and land use and its relationship with transportation planning, various studies have been conducted regarding the factors affecting people's residential location choice. (Hunt, 2001) conducted a study to analyze the sensitivity of various factors in urban transportation by interviewing households in Canada using the stated preference (SP) method. These factors were housing type, street type, walking condition, traffic noise, air quality, municipal taxes, and trip cost and time. The logit choice model examined the effect of each factor. It was indicated that housing type was the most important factor, followed by traffic noise, air quality, and municipal taxes. The results also showed that people prefer local streets, especially with the speed bumps in front of their residential units, compared to collector roads. (Khattak & Rodriguez, 2005) examined travel behaviors in neo-traditional neighborhoods in the United States of America (USA). They applied regression models for a household behavioral survey and showed that in comparison with the conventional neighborhood, single-family households in the neo-traditional neighborhood made fewer automobile and external travel. They also revealed that the proximity of stores to sidewalks positively affected the decision to stay in a neighborhood, and the presence of stores and similar facilities that are easily accessible on foot positively affected the neighborhood's attractiveness. (Brown & Robinson, 2006) presented agent-based models to indicate the process of residential developments in urban areas. The sensitivity analysis of environmental variation patterns showed that adding heterogeneity in an agent considerably influenced the results. Moreover, the positive role of public transportation and people's attitude toward public transit in the residential location choice was illustrated. (Walker & Li, 2007) applied choice models to indicate lifestyle preferences and decisions for household locations. Intriguing policy implications were observed in three lifestyles: transit-riders, urban dwellers, and suburban dwellers. Also, the effect of the difference in lifestyle and attitude toward public transportation for trips with the purpose of education and work in the residential location was examined. Additionally, (Hunt, 2010) used the SP survey in Canada to investigate the effect of various urban form and transportation factors, such as taxes, development density, treatment of neighborhood streets, traffic noises, air quality, and mobility. Results indicated the positive effect of a private house as a place of residence and the type of streets leading to the building of the place of residence on the attractiveness of a house for the residential location choice.

(Hoshino, 2013) estimated the preference heterogeneities in stated choice data using semiparametric varying-coefficient techniques and conducted an economic valuation of landscapes with dichotomous choice contingent valuations. The results indicated access to non-motorized transportation as a determining factor in the residential location choice. (Fu et al., 2014) carried out a study to understand multiple dimensions of residential choices. They considered various residential choice dimensions and found that social interactions significantly affect choosing a place of residence. Social interactions between current neighbors also affect their next neighborhood choice, especially in households with higher education and income levels. Older people pay more attention to social interactions when they decide to choose the next neighborhood for their residence. (Lasley, 2017) explored the effect of transportation and variables affecting the residential location choice in urban areas of the USA. The results showed that house attributes mainly influenced decisions, and in most cases, the price was the most important factor, followed by neighborhood quality, including aesthetics, reputation, and amenities. Transportation and traffic concerns also ranked near the middle. Moreover, neighborhoods that were more accessible by any mode of transportation were more desirable to buyers. (Cockx & Canters, 2020) analyzed household location preferences in Belgium using discrete choice models. It was illustrated that household type, nationality, education level, and tenure status discriminated heterogeneous residential location preferences. Also, along with the socio-economic characteristics, the characteristics of the residential unit, transportation, and access had a heterogeneous effect on the residential location choices in this country. (Masoumi, 2021) investigated the residential location choice and the impact of urban travels on house location decision-making in Cairo, Tehran, and Istanbul using binary probit regression. The results revealed that the accessibility to facilities, number of accessed facilities, commuting distance, number of driving licenses in a household, age, neighborhood attractiveness perceptions, frequency of public transits, and entertainment-shopping mode choice in a faraway place influenced the residential self-selection.

As represented in previous studies, researchers mostly have focused on a single or a few aspects of transportation factors and did not consider the comprehensive and multidimensional nature of transportation factors that affect residential location choice. Also, they often have used conventional discrete choice models that assume

homogeneous preferences among individuals and ignore the latent factors that may influence the choice behavior, such as attitudes, lifestyles, or perceptions. Some studies also investigated various factors in a specific area as a case study. Therefore, limited research has been conducted to examine the choice behavior of citizens for a residential location choice considering accessibility and public facilities, as well as transportation, residential, social, and environmental factors using the stated preference (SP) method and structural equation modeling (SEM) in Qazvin city, Iran. This study outlined the changes in the behavior of households in understanding the factors affecting their residential location choice. In this regard, using a questionnaire based on the SP method, the relationship between factors related to the residential location choice was investigated, and the importance of each factor was evaluated. This study also applied SEM, which is a powerful technique that can simultaneously estimate the choice model and the measurement model, and account for the heterogeneity and the latent factors in the choice behavior. In general, the goals of this research can be summarized as follows:

- Investigating the effective factors on the residential location choice among households of Qazvin city, Iran,
- Investigating the interaction between different factors in the residential location choice,
- Analysis of the impact of each factor on the residential location choice among households using SEM.

In this research, a questionnaire was prepared to investigate the residential location choice behavior of citizens in all areas of Qazvin city. The necessary information was collected by conducting field operations, including personal information, such as the amount of monthly income and the level of education, and household information, such as vehicle ownership in the household and the number of children, and other information related to households, as well as questions to understand the impact of various transportation factors on the residential location choice. Using the obtained data and SEM, the impact of various factors on the attractiveness or unattractiveness of choosing a residence place was estimated. Also, the accuracy of these estimates was examined, and finally, the importance of each factor was represented.

2 Methodology

In this section, the examined sample and the questionnaire prepared to investigate the factors affecting the residential location choice are first explained. Then, SEM was used to perform modeling considering the effect of various factors on the residential location choice, the existence of various variables and the nature of existing variables, such as accessibility, social factors, and society's insight regarding transportation factors, which were latent variables. Figure 1 shows the general process of this study.

2.1 Questionnaire design

The population studied in this research was the residents of Qazvin city, Iran, and based on the stated preference (SP) approach and through the preparation of questionnaires, data were collected in order to determine the factors effective in the residential location choice of residents. The questionnaire was designed in such a way that latent variables influential in choosing a residential location were collected through questions based on revealed variables.

The questionnaire was developed in two different parts. The first part of the questionnaire contains personal and demographic questions of the respondents, including 16 questions such as gender, marital status, job, education level, number of children and family members under 18 years, number of children in education, house type and ownership, and monthly income of the family. Table 1 summarizes the descriptive statistics of the respondents.

The second part of the questions was related to the factors affecting the choice of residential location and determining the importance of each factor on the questioner's location choice. In order to use evident variables to determine latent variables, the questionnaire questions were applied in the categories according to Table 2. The latent variables used in the modeling are as follows:

1. The importance of residential location accessibility based on household activities, including workplaces, schools, and shopping centers,
2. The importance of public facilities near the residential location,
3. The importance of transportation facilities near the residential location,
4. The importance of house location,
5. The importance of house type based on the structural nature of the house (private house, apartment, residential complexes, and tower),
6. The importance of traffic and pollution,
7. The importance of social factors.

The information obtained from the questionnaire was entered into AMOS software. Table 3 shows the items of the questionnaire and the names of the variables used in the model.

