Article

Urban Studies

Urban Studies
1–20

DOI: 10.1177/0042098015593023
usj.sagepub.com
SAGE

Dwelling unit choice in a condominium complex: Analysis of willingness to pay and preference heterogeneity

Hai Jiang
Tsinghua University, China

Shuiping Chen
Tsinghua University, China

Abstract
We study the choice of dwelling units in a condominium complex, a problem that has not been thoroughly investigated in the housing research literature. We take a revealed preference approach to quantify home buyers' preference toward attributes associated with a dwelling unit. In particular, we estimate a mixed logit model using the hierarchical Bayes approach based on the Markov Chain Monte Carlo method. We derive the willingness to pay measures for detailed attributes associated with a dwelling unit including its floor level, orientation, location in the complex, location on a floor level, and the type of bathrooms. We also find strong evidence for the preference heterogeneity among home buyers and conduct regression analysis to explain the preference heterogeneity using home buyers' socioeconomic characteristics. Our results show that home buyers with older ages and higher annual household income tend to focus more on the quality of the dwelling units, while first-time home buyers are more willing to accept dwelling units with less-desirable attributes.

Keywords
condominium complex, dwelling unit choice, hierarchical Bayes, mixed logit, regression analysis

Received August 2013; accepted May 2015

Introduction

Condominium living is the most common residential choice for homeowners in China.[1] It carries many desirable characteristics, such as affordable housing, low maintenance requirements, enhanced security features and, often times, easy access to restaurants, malls and other urban amenities. A condominium complex typically consists of low-rise and high-rise residential buildings, which are divided into a collection of private

Corresponding author:
Hai Jiang, Department of Industrial Engineering, Tsinghua University, Shunde Building, Room 517, Beijing 100084, China.
Email: haijiang@tsinghua.edu.cn

dwelling units. Each dwelling unit is owned by and registered in the name of the purchaser of the unit. Real estate developers price these dwelling units differently based on the popularity of their attributes, for example, size, floor level, orientation, proximity to street and so on. Currently, real estate developers in China rely heavily on the experience of the salespersons to set the prices of dwelling units (Liao, 2011). Hence, they are very interested in understanding how prospective homeowners make their choices of dwelling units within a condominium complex, particularly how they value the different attributes associated with these units. Such knowledge can help real estate developers better differentiate the prices of dwelling units and provide valuable insights into the design of new residential complexes in the future. It is also of interest to real estate developers to quantify and explain the preference heterogeneity, if any, among home buyers, so that the developers can better match prospective home buyers with their preferred dwelling units in the sales process.

Literature on dwelling unit choice in a condominium complex is very scarce. In an effort to ascertain the influence of the Green Building Certification Scheme on prospective homeowners, Heinzle et al. (2013) conducted a stated preference survey to study home buyers' preferences toward various attributes of condominium complexes and the dwelling units inside. The focus was on determining home buyers' willingness to pay (WTP) for properties with the Green Building Certification. Ishikawa et al. (2011) examined how individuals read and comprehend the floor plan of a condominium unit. Subjects were asked to classify various types of floor plans and results were analysed to identify floor plan attributes that are important. There are also a few remotely related studies investigating the choice of home types, for example, single-family detached house, townhouse, apartment and so on. However, they do not model the detailed characteristics of the dwelling units. A sampling of this research includes Quigley (1976), Tu and Goldfinch (1996), Skaburskis (1999) and Habib and Kockelman (2008). For a more comprehensive survey on the residential choice literature in general, interested readers are referred to Yates and Mackay (2006), Barrios Garcia and Rodriguez Hernandez (2007), Hoshino (2011) and Wu et al. (2013).

To the best of the authors' knowledge, this paper is the first of its kind to investigate the choice of dwelling units within a condominium complex. Studies of residential preferences generally employ one of two types of data, that is, stated preference (SP) data based on respondents' choice under hypothetical situations; or revealed preference (RP) data based on actual market behaviour. Their respective strengths and limitations are thoroughly discussed in Train (2003). In this research, we take a RP approach and use actual sales data from a real estate developer in Beijing to analyse dwelling unit choice. The data were collected in the second quarter of 2012 and contain detailed information about the dwelling units sold as well as the socioeconomic characteristics of the home buyers.

We first estimate a multinomial logit model, which is a widely used discrete choice model in the housing research literature and use it as the base model for comparison. The attributes we examine include location of the dwelling unit in the complex (whether it is in the centre of the complex or adjacent to city streets), location of the dwelling unit in the residential building (its floor level and location on the floor), orientation, characteristics of bathrooms, size of bay windows, size of the unit, purchase price and so on. We then estimate a mixed logit model that allows us to assess the heterogeneity among home buyers. The model is estimated using a hierarchical Bayes approach based on the Markov

Chain Monte Carlo (MCMC) method. Results from the mixed logit model reveal strong evidence for the behavioural heterogeneity among home buyers. Besides interpreting the results for the mixed logit model through the conventional 'random coefficients' perspective, we also take the 'error components' approach (Train, 2003) to analyse the substitution or correlation patterns among dwelling units. We show that complex correlations exist among dwelling units and the mixed logit model is more appropriate in this problem context.