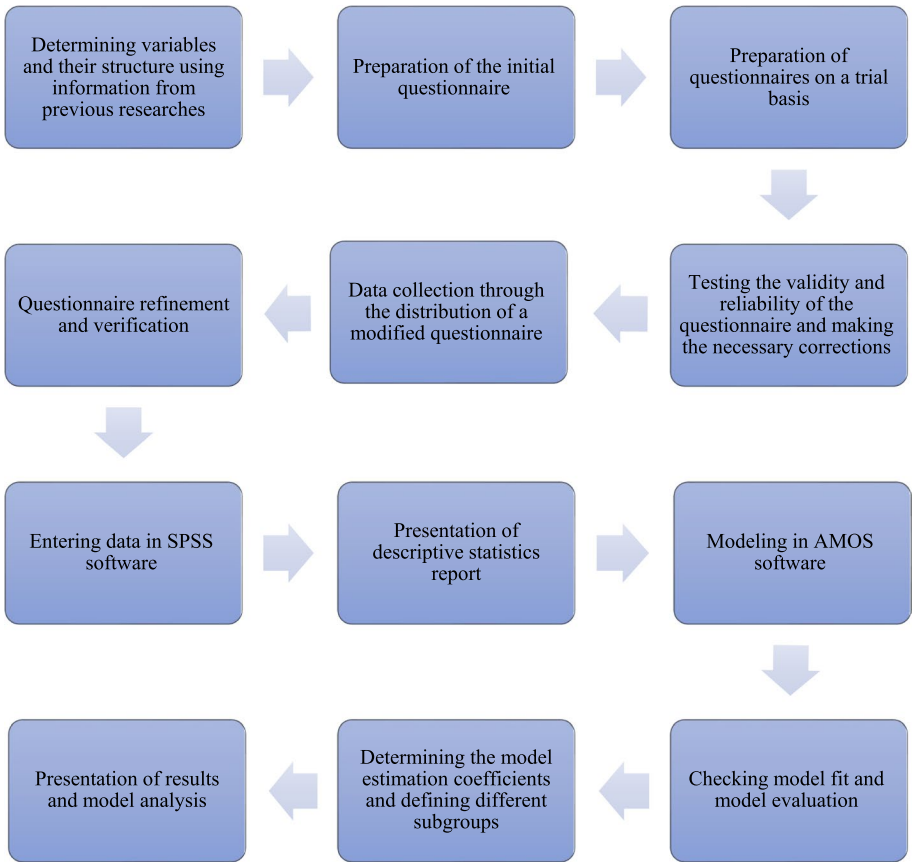


Fig. 1 The general process of the study

2.2 Sample size

Sampling is the selection of a small part of the target population as a representative of the target population. The completion of the questionnaire and the collection of information are conducted from this sample in order to carry out the research. One of the most widely used methods in choosing the sample size from the target population is Cochran's formula as Eq. 1, in which n is the minimum number of required samples, N is the size of the study population, a is the confidence interval, and p is the estimated proportion of the population (Yarahmadi, 2020).

$$n = \frac{1.96^2 p(1-p)}{\left(\frac{a}{100}\right)^2 + \frac{1.96^2 p(1-p)}{N}} \quad (1)$$

In this study, considering the confidence interval of 95% (error value of 0.05) and according to the statistical population, which are the residents of Qazvin city with 597,000 people based on the statistical information of Iran Statistics Center, the required sample size was 384.

In the process of statistics and data collection, a total of 450 questionnaires were collected, considering that some of the questionnaires had incomplete or no answers. Finally, the number of complete questionnaires was 407.

2.3 Data collection

Qazvin city, with an area of 64.13 square kilometers and a population of 597,000 people, was the study area of this research. The transportation facilities available in this city include buses and taxis. In recent years, the development of bicycle lanes has made it possible for citizens to use this type of facility. In this research, data collection was carried out through face-to-face interviews and an online questionnaire in the 12 traffic zones of Qazvin city based on the division of Qazvin Municipal Transportation and Traffic Deputy, as illustrated in Fig. 2.

2.4 Structural equation modeling

Structural equation modeling (SEM) is a perspective in which hypothetical patterns of direct and indirect relationships among a set of revealed and latent variables are examined, and its main application is in multivariable issues that cannot be examined in a two-variable way by considering an independent variable with a dependent variable each time (Cheng et al., 2019; Hair

Table 1 Personal and demographic characteristics of the respondents

Characteristic	Number	Frequency (%)	Characteristic	Number	Frequency (%)
Gender			**Education level**		
Male	223	55	High school	19	4.5
Female	184	45	Diploma	48	12
Marital status			Post-diploma	37	9
Single	143	35	Bachelor's degrees	150	37
Married	264	65	Master's degree	138	34
Number of children			PhD	15	4.5
0	202	50	**Job**		
1	78	19	Employee	169	41
2	85	21	Manual worker	36	9
3	27	6	Self-employment	93	23
More than 3	15	4	Retired	12	3
Number of family members under 18 years old			Housekeeper	41	10
0	284	70	University student	33	8
1	80	20	Unemployed	23	6
2	37	9	**Monthly family income ($)**		
3	6	1	Less than 200	53	13
More than 3	0	0	200–400	161	40
Number of children in education			400–600	93	23
0	284	70	600–800	55	13
1	80	20	More than 800	45	11
2	37	9	**Ownership type of current residence**		
3	6	1	Owner	265	65
More than 3	0	0	Tenant	142	35
Type of house of current residence					
Private house	101	25			
Apartment	306	75			

et al., 2019). Also, SEM is very useful for displaying multiple relationships between a set of variables, in which a variable that is the result (dependent variable) in one set of relationships may be a predictor of the results (explanatory variable) in other relationships (Haoran et al., 2019). This model includes exogenous variables and endogenous variables. Exogenous variables are variables that affect at least one other structure but are not affected by any structure. However, endogenous variables are variables that are affected by at least one other structure and can affect other structures (Wang et al., 2020). SEM can be defined as Eq. 2.

$$Y = BY + \Gamma X + \zeta \tag{2}$$

in which, Y is the column vector of endogenous variables, X is the column vector of exogenous variables, B is the matrix of coefficients representing the direct effects of endogenous variables on other endogenous variables, Γ is the matrix of coefficients representing the direct effects of exogenous variables on other endogenous variables, and ζ is error vector column (Nesamani et al., 2017).

According to Fig. 3, the factors affecting the residential location choice were employed for modeling by the SEM method. In this regard, three fit indices of the model were presented. One of the indices that minimize the effect of the sample size on the chi-square index is the index of the ratio of chi-square to the degrees of freedom (CMIN/DF). If this value is less than 3, it indicates a good fit (García-Santillán et al., 2012). On the other hand, the goodness of fit (GFI) index calculates the reproduced variance ratio by estimating the observed covariance value. The GFI index indicates the model's accuracy in reproducing the observed covariance matrix. This index changes between 0 and 1, where 0 indicates that the estimated model has no fit with the data, and 1 indicates the perfect fit of the model with the available data (Izadi et al., 2020; Mak et al., 2005). Moreover, the root mean square error of approximation (RMSEA) index measures the difference

Table 2 Latent and evident variables used in the model

Latent variable	Evident variable	Variable
Accessibility	Access to the workplace	wrk
	Access to schools	sch
	Access to shopping centers	shp
Public facilities	Proximity to health centers	hth
	Proximity to gardens and parks	prk
	Proximity to amenities	ame
	Proximity to government offices and service providers	ofc
Transportation facilities	Proximity to the bus station	bus
	Proximity to the taxi station	txi
	Proximity to bicycle lanes	bcy
House location	Proximity to the central business district (CBD)	cbd
	Proximity to the main street	str
House type	Private house	hos
	Apartment	apt
	Residential complexes	cpx
	Tower	twr
Traffic and pollution	Main street traffic	trf
	Air pollution	pol
Social factors	Proximity to relatives and acquaintances	rel
	Reputation of the neighborhood	rep
	Security	sec