Based on the the WTP estimates produced by the mixed logit model, we derive the price adjustments real estate developers make for dwelling units with different attributes and characterise the ideal dwelling unit in Chinese home buyers' minds. We also identify factors that influence home buyers' choice and discover that the floor level and orientation are the top two factors affecting dwelling unit choice, followed by the location in the complex, the location on a floor level and, finally, the types of bathrooms in the dwelling unit.

To uncover the socioeconomic causes for the preference heterogeneity among home buyers, we estimate each individual's WTP conditional upon his or her observed choice. We then conduct regression analysis to identify the correlation between home buyers' WTP estimates and their socioeconomic characteristics. We find a reasonable relationship between the WTP estimates and the explanatory variables. For example, the regression results suggest that home buyers with older ages and higher annual household income focus more on the quality of the dwelling units, that is, they prefer dwelling units located on good floor levels and whose orientation is north- and south-facing. It is also found that first-time home buyers are more willing to accept dwelling units with less-desirable attributes. These results can assist real estate developers in tailoring their offerings to different customer segments to improve customer satisfaction as well as their profitability.

Methodology

This paper investigates dwelling unit choice in a condominium complex. To address this discrete choice problem, we use the multinomial and mixed logit models (Ben-Akiva and Lerman, 1985; Train, 2003). The multinomial logit model is a widely used discrete choice model in the housing choice literature, while the mixed logit model is well suited to discover behavioural heterogeneity among decision makers.

From the discrete choice perspective, home buyers are assumed to attain some level of utility when purchasing a dwelling unit in the condominium complex. Specifically, for home buyer $n (n = 1, 2, \cdots, N)$, let C_n be the set of dwelling units available to him or her, that is, the set of unsold dwelling units in the condominium complex before the home buyer finally purchases. Let y_{in} take value 1 if home buyer n chooses dwelling unit $i \in C_n$; and 0, otherwise. Define $\boldsymbol{y}_n = (y_{1n}, y_{2n}, \cdots, y_{|C_n|n})$, where $|C_n|$ is the cardinality of set C_n. For dwelling unit $i \in C_n$, its utility function for home buyer n can be written as:

$$U_{in} = \boldsymbol{\beta}'_n \boldsymbol{x}_{in} + \in_{in} , \qquad (1)$$

where $\boldsymbol{x}_{in}$ is a vector of observed attributes related to dwelling unit i for home buyer n, $\boldsymbol{\beta}_n$ is the coefficient vector of these variables for this buyer, and $\in_{in}$ is a random term, representing the unobserved utility component. Let κ be the length of vector $\boldsymbol{\beta}_n$. The probability for home buyer n to choose dwelling unit i is then defined as:

$$P_n(i|C_n) = P(U_{in} \geq U_{jn}|j \neq i, j \in C_n). \qquad (2)$$

Depending on the assumptions on the distributions of $\boldsymbol{\beta}_n$ and $\in_{in}$, we can obtain various discrete choice models.

The multinomial logit model

The multinomial logit model is a widely applied discrete choice model in the housing choice literature because of its simplicity in estimation, its closed form specification and its robustness to the very strong Independent from Irrelevant Alternatives (IIA) assumption imposed on the error terms.

If we assume that the coefficient vector $\boldsymbol{\beta}_n$ is constant across home buyers (that is, $\boldsymbol{\beta}_n = \boldsymbol{\beta}, \forall n$) and $\in_{in}$ is independently identically Gumbel distributed across home buyers and alternatives, we obtain the multinomial logit model. The probability that home buyer n chooses dwelling unit i is:

$$P_n(i|C_n) = \frac{\exp(\boldsymbol{\beta}' x_{in})}{\sum_{j \in C_n} \exp(\boldsymbol{\beta}' x_{jn})}. \tag{3}$$

The coefficient vector $\boldsymbol{\beta}$ is estimated by maximising the likelihood function based on all individuals in the sample, that is, $L(\boldsymbol{\beta}) = \Pi_{n=1}^{N} L(\boldsymbol{y}_n)$, where $L(\boldsymbol{y}_n)$ is the probability of home buyer n's observed choices and can be expressed as:

$$L(\boldsymbol{y}_n) = \prod_{i \in C_n} \left(\frac{\exp(\boldsymbol{\beta}' \boldsymbol{x}_{in})}{\sum_{j \in C_n} \exp(\boldsymbol{\beta}' \boldsymbol{x}_{jn})} \right)^{y_{in}}. \tag{4}$$

The mixed logit model

The mixed logit model relaxes the assumption that $\boldsymbol{\beta}_n$ is constant across home buyers and allows us to assess the heterogeneity among home buyers as well as the correlation among dwelling units (Train, 2003). In the mixed logit model, $\boldsymbol{\beta}_n$ is assumed to vary over individuals to capture their taste variations. That is, β_n is decomposed into mean $\boldsymbol{b}$ and deviation ζ_n. Often times ζ_n is assumed to conform to a multivariate normal distribution $N(\mathbf{0}, \boldsymbol{W})$ where $\boldsymbol{W}$ is the variance-covariance matrix. The density function of $\boldsymbol{\beta}_n$ can therefore be written as $\phi(\boldsymbol{\beta}_n|\boldsymbol{b}, \boldsymbol{W})$.