Table 3 Questionnaire items and variables used in the model

No	Questionnaire items	Latent variable
1	How much does the proximity to the workplace affect your residential location choice?	Accessibility
2	How much influence does the presence of schools and colleges near your place of residence have on your residential location choice?	
3	How much influence does proximity to shopping centers and stores have on your residential location choice?	
4	How much does proximity to gardens and parks affect your residential location choice?	Public facilities
5	How much does proximity to medical and health centers affect your residential location choice?	
6	How much does amenities and access to service shops, such as laundry, etc., affect your residential location choice?	
7	How much does proximity to banks, government offices and service providers affect your residential location choice?	
8	How much does easy access to the bus station on foot affect your residential location choice?	Public transportation
9	How much does easy access to a taxi on foot affect your residential location choice?	
10	How much does the easy access to bicycle lanes affect your residential location choice?	
11	How much does proximity to relatives and acquaintances affect your residential location choice?	Social factors
12	How much influence does the neighborhood's reputation have on your residential location choice?	
13	How much does neighborhood security affect your residential location choice?	
14	How much does proximity to the central business district (CBD) affect your residential location choice?	House location
15	How much does proximity to the main street affect your residential location choice?	
16	With the current budget, buying a private house is my priority for choosing a house	House type
17	With the current budget, buying an apartment is my priority for choosing a house	
18	With the current budget, buying an apartment in residential towers is my priority for choosing a house	
19	With the current budget, buying an apartment in residential complexes is my priority for choosing a house	
20	How much does low traffic affect your residential location choice?	Traffic and pollution
21	How much does air pollution in the neighborhood affect your residential location choice?	

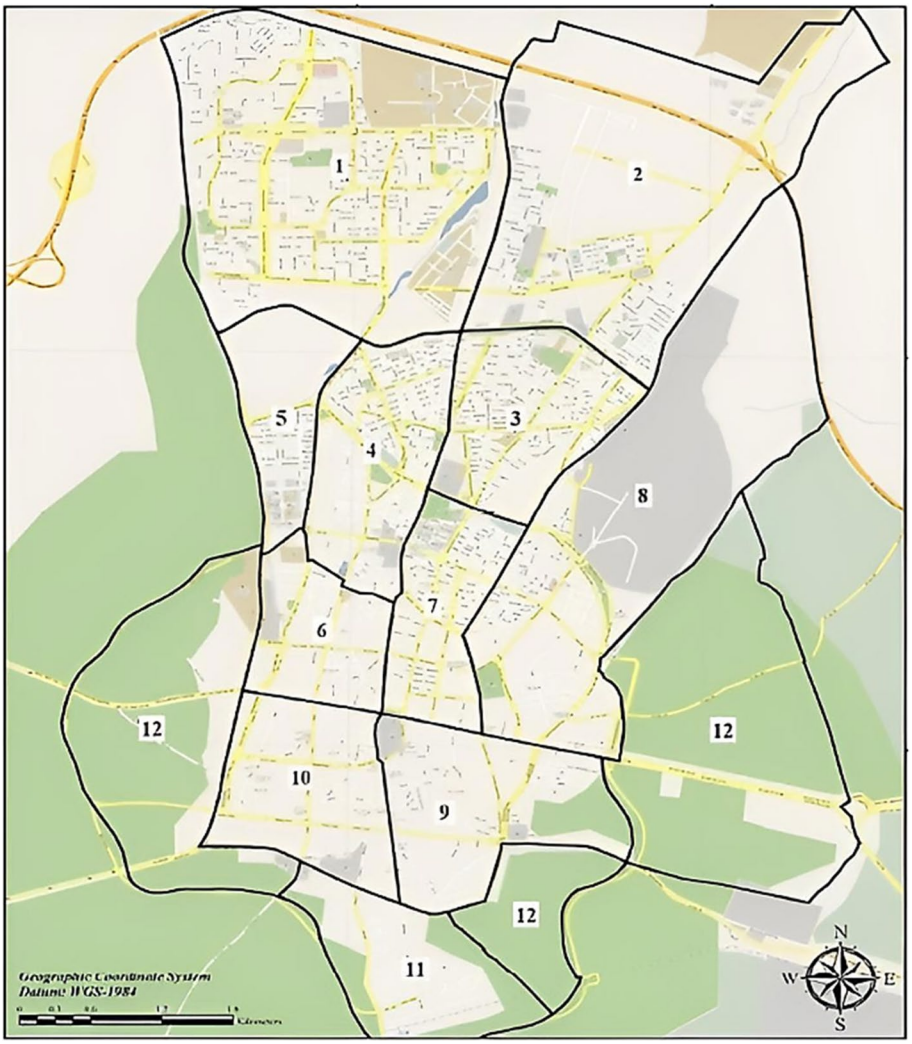


Fig. 2 Traffic division of study area

between the observed covariance matrix at each degree of freedom and the predicted covariance matrix. If the value of this index is smaller than 0.1, the model's fit is excellent (Jafarzadeh et al., 2023).

3 Results

In this section, after collecting data from the target population through sampling using the described methodology, modeling the residential location utility for different groups of households, based on their income level and the presence of children in education, was carried out through the SEM method.

3.1 Modeling the residential location utility

The values of the fit indices of the model are shown in Table 4. In the model's output, the value of the CMIN/DF index is reported as 2.58, which indicates that the presented model has the necessary usefulness for estimating the statistical population. Also, the value of the GFI index is 0.907 based on the results of the model, which is considered appropriate because it is higher than 0.9. In addition, based on the output of the model, the estimated value of 0.062 is calculated for the RMSEA index, which indicates the appropriateness of the model.

3.2 Estimation of model coefficients

In this section, the estimation of model coefficients is presented for all households, households with income less than 200$ per month, households with income above 800$ per month, households without children in education, and households with children in education.

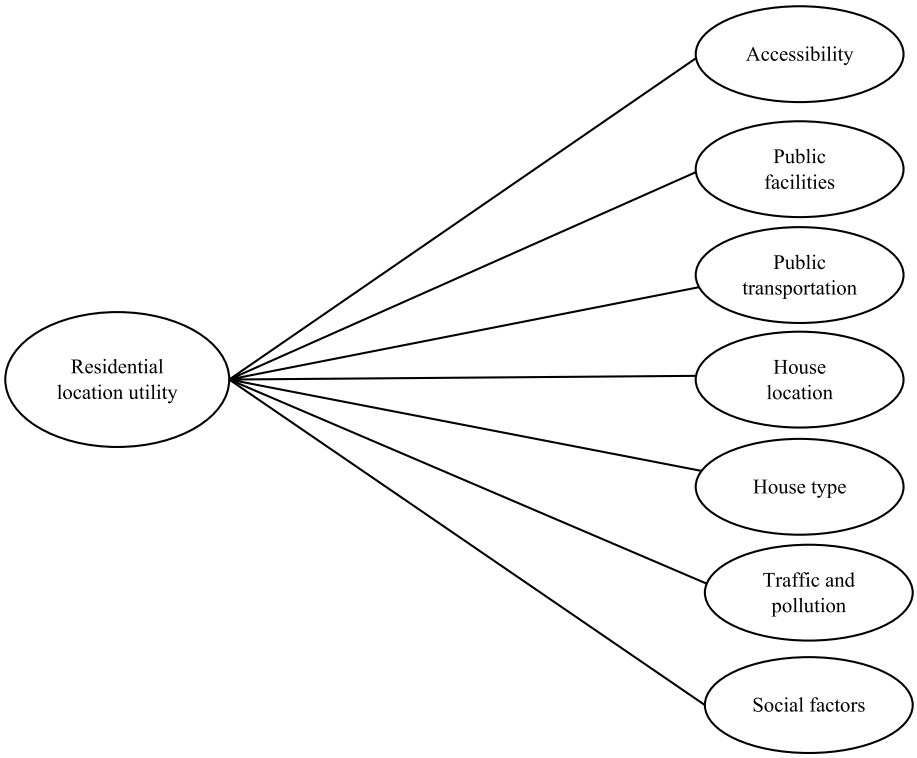


Fig. 3 Modeling factors affecting the residential location utility based on the SEM method