The probability that home buyer n chooses alternative i, conditional on $\boldsymbol{\beta}_n$, is:

$$P_n(i|\boldsymbol{\beta}_n, C_n) = \frac{\exp(\boldsymbol{\beta}'_n \boldsymbol{x}_{in})}{\sum_{j \in C_n} \exp(\boldsymbol{\beta}'_n \boldsymbol{x}_{jn})}. \tag{5}$$

The probability of home buyer n's observed choices, conditional on $\boldsymbol{\beta}_n$ is:

$$L(\boldsymbol{y}_n|\boldsymbol{\beta}_n) = \prod_{i \in C_n} \left(\frac{exp(\boldsymbol{\beta}'_n \boldsymbol{x}_{in})}{\sum_{j \in C_n} exp(\boldsymbol{\beta}'_n \boldsymbol{x}_{jn})} \right)^{y_{in}}. \tag{6}$$

The unconditional probability is therefore the integral of $L(\boldsymbol{y}_n|\boldsymbol{\beta}_n)$ over $\boldsymbol{\beta}_n$:

$$L(\boldsymbol{y}_n|\boldsymbol{b}, \boldsymbol{W}) = \int L(\boldsymbol{y}_n|\boldsymbol{\beta}_n)\phi(\boldsymbol{\beta}_n|\boldsymbol{b}, \boldsymbol{W})\mathrm{d}\boldsymbol{\beta}_n \tag{7}$$

and the final likelihood function is:

$$L(\boldsymbol{b}, \boldsymbol{W}) = \Pi_{n=1}^{N} L(\boldsymbol{y}_n|\boldsymbol{b}, \boldsymbol{W}). \tag{8}$$

Parameters $\boldsymbol{b}$ and $\boldsymbol{W}$ can be estimated using the hierarchical Bayes approach based on the MCMC method, which offers two advantages over the maximum simulated likelihood approach: (1) it does not attempt to maximise the likelihood function, which can be very challenging numerically and convergence can be hard to achieve; and (2) desirable estimation properties, such as consistency and efficiency, can be obtained under more relaxed conditions (Train, 2003). Let $k(\boldsymbol{b}, \boldsymbol{W})$ be the prior density function for $\boldsymbol{b}$ and $\boldsymbol{W}$, and $\boldsymbol{Y}$ be the vector $(\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_N)$, then the posterior distribution of $\boldsymbol{b}$ and $\boldsymbol{W}$ is:

$$K(\boldsymbol{b}, \boldsymbol{W}|\boldsymbol{Y}) \propto \prod_{n=1}^{N} L(\boldsymbol{y}_n|\boldsymbol{b}, \boldsymbol{W})k(\boldsymbol{b}, \boldsymbol{W}). \tag{9}$$

It is customary to assume that $\boldsymbol{b}$ and $\boldsymbol{W}$ are independent, the prior on $\boldsymbol{b}$ is normal with an unboundedly large variance, and the prior on $\boldsymbol{W}$ is inverted Wishart with κ degrees of freedom and scale matrix $\boldsymbol{I}$, which is an identity matrix with rank κ. To improve computational efficiency, the hierarchical Bayes approach treats each $\boldsymbol{b}_n$ as a parameter (that is, a random variable) along with $\boldsymbol{b}$ and $\boldsymbol{W}$, and the corresponding posterior for $\boldsymbol{b}$, $\boldsymbol{W}$, and $\boldsymbol{\beta}_n \forall n$ is then given by:

$$K(\boldsymbol{b}, \boldsymbol{W}, \boldsymbol{\beta}_n \forall n | \boldsymbol{Y}) \propto \prod_{n=1}^{N} L(\boldsymbol{y}_n | \boldsymbol{\beta}_n) \phi(\boldsymbol{\beta}_n | \boldsymbol{b}, \boldsymbol{W}) k(\boldsymbol{b}, \boldsymbol{W}). \tag{10}$$

In the hierarchical Bayes approach, drawing from $K(\boldsymbol{b}, \boldsymbol{W}, \boldsymbol{\beta}_n \forall n | \boldsymbol{Y})$ can be efficiently achieved by the Gibbs sampler, which is detailed in Train (2003). Essentially, the method starts with any initial values $\boldsymbol{b}^0$, $\boldsymbol{W}^0$ and $\boldsymbol{\beta}_n^0 \ \forall n$. The r-th iteration ($r \geq 1$) consists of the following steps:

1. Draw $\boldsymbol{b}^r$ from the normal distribution $N(\overline{\boldsymbol{\beta}}^{r-1}, \boldsymbol{W}^{r-1})$, where $\overline{\boldsymbol{\beta}}^{r-1} = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{\beta}_n^{r-1}$;
2. Draw $\boldsymbol{W}^r$ from the inverted Wishart distribution $IW(\kappa + N, (\kappa \boldsymbol{I} + N\boldsymbol{S}^{r-1})/(\kappa + N))$, where $\boldsymbol{S}^{r-1} = \frac{1}{N}\sum_{n=1}^{N} (\boldsymbol{\beta}_n^{r-1} - \boldsymbol{b}^r)(\boldsymbol{\beta}_n^{r-1} - \boldsymbol{b}^r)'$;
3. For each $n \in \{1, 2, \cdots, N\}$, draw $\boldsymbol{\beta}_n^r$ from its posterior conditional on $\boldsymbol{y}_n$, $\boldsymbol{b}^{r-1}$, and $\boldsymbol{W}^{r-1}$, that is, from $K(\boldsymbol{\beta}_n | \boldsymbol{y}_n, \boldsymbol{b}^{r-1}, \boldsymbol{W}^{r-1}) \propto L(\boldsymbol{y}_n | \boldsymbol{\beta}_n) \ \phi \ (\boldsymbol{\beta}_n | \boldsymbol{b}^{r-1}, \boldsymbol{W}^{r-1})$, which is achieved by applying the Metropolis-Hasting algorithm.

The above steps are repeated for many iterations and the resulting values converge to draws from the joint posterior of $\boldsymbol{b}$, $\boldsymbol{W}$ and $\boldsymbol{\beta}_n \forall n$. Consequently, the mean and standard deviation of the converged draws can be calculated to obtain estimates and standard errors of the parameters.

Willingness to pay analysis

One of the key outputs of a discrete choice model is the WTP measurement for various attributes of the alternatives. An advantage of the mixed logit model is that it enables us to estimate each individual's WTP for these attributes. Once we obtain the individual specific WTP estimates, we can conduct regression analysis to infer the socioeconomic sources that cause such differences across individuals. This type of two-step analysis, that is, a mixed logit model followed by a regression analysis, was first proposed by Campbell (2007) to estimate the economic benefits associated with rural landscape improvement. Recently, Hoshino (2011) applied this method to analyse preference heterogeneity in a residential choice problem.

Recall that when estimating the mixed logit model through the hierarchical Bayes method, each individual's $\boldsymbol{\beta}_n$ is treated as a random variable along with $\boldsymbol{b}$ and $\boldsymbol{W}$. Based on the joint posterior for $\boldsymbol{b}$, $\boldsymbol{W}$, and $\boldsymbol{\beta}_n \forall n$ defined in *equation*(10), we can obtain the posterior distribution of $\boldsymbol{\beta}_n$ by integrating out variables $\boldsymbol{b}$, $\boldsymbol{W}$, $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, $\cdots$, $\boldsymbol{\beta}_{n-1}$, $\boldsymbol{\beta}_{n+1}$, $\cdots$, $\boldsymbol{\beta}_N$ in the joint posterior:

$$h(\boldsymbol{\beta}_n | \boldsymbol{Y}) = \int\int\cdots\int K(\boldsymbol{b}, \boldsymbol{W}, \boldsymbol{\beta}_n \forall n | \boldsymbol{Y}) \ d\boldsymbol{b} \ d\boldsymbol{W} \ d\boldsymbol{\beta}_1 \ d\boldsymbol{\beta}_2 \cdots \ d\boldsymbol{\beta}_{n-1} \ d\boldsymbol{\beta}_{n+1} \cdots \ d\boldsymbol{\beta}_N. \tag{11}$$

As a result, the mean $\boldsymbol{\beta}_n$ for individual n in the sampled population can be written as:

$$\overline{\boldsymbol{\beta}}_n = E[\boldsymbol{\beta}_n | \boldsymbol{Y}] = \int \boldsymbol{\beta}_n h(\boldsymbol{\beta}_n | \boldsymbol{Y}) \ d\boldsymbol{\beta}_n = \int\int\cdots\int \boldsymbol{\beta}_n K(\boldsymbol{b}, \boldsymbol{W}, \boldsymbol{\beta}_n \forall n | \boldsymbol{Y}) \ d\boldsymbol{b} \ d\boldsymbol{W} \ d\boldsymbol{\beta}_1 \ d\boldsymbol{\beta}_2 \cdots \ d\boldsymbol{\beta}_N \tag{12}$$

The last equality in *equation*(12) is obtained by plugging in the result for $h(\boldsymbol{\beta}_n | \boldsymbol{Y})$ from *equation*(11). The above multi-dimensional integral does not have a closed form and is calculated by simulation. Note that in the hierarchical Bayes approach, we have already obtained the draws for $\boldsymbol{b}$, $\boldsymbol{W}$, and $\boldsymbol{\beta}_n \forall n$ from their joint posterior defined in *equation*(10). Let R be the number of draws after the burn-in period, and $\boldsymbol{\beta}_n^r$ be the sampled value in the r-th draw, we have:

$$\overline{\boldsymbol{\beta}}_n = \frac{1}{R}\sum_{r=1}^{R}\boldsymbol{\beta}_n^r. \tag{13}$$