Table 4 Fit indices based on model output

Index	Values
CMIN/DF	2.58
GFI	0.907
RMSEA	0.062

3.2.1 Model estimation for all households

Table 5 shows the values of SEM coefficients in the output of AMOS software based on the data of all households.

According to the coefficients related to the *p*-value and t-statistic, all the variables except the tower variable in the house type category have statistical significance at the 95% confidence level. Also, according to the type of questions asked in the questionnaire, all the coefficients, except those for the residential complexes and towers in the house type category, are positive, which means, for example, proximity to transportation facilities as well as low pollution in the place of residence increases the residential location utility. Moreover, the highest coefficient among the variables defined in the model is related to the private house with a value of 0.837, which means that a private house has the highest impact on people's residential location choices. This result may reflect the problems associated with apartment living and the cultural and social issues that sometimes arise from living in certain types of buildings. Also, living in a private house gives people more independence in managing things related to their place of residence.

The coefficient related to air pollution is the second most important factor with a value of 0.817, which means that living in a place with less air pollution will interest the respondents. On the other hand, in the category related to accessibility, which is classified based on the activities of household members (access to shopping centers, workplace, and schools), access to shopping centers with a coefficient of 0.635 indicates the importance of accessibility of shopping centers for daily household purchases, more so than the accessibility of the workplace and schools of family members. Also, the proximity to the workplace has a coefficient of 0.354, which makes it the third most important factor, which can be due to the small size of Qazvin city and has caused the distance to the workplace in this city to be less important than the distance to shopping centers and stores as well as schools.

In the category of public facilities available near the place of residence, the health center has the highest importance with a coefficient of 0.727, and in fact, proximity to health centers is the third most important factor in the utility of choosing a house for a

Table 5 Values of model coefficients based on all household data

Variable	Standardized estimate	Unstandardized estimate	Standard error	t-statistic	*p*-value
hth	0.727	1.179	0.104	11.383	< 0.001
ame	0.539	0.746	0.082	9.149	< 0.001
ofc	0.537	0.851	0.102	8.343	< 0.001
wrk	0.384	1			
bus	0.59	1			
txi	0.624	0.923	0.069	13.426	< 0.001
bcy	0.447	0.715	0.099	7.213	< 0.001
rel	0.354	1			
rep	0.592	0.983	0.18	5.462	< 0.001
sec	0.545	0.712	0.133	5.341	< 0.001
cbd	0.624	1			
str	0.636	0.788	0.095	8.313	< 0.001
hos	0.838	1			
apt	0.281	0.258	0.078	3.322	< 0.001
trf	0.693	1			
pol	0.817	1.13	0.119	9.452	< 0.001
shp	0.635	1.219	0.168	7.27	< 0.001
cpx	-0.432	-0.52	0.139	-3.729	< 0.001
twr	-0.101	-0.115	0.073	-1.592	0.111
prk	0.594	1			
sch	0.553	1.401	0.204	6.905	< 0.001

place of residence. Part of the high importance of this variable can be the spread of the Coronavirus disease in recent years and the increase in visits of households to health centers. Access to gardens and parks is placed in the next category with a coefficient of 0.594. Moreover, in the category of public transportation, access to the taxi station is the most important with a coefficient of 0.624, and then, access to the bus station and bicycle lanes with a coefficient of 0.59 and 0.447 are ranked next.

In the house type category, the private house variable with a coefficient of 0.837 has the highest impact on choosing a place of residence for people, and the apartment variable with a coefficient of 0.281 and the tower variable with a coefficient of -0.101 are placed in the next categories, although the coefficient of the tower variable is not interpreted because its p-value is greater than 0.05 and it is not significantly different from zero. Also, living in the residential complexes has a negative effect on the utility of the place of residence with an importance coefficient of -0.432. Finally, in the category of social factors, the reputation variable has a coefficient of 0.594, which gives it the greatest impact on the utility of the place of residence, and proximity to relatives and acquaintances with a coefficient of 0.354 shows the least importance in the utility of the place of residence.

3.2.2 Model estimation for low-income households

In estimating the model of low-income households, households with income less than 200$ per month are used in the modeling. Based on the modeling output, the values of the coefficients are presented in Table 6.

According to the coefficients related to the p-value and t-statistic, all the variables except tower and apartment variables in the house type category, and security in the social factors category have statistical significance at the 95% confidence level.

According to the obtained results, the highest coefficient among the variables defined in the model for people with low income is related to the proximity to CBD with a coefficient of 0.821, which means that the proximity to CBD has the highest impact on choosing a place of residence for people with low income, which can be due to the cultural and geographical factors of Qazvin city. The next most important coefficients are proximity to schools with a coefficient of 0.793 in the utility of choosing a place of residence. Also, the coefficient related to air pollution has the second highest value of 0.817, which means that living in a place with less air pollution will be of interest to low-income households.

In the public transportation category, the proximity of the residential location to the taxi station and bus station are estimated respectively with a coefficient of 0.636 and 0.607, both of which are higher than the coefficient

Table 6 Values of model coefficients based on the data of low-income households

Variable	Standardized estimate	Unstandardized estimate	Standard error	t-statistic	*p*-value
hth	0.681	1.338	0.373	3.591	< 0.001
ame	0.617	0.992	0.291	3.407	< 0.001
ofc	0.629	1.33	0.436	3.05	0.002
wrk	0.429	1			
bus	0.607	1			
txi	0.636	0.812	0.17	4.781	< 0.001
bcy	0.599	1.215	0.33	3.684	< 0.001
rel	0.373	1			
rep	0.548	0.849	0.399	2.126	0.033
sec	0.442	0.791	0.407	1.845	0.052
cbd	0.821	1			
str	0.612	0.65	0.165	3.938	< 0.001
hos	0.503	1			
apt	0.26	0.384	0.207	1.851	0.064
trf	0.678	1			
pol	0.715	0.799	0.191	4.19	< 0.001
shp	0.685	1.076	0.379	2.84	0.005
cpx	-0.996	-1.96	0.702	-2.791	0.005
twr	-0.453	-0.939	0.494	-1.901	0.057
prk	0.499	1			
sch	0.793	1.735	0.583	2.976	0.003

estimated in the modeling of all households, and this issue can indicate the greater importance of access to public transportation in low-income households due to cost reduction and the economic nature of this issue for this group. Moreover, In the category of variables related to the house type, the private house shows more importance in determining the utility of choosing a place of residence than the other cases. Also, among the factors, access to shopping centers with a coefficient of 0.685 is more important for the respondents, and compared to all households, access to the workplace shows more importance among the group of all households.