Let $\overline{\beta}_{ns}$ be the coefficient for attribute s in vector $\overline{\boldsymbol{\beta}}_n$, then home buyer n's WTP for attribute s is given by:

$$WTP_{ns} = -\frac{\bar{\beta}_{ns}}{\beta_p}, \tag{14}$$

where β_p is the price coefficient. To explain the variability in WTP_{ns}, we assume that individual n's WTP for attribute s can be expressed as:

$$WTP_{ns} = \alpha_s + \boldsymbol{\gamma}_s'\mathbf{z}_{ns} + \omega_{ns}, \tag{15}$$

where α_s is the intercept term, which is constant for all individuals; $\mathbf{z}_{ns}$ is a vector of individual n's socioeconomic characteristics related to attribute s, and $\boldsymbol{\gamma}_s$ is the corresponding coefficient vector; and ω_{ns} is the unobserved random term which models the effect of omitted variables. The estimates for the coefficient vector $\boldsymbol{\gamma}_s$ can then help us uncover the determinants for the preference heterogeneity toward attribute s.

Data description

The data used in this research come from a condominium complex in the city of Beijing. The complex comprises ten residential buildings and covers a total area of around 46,200 m^2, as illustrated in Figure 1. The complex is surrounded by a river on the east, a minor street on the south, and two major streets on the west and the north, respectively. Buildings are identified by the numbers in the figure. Buildings 1, 4 and 8 are 34-storey buildings; Buildings 2, 3, 5, 7, 9 and 10 are 18-storey buildings; and Building 6 is a 16-storey building. Given a residential building, the number of dwelling units on each floor can also be identified. For example, there are eight dwelling units on each floor of Building 6, each of which corresponds to a dark gray shape in the figure. Altogether there are 1092 dwelling units in the condominium complex. Our data contain 818 units sold in the second quarter of 2012. The sales data clearly record the date and time each unit is sold and the detailed information of each dwelling unit, including its location, floor level, floor plan, size, price and so on. Note that if there was a change to the price of a dwelling unit, it is also documented in our data. In addition, for each dwelling unit sold, we know the corresponding home buyer's age, gender and other socioeconomic characteristics.

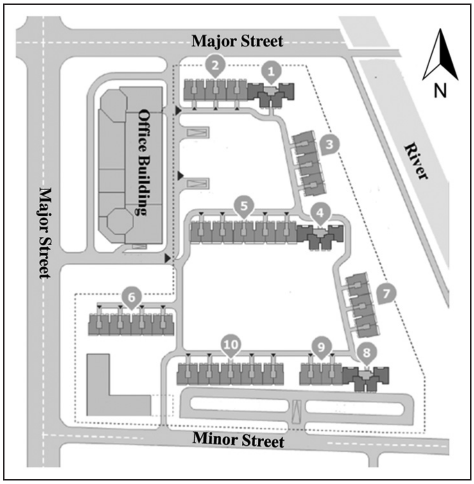


Figure 1. The layout of the condominium complex.

Based on discussions with experts in the real estate sector and potential home buyers, we came up with an initial set of variables to use in our models. Through exploration and testing of the multinomial logit model, the final set of variables are determined and summarised in Table 1. There are eight

Table 1. Description of variables used in the models.

Variable	Description
Location of the dwelling unit in the complex	
BLD_CTR	Binary: 1 = the dwelling unit is in a building in the centre of the complex
BLD_MJR (base)	Binary: 1 = the dwelling unit is in a building adjacent to major streets
BLD_MNR	Binary: 1 = the dwelling unit is in a building adjacent to minor streets
Floor level of the dwelling unit in a building	
FLR_1	Binary: 1 = the dwelling unit is on the first floor
FLR_2	Binary: 1 = the dwelling unit is on the second floor
FLR_3	Binary: 1 = the dwelling unit is on the third floor
FLR_4_5	Binary: 1 = the dwelling unit is on the fourth or the fifth floors
FLR_6_10 (base)	Binary: 1 = the dwelling unit is located between the sixth and the tenth floors
FLR_11_SH	Binary: 1 = the dwelling unit is located between the eleventh and the second highest floor
FLR_H	Binary: 1 = the dwelling unit is located on the highest floor
Location of the dwelling unit on a floor	
UL_END	Binary: 1 = the dwelling unit is located at the end of the floor
Orientation of the dwelling unit	
ORNT_SN (base)	Binary: 1 = the orientation of the dwelling unit is south- and north-facing
ORNT_S	Binary: 1 = the orientation of the dwelling unit is south-facing
ORNT_EW	Binary: 1 = the orientation of the dwelling unit is east- and west-facing
Bathroom	
BATH_MW	Binary: 1 = both bathrooms in the dwelling unit are *MingWei* bathrooms
Bay window	
BAYWND	The sum of the widths of bay windows in the dwelling unit, measured in metres
Size and price	
SIZE	The total size of the dwelling unit, measured in square metres
PRICE	The purchase price of the dwelling unit, measured in 10,000 CNY
Socioeconomic characteristics of the home buyer	
AGE	The home buyer's age
GNDR	Binary: 1 = the individual is male
INC	The home buyer's annual household income in 10,000 CNY
FTHB	Binary: 1 = the individual is a first-time home buyer