3.2.3 Model estimation for high-income households

In estimating the model of high-income households, households with an income of more than 200$ per month are used in the modeling. Based on the modeling output, the values of the coefficients are presented in Table 7.

According to the coefficients related to the p-value and t-statistic, all the variables except for the variables of residential complexes in the category of house type and reputation variable in the category of social factors, as well as the variable of bicycle lanes in the category of public transportation are statistically significant at the 95% confidence level.

According to the obtained results, the highest coefficient among the variables defined in the model is related to the private house with a value of 1.23 value 6, which, like the group of all households, has the highest coefficient and is also more than it. This means that the personal house has the highest impact on the choice of residence for high-income households. This coefficient is more important than the group of all households. Regarding the tower variable, it can be seen that the coefficient of this variable is reported as 0.031, which, like the previously mentioned groups, is of low importance, but the significant point is that this coefficient is higher than that for low-income and all households. This issue can be due to the mentality of wealthy people towards towers that have more amenities. Residential complexes are also not statistically significant due to the p-value being more than 0.05 and it can be said that this variable can be removed from the model.

The air pollution variable related to the traffic and pollution category with a coefficient of 0.921 is the second most important variable, indicating the importance of air pollution for high-income households, like the previous groups. On the other hand, in the category of transportation factors, the proximity of the residence to the taxi station and bus station is estimated, respectively, with a coefficient of 0.46 and 0.617. Compared to low-income people, where the taxi station coefficient is 0.636, it can be said that access to a taxi station is less important for people with higher incomes than those with lower

Table 7 Values of model coefficients based on the data of high-income households

Variable	Standardized estimate	Unstandardized estimate	Standard error	t-statistic	*p*-value
hth	0.798	1.362	0.331	4.108	< 0.001
ame	0.297	0.516	0.281	2.835	0.024
ofc	0.414	0.606	0.277	2.189	0.029
wrk	0.267	1			
bus	0.617	1			
txi	0.46	0.727	0.128	5.699	< 0.001
bcy	0.178	0.273	0.215	1.224	0.221
rel	0.296	1			
rep	0.201	0.381	0.333	1.146	0.252
sec	0.704	1.489	0.751	1.982	0.048
cbd	0.679	1			
str	0.458	0.422	0.173	2.44	0.015
hos	1.236	1			
apt	0.027	0.003	0.065	4.652	< 0.001
trf	0.601	1			
pol	0.921	1.901	0.175	3.799	< 0.001
shp	0.627	1.838	1.064	2.727	0.025
cpx	-0.054	-0.009	0.198	-0.045	0.964
twr	0.031	0.005	0.102	2.145	0.034
prk	0.596	1			
sch	0.073	2.931	1.058	2.768	0.037

incomes, which can be naturally explained by the fact that wealthy people own more personal cars and have a greater tendency to use personal cars among other groups. The noteworthy point in the comparison of the bus station coefficient is the very close value of this coefficient among the groups of all, low-income and high-income households, which indicates the almost equal sensitivity of this variable in all these groups regarding this variable. Finally, in the category of public facilities, health centers have the highest importance with a coefficient of 0.798, and in fact, proximity to health centers is the third most important factor in the utility of choosing a house for a place of residence in this group. This issue shows the great importance of this variable among high-income households.

3.2.4 Model estimation for households without children in education

In the estimation of this model, households without children in education are used in modeling. Based on the modeling output, the coefficient values are illustrated in Table 8.

According to the coefficients related to the p-value and t-statistic, all the variables except the tower variable in the house type category have statistical significance at the 95% confidence level. Also, the highest coefficient among the variables defined in the model is related to air pollution with a value of 0.833 in the traffic and pollution category, which indicates the priority of this variable in the utility of the place of residence for this group of households. In the next level, a private house with a coefficient of 0.708 is in the second place of importance for the utility of the place of residence, which, like the previous groups, shows the importance of this variable. Among the accessibility factors, the proximity to shopping centers has a higher coefficient (0.613) than proximity to schools (0.554) and proximity to the workplace (0.356), indicating that it is a more important factor. In this group of households, the proximity to the workplace has the least effect on the utility of the place of residence, and the proximity to shopping centers is more important.

In the public facilities category, the health center has the highest importance with a coefficient of 0.666. This issue shows the great importance of this variable among this group of households, like other groups. Also, in the category of public transportation, the proximity of the residence to the taxi station and bus station are estimated respectively with a coefficient of 0.553 and 0.449, which is less important compared to other investigated groups, indicating that households without children in education are less sensitive to the proximity to public transportation stations than other groups. Moreover, in the category of house type for the place of residence, like all the previous groups, the private house variable is the most important

Table 8 Values of model coefficients based on the data of households without children in education

Variable	Standardized estimate	Unstandardized estimate	Standard error	t-statistic	*p*-value
hth	0.66	1.1	0.126	8.74	< 0.001
ame	0.526	0.74	0.1	7.397	< 0.001
ofc	0.527	0.868	0.129	6.752	< 0.001
wrk	0.356	1			
bus	0.449	1			
txi	0.553	1.082	0.116	9.331	< 0.001
bcy	0.439	0.96	0.169	5.675	< 0.001
rel	0.318	1			
rep	0.605	1.163	0.279	4.171	< 0.001
sec	0.571	0.911	0.221	4.123	< 0.001
cbd	0.68	1			
str	0.681	0.768	0.103	7.475	< 0.001
hos	0.708	1			
apt	0.383	0.407	0.103	3.958	< 0.001
trf	0.673	1			
pol	0.833	1.158	0.146	7.917	< 0.001
shp	0.613	1.178	0.201	5.397	< 0.001
cpx	-0.498	-0.703	0.163	-4.302	< 0.001
twr	-0.079	-0.107	0.109	-0.985	0.324
prk	0.563	1			
sch	0.554	1.542	0.28	5.507	< 0.001

with a coefficient of 0.708, followed by the apartment variable with a coefficient of 0.383 and the residential complexes variable with a coefficient of -0.489.

3.3 Model estimation for households with children in education

In estimating this model, the information on households with children in education is used in modeling. Based on the modeling output, the values of the coefficients are presented in Table 9.

According to the coefficients related to the *p*-value and t-statistic, all the variables, except for the variables of tower and residential complexes, in the category of house type, have statistical significance at the 95% confidence level. According to the obtained results, the highest coefficient among the variables is related to air pollution with a value of 0.922 in the traffic and pollution category, followed by the private house and proximity to a bus station with coefficients of 0.895 and 0.892, respectively. The noteworthy point is the high importance of proximity to the bus station in this group (households with children in education) compared to the other groups, which can be due to the importance of this item for children's commuting to schools in this group of households.

Among the accessibility factors, proximity to shopping centers has a higher coefficient (0.674) than proximity to schools (0.624), indicating that it is a slightly more important factor. The degree of importance of proximity to educational centers in this group of households is predictable, and naturally, households with children in education have a greater desire for schools to be close to their place of residence. Also, the proximity to the workplace with a coefficient of 0.377 is in the last rank in this category. On the other hand, in the public transit category, the proximity of the residence to the taxi station and bus station are estimated respectively with coefficients of 0.775 and 0.892. Compared to the group of households without children in education, both variables have higher coefficients, and this issue shows the great importance of access to bus and taxi stations in the desirability of the residence for this group of households, which naturally this issue can be due to the use of public transportation by children in education for commuting to schools.