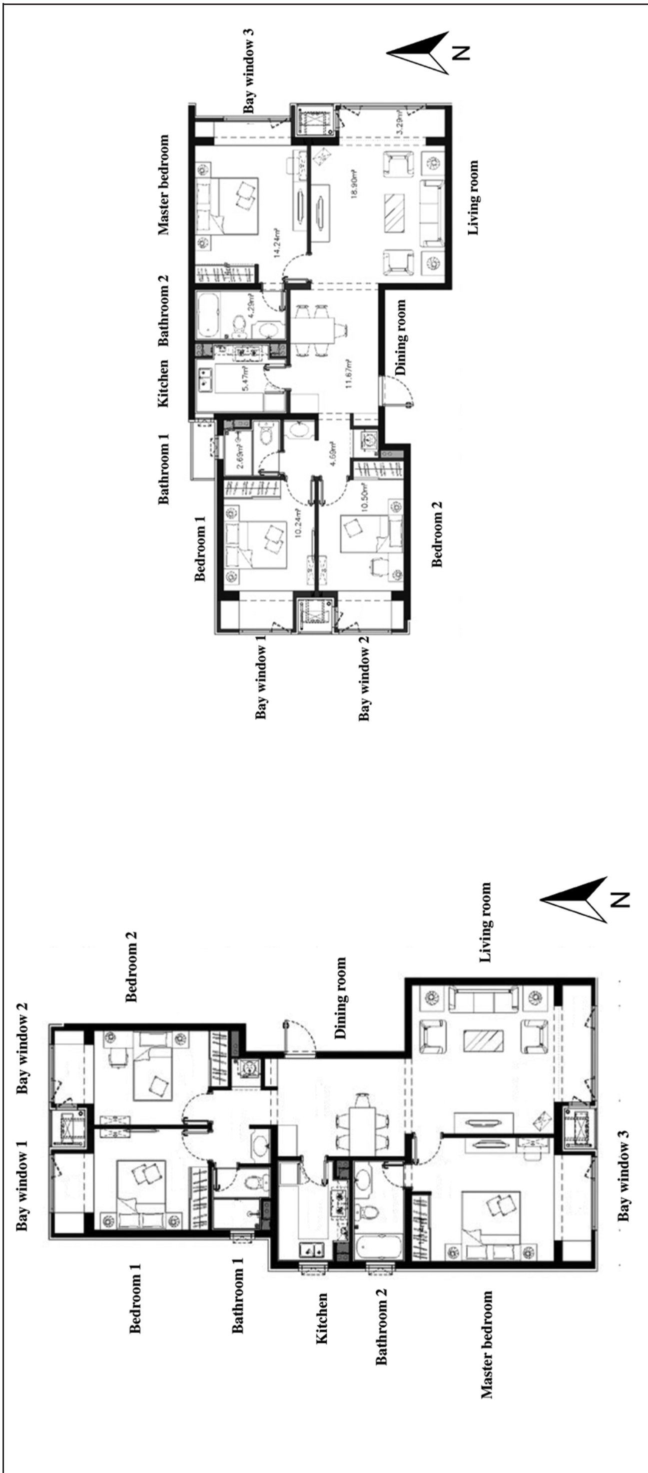
groups of explanatory variables, where the first seven groups describe a dwelling unit and the last group is related to the socioeconomic characteristics of the home buyers.

The first group of variables deals with the location of the dwelling unit in the residential complex. Dwelling units adjacent to city streets may suffer from excessive noise compared with those located in the centre of the complex. To capture home buyers' preferences, we classify the locations of the dwelling units into three categories: in the centre of the complex (Buildings 3, 4, 5 and 7), adjacent to major streets (Buildings 1, 2 and 6), or adjacent to minor streets (Buildings 8, 9 and 10). In this table we also indicate the variables used as the base during model estimation.

The second group of variables quantifies home buyers' preferences toward floor levels in a building. In general, home buyers in China dislike living on the first floor or lower floors because of safety concerns, privacy issues and lack of daylight. For example, residents on the first floor need to install burglar-proof windows and lower their curtains most of the day to prevent passers-by from seeing further into the rooms. In addition, residents on lower floors often need to worry about the overflow of waste water in their bathrooms when the sewerage system for the building is choked. Dwelling units on mid or higher floor levels are much preferred by Chinese home buyers, because they can get more daylight and these floors offer nicer views out of the windows. However, when we go to dwelling units on the highest floor, there does not seem to be a consensus. On one hand, dwelling units on the top floor provide homeowners with a best-in-the-building view and they are in general considered as the quietest place because one is spared the sounds of children and pets running around upstairs, or early risers stumbling from their beds at 5 o'clock in the morning. On the other hand, these dwelling units can get very hot in the summer and the electric bill may be high. Moreover, roof wear and tear can also create leak problems, which can become a big headache. In this research, several different specifications for the floor level variable are examined and compared (for example, linear, piecewise linear, log linear and step functions), and the step function specified in this table provides the best fit.