In the public facilities category, health centers have the highest importance with a coefficient of 0.771. In the house type category, like all the previous groups of households, the private house variable is the most important with a coefficient of 0.895, followed by the residential complexes variable with a coefficient of 0.082, the apartment variable with a coefficient of 0.048, and the tower variable with a coefficient of 0.015. According to the coefficients related to the *p*-value statistic, the coefficients related to the residential complexes and tower variables

Table 9 Values of model coefficients based on the data of households with children in education

Variable	Standardized estimate	Unstandardized estimate	Standard error	t-statistic	*p*-value
hth	0.771	1.495	0.309	4.845	< 0.001
ame	0.63	0.994	0.231	4.301	< 0.001
ofc	0.536	0.902	0.26	3.476	< 0.001
wrk	0.337	1			
bus	0.892	1			
txi	0.775	0.764	0.105	7.266	< 0.001
bcy	0.506	0.488	0.132	3.682	< 0.001
rel	0.532	1			
rep	0.259	0.244	0.116	3.103	0.035
sec	0.251	0.162	0.079	2.044	0.041
cbd	0.547	1			
str	0.435	0.643	0.181	3.543	< 0.001
hos	0.895	1			
apt	0.048	0.008	0.138	3.044	0.003
trf	0.561	1			
pol	0.922	1.765	0.576	3.066	0.002
shp	0.674	1.496	0.503	2.977	0.003
cpx	-0.082	-0.017	0.3	-0.056	0.956
twr	0.015	0.003	0.047	0.056	0.956
prk	0.545	1			
sch	0.624	1.551	0.533	2.909	0.004

have no statistical significance, and these two variables can be removed from the model.

4 Discussion

According to the modeling results, a comparison was performed regarding the changes in the model estimation coefficients in the household groups, including low and high-income households, households with and without children in education, as well as all household groups.

4.1 Comparison of low and high-income households

Figure 4 shows the changes in the model estimation coefficient in low and high-income household groups.

In the initial comparison, a large difference in the coefficient related to a private house can be seen in the high-income group compared to the low-income group. Considering the high price of private houses in Qazvin city, this difference in the coefficient can naturally be explained by the financial ability of high-income households, as well as the cultural problems related to apartment living, and the feeling of more privacy in private houses can explain this difference in the group that can buy or rent such houses. In terms of air pollution, it is evident that high-income groups are more sensitive to this factor than low-income groups, and in addition to the comparison of the coefficient of health centers between these two groups, where high-income households have a higher coefficient, it may be concluded that with the increase in income, the importance of access to living places with lower pollution and access to health centers in high-income households will be much greater.

In the group of low-income households, access to the taxi station, as well as the proximity to the workplace, have a relatively large difference compared to the high-income group, and the importance of proximity to the workplace and access to the taxi station can be attributed to the lower rate of personal car ownership among low-income people compared to high-income people, for this reason, low-income households tend to choose a place of residence that is closer to the workplace and has easier access to the taxi station due to transportation costs to the workplace. This issue also applies to the coefficients related to access to schools and shopping centers. In the category of social factors, the security of the neighborhood has been more important for high-income people than for low-income people, while in terms of proximity to relatives and acquaintances, the coefficient of this factor is higher in low-income households than in higher-income households.

4.2 Comparison of households with and without children in education

Figure 5 shows the changes in the model estimation coefficient in households with and without children in education.

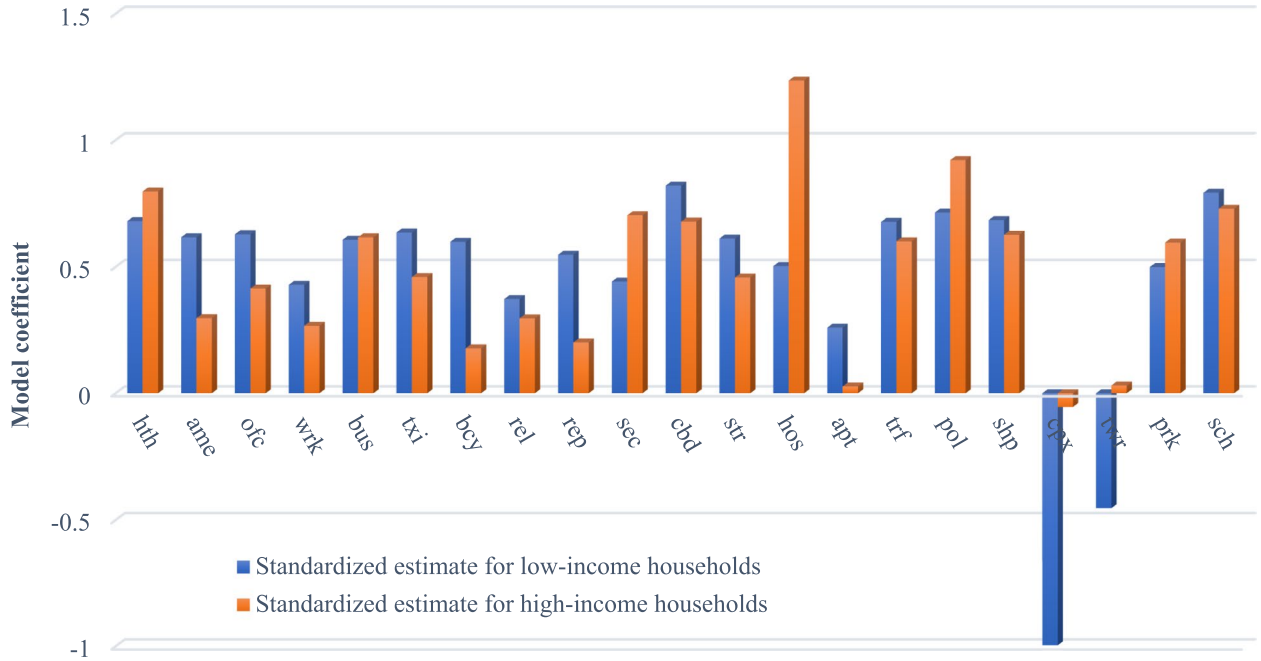


Fig. 4 Comparison of estimation coefficients of low-income households and high-income households