The third group of variables measures home buyers' preferences toward dwelling units on the same floor. Units at the end of each floor are in general not preferred because a higher percentage of the exterior walls of these units is exposed to outside environment, often causing higher energy consumption. For example, units at the east end of the floor get full morning sun, while units at the west end get full afternoon sun. Binary variable UL_END is introduced to indicate whether a unit is at the end of its floor. Take Building 6 in Figure 1 as an example. There are two dwelling units at the end of each floor: one on the east end and the other on the west end.

The fourth group of variables describes the floor plan of a dwelling unit. An important attribute related to a floor plan is its orientation. Good orientation improves the energy efficiency of a home, making it more comfortable to live in and cheaper to run. The orientation of a dwelling unit is determined primarily by the facing of its exterior windows. There are three common types of orientations for dwelling units in China: south- and north-facing, east- and west-facing, and south-facing. When at least one exterior window faces the south and the north, respectively, this dwelling unit is classified as south- and north-facing, which is indicated by variable ORNT_SN. The orientation of the dwelling unit shown in Figure 2(a) is south- and north-facing. This is the most popular orientation among Chinese home buyers because it has good exposure and the best natural ventilation. In the winter days when the sun is low, the windows on the south side allow natural light to bathe the rooms throughout the day; while in the summer days when the sun is high, only limited sunlight could pass through the southern windows, which helps to keep the rooms cool. When at least one exterior window faces the east and the west, respectively, this dwelling unit is defined as east- and west-facing, which is indicated by variable ORNT_EW. Figure 2(b) shows the floor plan of a dwelling unit whose orientation is east- and west-facing. Although this type of floor plan still can offer proper natural ventilation, the major drawback is that rooms having west-facing windows can get very hot in the summer because of the extended exposure to sunlight in the afternoon. When most of the exterior windows face the south, this unit is classified as south-facing, which is indicated by variable ORNT_S. The major concern with this type of orientation is ventilation. Since most of the exterior windows face the south, there is no passage to allow cool breezes to get through the rooms. All three types of orientation are found in this condominium complex.



(a) The floor plan of a dwelling unit whose orientation is south- and north- facing.

(b) The floor plan of a dwelling unit whose orientation is east- and west- facing.

Figure 2. Illustration of floor plans.

Table 2. Parameter estimates for the multinomial logit and mixed logit models.

		Multinomial logit			Mixed logit	
		Coefficient	SE		Parameter	SE
(1)	BLD_CTR	3.9758‡	0.1663	Mean of coefficient	7.5300‡	0.4295
(2)		–	–	Std. dev. of coefficient	2.5051‡	0.5408
(3)	BLD_MNR	1.0430‡	0.0981	Mean of coefficient	1.4297‡	0.2049
(4)		–	–	Std. dev. of coefficient	3.0270‡	0.4768
(5)	FLR_1	−40.1278†	18.2524	Mean of coefficient	−48.1156‡	1.5052
(6)		–	–	Std. dev. of coefficient	9.0330‡	1.9080
(7)	FLR_2	−27.1530†	12.1833	Mean of coefficient	−30.2771‡	0.8542
(8)		–	–	Std. dev. of coefficient	6.6485‡	1.2175
(9)	FLR_3	−19.5892†	9.1432	Mean of coefficient	−21.3956‡	0.9489
(10)		–	–	Std. dev. of coefficient	6.5600‡	1.3421
(11)	FLR_4_5	−12.9468†	6.113	Mean of coefficient	−13.9328‡	0.6736
(12)		–	–	Std. dev. of coefficient	6.4627‡	1.4136
(13)	FLR_11_SH	6.8563†	3.0259	Mean of coefficient	8.5487‡	0.5221
(14)		–	–	Std. dev. of coefficient	3.2611‡	0.8027
(15)	FLR_H	−0.8630‡	0.179	Mean of coefficient	−8.2632‡	1.6121
(16)		–	–	Std. dev. of coefficient	11.0139‡	2.0317
(17)	UL_END	−0.9451‡	0.1551	Mean of coefficient	−4.1848‡	0.4597
(18)		–	–	Std. dev. of coefficient	4.1489‡	0.5730
(19)	ORNT_S	−12.3321†	5.7281	Mean of coefficient	−13.1265‡	0.6991
(20)		–	–	Std. dev. of coefficient	3.2699‡	0.4902
(21)	ORNT_EW	−27.4623†	12.1847	Mean of coefficient	−28.8199‡	0.3052
(22)		–	–	Std. dev. of coefficient	2.9164‡	0.4408
(23)	BATH_MW	1.1013‡	0.1381	Mean of coefficient	0.9504‡	0.2905
(24)		–	–	Std. dev. of coefficient	2.6811‡	0.5733
(25)	BAYWND	0.6038‡	0.0429	Mean of coefficient	1.1982‡	0.0880
(26)		–	–	Std. dev. of coefficient	1.0759‡	0.1067
(27)	SIZE	10.1032†	4.9639	Mean of coefficient	10.1944‡	0.0302
(28)		–	–	Std. dev. of coefficient	0.2036‡	0.0195
(29)	PRICE	−3.0926†	1.5294	Coefficient	−3.0679‡	0.0165
	No. of observations	818			818	
	Loglikelihood null	−5284			−5284	
	Loglikelihood at convergence	−4329			−4184	
	ρ^2	0.181			0.208	
	$\bar{\rho}^2$	0.178			0.203	

Notes: *: p-value <0.10, †: p-value <0.05, ‡: p-value <0.01.