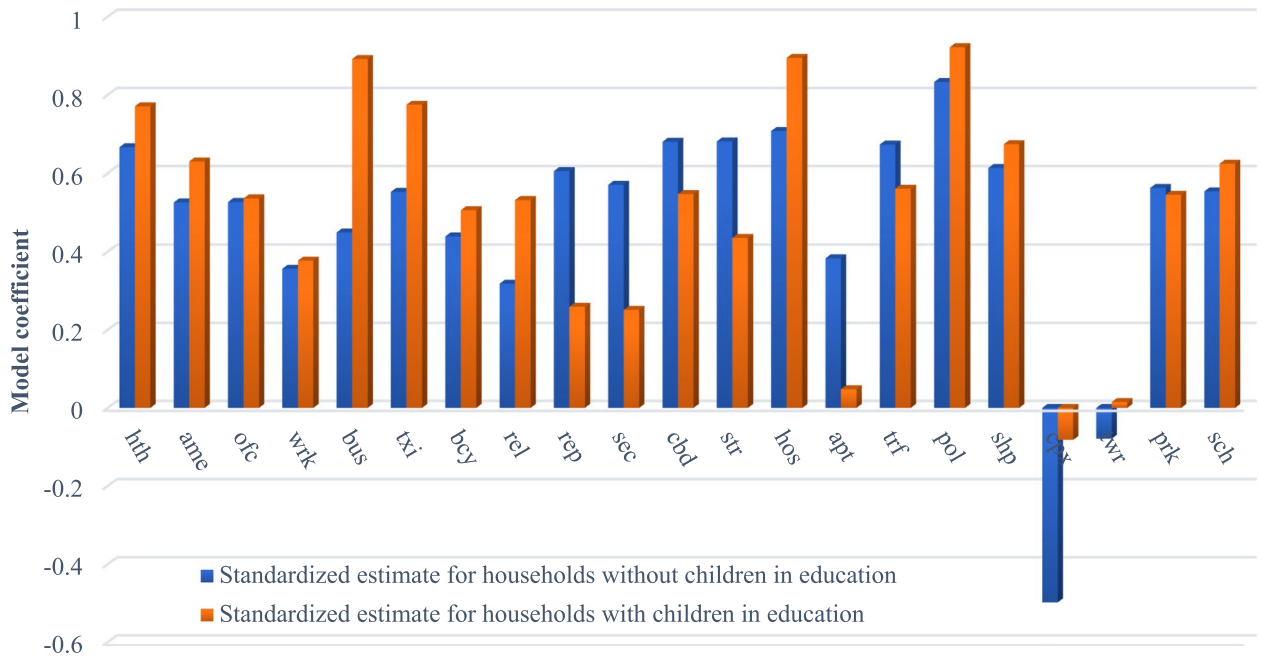


Fig. 5 Comparison of estimation coefficients of households with and without children in education

In the initial comparison, a large difference in the coefficient related to access to taxi and bus stations is evident in the group of households with children in education compared to households without children in education. This issue can be due to the greater importance of public transportation for households with children in education, which is predictable. Also, regarding access to schools and educational centers, as can be expected, in households with children in education, the proximity of the house to educational centers has a greater effect on

increasing the utility of the place of residence. Moreover, in the category of social factors, the coefficient of proximity to relatives and acquaintances in households with children in education is higher than in households without children in education, which can be explained by the fact that if the parents are working, relatives and acquaintances such as grandfather or grandmother can be more easily accessible to take care of the children after school hours and when they return from school.

4.3 Comparison of all household groups

Figure 6 shows the model estimation changes in different household groups under study. In this figure, the rate of change of the coefficients related to each effective factor in the utility of the place of residence is displayed.

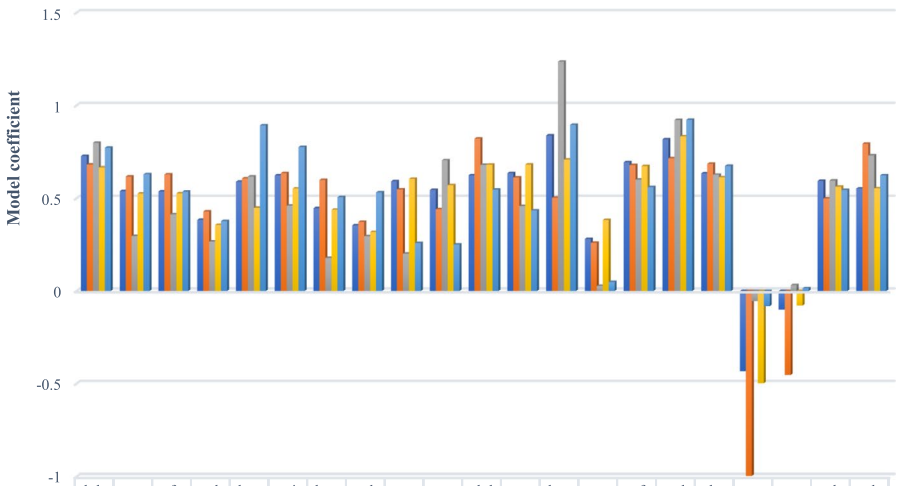
According to Fig. 6, the degree of importance of the house type can be seen in different groups, which indicates the greater impact of a private house among high-income people. It can also be seen that residential complexes and towers have negative coefficients in all groups, which indicates the unpopularity of this house type among the residents of Qazvin city. In the public facility category, access to health centers in all groups is almost the same and has high importance. Regarding access to bicycle lanes, a significant difference can be observed between the high-income group and other groups, which indicates that this type of public transportation facility is less important in this group. Finally, considering access to taxi and bus stations, the importance of these types of facilities in groups with children in education is significant.

5 Conclusion

This study aimed to investigate the effective factors in residential location choice in Qazvin city, Iran. Using a questionnaire based on the stated preference (SP) method, the relationship between various transportation factors and factors related to residential location choice was examined, and the importance of each factor was evaluated using structural equation modeling (SEM).

The findings indicate that among various factors, the house type, specifically the private house, is more important in increasing the utility of the place of residence in all household groups. The private house was ranked first in creating the utility of the place of residence for all households, with a coefficient of 0.838. In both low and high-income groups, this factor was of great importance and ranked highly among choice priorities. In the high-income group, with a coefficient of 1.236, the effect of the private house on the utility of the place of residence was significant. Conversely, the negative impact of residential complexes on different groups shows that their construction is not desirable for the residents of Qazvin city.

Proximity to health centers was a concern for respondents in terms of the utility of the place of residence, ranked second among other public facility variables with a coefficient of 0.727. Air pollution was also an important variable in all models, with a coefficient of 0.922 for households with children in education. In terms of access



	hth	ame	ofc	wrk	bus	txi	bcy	rel	rep	sec	cbd	str	hos	apt	trf	pol	shp	cpx	twr	prk	sch
Standardized estimate for all households	0.73	0.54	0.54	0.38	0.59	0.62	0.45	0.35	0.59	0.55	0.62	0.64	0.84	0.28	0.69	0.82	0.64	-0.4	-0.1	0.59	0.55
Standardized estimate for low-income households	0.68	0.62	0.63	0.43	0.61	0.64	0.6	0.37	0.55	0.44	0.82	0.61	0.5	0.26	0.68	0.72	0.69	-1	-0.5	0.5	0.79
Standardized estimate for high-income households	0.8	0.3	0.41	0.27	0.62	0.46	0.18	0.3	0.2	0.7	0.68	0.46	1.24	0.03	0.6	0.92	0.63	-0.1	0.03	0.6	0.73
Standardized estimate for households without studying children	0.67	0.53	0.53	0.36	0.45	0.55	0.44	0.32	0.61	0.57	0.68	0.68	0.71	0.38	0.67	0.83	0.61	-0.5	-0.1	0.56	0.55
Standardized estimate for households with studying children	0.77	0.63	0.54	0.38	0.89	0.78	0.51	0.53	0.26	0.25	0.55	0.44	0.9	0.05	0.56	0.92	0.67	-0.1	0.02	0.55	0.62

Fig. 6 Comparative chart of model estimation coefficients in all household groups

to public transportation, access to taxi and bus stations was estimated to have higher coefficients for households with children in education than those without children, with coefficients of 0.775 and 0.892, respectively. This indicates that the use of public transportation by children and proper access to it are of interest to households with children.

This study found that proximity to the workplace had relatively low importance in the utility of the place of residence, with a coefficient of 0.384 for all households, and less importance in other groups. This can be attributed to urban conditions and the relatively small size of Qazvin city. In terms of social factors, proximity to relatives and acquaintances had a lower effect on the utility of the place of residence than the reputation and security of the neighborhood.

Future research can benefit from collecting information from heads of households and considering additional variables and personality factors to better understand the influence of these characteristics on residential location choice. Additionally, examining the effects of land use changes on neighborhood attractiveness can provide insights into urban policies and their consequences. Also, another direction for future research is to apply some spatial regression models for comparative performance studies, such as the spatial error model, spatial lag model, and geographically weighted regression. These models can account for the spatial dependence and heterogeneity in the data, and provide more accurate and efficient estimates of the regression parameters. Moreover, different methods of statistical analysis and machine learning methods can be applied in the continuation of this study (Han & Fu, 2023; Jandaghi et al., 2023; Shen et al., 2018), which can help to discover the patterns, relationships, or trends in the data, and to predict or explain the outcomes of the residential location choice.

Author contributions

H. Rajabi and H. Mirzahossein conceived of the presented idea. H. Rajabi and H. Mirzahossein developed the theory and performed the computations H. Rajabi, H. Mirzahossein and S.M. Hosseinian verified the analytical methods. X. Jin and H. Mirzahossein supervised the findings of this work. All authors discussed the results and contributed to the final manuscript.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Availability of data and materials

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Consent for publication

Authors consent for the publication of the submitted paper and any associated data and accompanying images.

Competing interests

The authors reported no potential conflict of interest.

Received: 17 October 2023 Revised: 24 December 2023 Accepted: 4 January 2024
Published online: 17 January 2024

References

Aslam, A. B., Masoumi, H. E., Naeem, N., & Ahmad, M. (2019). Residential location choices and the role of mobility, socioeconomics, and land use in Hafizabad. *Pakistan. Urbani Izziv, 30*(1), 115–128.

Ballesteros, L., Ourang, S., Pazdon, J., & Kuffel, K. (2023). Increasing safety in residential construction through a simplified earthquake-and typhoon-resistant guidelines. *Journal of Future Sustainability, 3*(3), 119–124.

Brown, D. G., & Robinson, D. T. (2006). Effects of heterogeneity in residential preferences on an agent-based model of urban sprawl. *Ecology and society*, *11*(1).

Cheng, L., Chen, X., Yang, S., Wu, J., & Yang, M. (2019). Structural equation models to analyze activity participation, trip generation, and mode choice of low-income commuters. *Transportation Letters, 11*(6), 341–349.

Cockx, K., & Canters, F. (2020). Determining heterogeneity of residential location preferences of households in Belgium. *Applied Geography, 124*, 102271.

Dehghani, A., Alidadi, M., & Sharifi, A. (2022). Compact development policy and urban resilience: a critical review. *Sustainability, 14*(19), 11798.

Dehghani, A., Alidadi, M., & Soltani, A. (2023). Density and urban resilience, cross-section analysis in an Iranian metropolis context. *Urban Science, 7*(1), 23.

Fu, X., Bhat, C. R., Pendyala, R. M., Vadlamani, S., & Garikapati, V. M. (2014). *Understanding the multiple dimensions of residential choice*.

García-Santillán, A., Escalera-Chávez, M. E., & Córdova-Rangel, A. (2012). Variables to measure interaction among mathematics and computer through structural equation modeling. *Journal of Applied Mathematics and Bioinformatics, 2*(3), 51.

Hair, J. F., Ringle, C. M., Gudergan, S. P., Fischer, A., Nitzl, C., & Menictas, C. (2019). Partial least squares structural equation modeling-based discrete choice modeling: an illustration in modeling retailer choice. *Business Research, 12*(1), 115–142.

Han, C., & Fu, X. (2023). Challenge and opportunity: deep learning-based stock price prediction by using Bi-directional LSTM model. *Frontiers in Business, Economics and Management, 8*(2), 51–54.

Haoran, L., Qiming, L., & Ying, L. (2019). Analysis of key safety risks in prefabricated building construction based on structural equation model. *China Safety Science Journal, 29*(4), 171.

Hoshino, T. (2013). Estimation of the preference heterogeneity within stated choice data using semiparametric varying-coefficient methods. *Empirical Economics, 45*(3), 1129–1148.

Hunt, J. D. (2001). Stated preference analysis of sensitivities to elements of transportation and urban form. *Transportation Research Record, 1780*(1), 76–86.

Hunt, J. D. (2010). Stated preference examination of factors influencing residential attraction. In *Residential Location Choice* (pp. 21–59). Springer.

Izadi, A., Jamshidpour, F., Safari, D., & Gilani, V. (2020). Accident analysis of bus rapid transit system: Before and after construction. *Eur. Transp. Trasp. Eur*, *79*(9).

Jafarzadeh, E., Kabiri-Samani, A., Boroomand, B., & Bohluly, A. (2023). Analytical modeling of flexible circular submerged mound motion in gravity waves. *Journal of Ocean Engineering and Marine Energy, 9*(1), 181–190.

Jandaghi, E., Chen, X., & Yuan, C. (2023, 28–30 June 2023). Motion Dynamics Modeling and Fault Detection of a Soft Trunk Robot. 2023 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM),

Khattak, A. J., & Rodriguez, D. (2005). Travel behavior in neo-traditional neighborhood developments: a case study in USA. *Transportation Research Part a: Policy and Practice, 39*(6), 481–500.

Khodayari, M., Razmi, J., & Babazadeh, R. (2019). An integrated fuzzy analytical network process for prioritisation of new technology-based firms in Iran. *International Journal of Industrial and Systems Engineering, 32*(4), 424–442.

Lasley, P. (2017). *The Influence of Transportation on Residential Choice: A Survey of Texas REALTORS® on Factors Affecting Housing Location Choice*

Le, T., & Le, C. (2022). Action Sequencing in Construction Accident Reports using Probabilistic Language Model. ISARC. *Proceedings of the International Symposium on Automation and Robotics in Construction, 39*, 653–660.
Li, T., Sun, H., Wu, J., & Lee, D.-H. (2020). Household residential location choice equilibrium model based on reference-dependent theory. *Journal of Urban Planning and Development, 146*(1), 04019024.
Mak, B., Sim, J., & Jones, D. (2005). Model of service quality: Customer loyalty for hotels. *Hospitality Review, 23*(1), 9.
Masoumi, H. (2021). Residential location Choice in Istanbul, Tehran, and Cairo: the importance of commuting to work. *Sustainability, 13*(10), 5757.
Nesamani, K., Saphores, J.-D., McNally, M. G., & Jayakrishnan, R. (2017). Estimating impacts of emission specific characteristics on vehicle operation for quantifying air pollutant emissions and energy use. *Journal of Traffic and Transportation Engineering (english Edition), 4*(3), 215–229.
Ourang, S. (2022). Evaluation of Inter-Organizational Coordination of Housing Services in Rural Alaska Through Social Network Analysis.
Schirmer, P. M., Van Eggermond, M. A., & Axhausen, K. W. (2014). The role of location in residential location choice models: a review of literature. *Journal of Transport and Land Use, 7*(2), 3–21.
Shen, G., Han, C., Chen, B., Dong, L., & Cao, P. (2018). Fault analysis of machine tools based on grey relational analysis and main factor analysis. *Journal of physics: conference series, 1069*(1), 012112.
Walker, J. L., & Li, J. (2007). Latent lifestyle preferences and household location decisions. *Journal of Geographical Systems, 9*(1), 77–101.
Wang, Y., Cao, M., Liu, Y., Ye, R., Gao, X., & Ma, L. (2020). Public transport equity in Shenyang: Using structural equation modelling. *Research in Transportation Business & Management, 42*, 100555.
Yarahmadi, F. (2020). *Multistage sampling technique and estimating sample size for a descriptive study on viewers' perception of tv commercials*. SAGE Publications Ltd.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.