The next group of variables describes an important characteristic of bathrooms in the dwelling unit. A dwelling unit in China typically has a guest bathroom accessible from the living room (or the dining room) and an ensuite bathroom attached to the master bedroom. Depending on whether it has an exterior window or not, a bathroom is classified into two types, where *MingWei* bathrooms are those with exterior windows and *AnWei* bathrooms are those with no exterior windows. For the dwelling unit shown in Figure 2(b), Bathroom 1 is a *MingWei* bathroom while Bathroom 2 is an *AnWei* bathroom. Since *MingWei* bathrooms offer better ventilation and more daylight, they are very appealing to home buyers in China. Unfortunately, because dwelling units on the same floor are arranged side by side, real estate developers often times cannot accommodate two *MingWei* bathrooms in a dwelling unit, instead they make one of them an *AnWei* bathroom. It is very rare to have two *AnWei* bathrooms in a dwelling unit. We introduce dummy variable BATH_MW which takes value 1 if both bathrooms are *MingWei* bathrooms; and 0 otherwise. The two bathrooms in the dwelling unit shown in Figure 2(a) are both *MingWei* bathrooms.

Bay windows project outward from the exterior walls of the dwelling unit. They offer panoramic views, bring in more natural light, provide good ventilation and give a feeling of spaciousness to a room. All of these factors make bay windows highly attractive to home buyers in China. In Figure 2(b), there are three rooms with bay windows. We introduce variable BAYWND to measure the sum of the widths of all bay windows in a dwelling unit. Presumably, the larger the value of BAYWND, the more preferred the dwelling unit when all else are equal.

The seventh group of variables includes the size of the dwelling unit measured in square metres and the purchase price for the unit measured in 10,000 Chinese Yuan. For the reader's reference, the exchange rate between US Dollar (USD) and Chinese Yuan (CNY) is approximately 1 USD = 6.3 CNY in the second quarter of 2012.

The last group of variables corresponds to the socioeconomic characteristics of the home buyer: age, gender, annual household income and whether he or she is a first-time home buyer.

Empirical results

The estimation of the multinomial logit model is straightforward. For the mixed logit model, all parameters are assumed to be random except the parameter for variable PRICE. This treatment is recommended by Train (2003) and Hoshino (2011) to avoid numerical difficulty when calculating the WTP measures. While estimating the mixed logit model, we perform 50,000 iterations in the MCMC process, of which the first 25,000 iterations are for the burn-in period. Afterwards, we retain every fifth draw to obtain a total of 5000 draws from the posterior density function.

Parameter estimates

The parameter estimates for both the multinomial logit and mixed logit models are reported in Table 2. We can find that all positive/negative signs of the estimated coefficients in the multinomial logit model are consistent with expectation. And all of them are significantly different from zero. The parameter estimates show that home buyers prefer residential buildings located in the centre of the complex than those located next to streets. Speaking of floor levels, dwelling units on the first floor are least preferred. Homeowners favour dwelling units on higher floor levels and this trend continues until we reach the second highest floor. Units located at the end of each floor are not as good as those located in the middle of each floor, which is suggested by the negative sign for UL_END. Homeowners pay a lot of attention to orientation, they prefer units that are south- and north-facing, followed by those that are south-facing, and finally those that are east- and west-facing. Homeowners also prefer units having two *MingWei* bathrooms and bigger bay windows.

Similar results are observed for parameters in the mixed logit model. In addition, the estimated standard deviations of coefficients in the mixed logit model are highly significant, indicating the presence of substantial preference heterogeneity in dwelling unit choice behaviour. The value of the log-likelihood is much greater in the mixed logit model than that in the multinomial logit model. The log-likelihood ratio test statistic is 290, which is greater than the critical value of the chi-square distribution for 14 degrees of freedom, that is, 29.141 at the 0.01 level of significance.

When we compare the parameter estimates in the multinomial logit model to the estimated mean values of the parameters in the mixed logit model on rows (1), (3), (5), $\cdots$, and (29), we notice that the estimated mean values of the parameters in the mixed logit model are generally larger in magnitude than those in the multinomial logit model. This phenomenon is expected because the scale of utility is determined by the normalisation of the i.i.d. term e_{in}. In the multinomial logit model, all stochastic terms not accounted for by the deterministic part is absorbed by this single error term. However, in the mixed logit model, a portion of the variance in the random component of utility is captured in ζ_n. Hence, the variance of e_{in} in the multinomial logit model is greater than that in the mixed logit model, that is, the scale of e_{in} in the multinomial logit model is smaller than that in the mixed logit model. Since we normalise the scale of e_{in} to one in both models, the utility (and hence the parameters) of the multinomial logit model are scaled down relative to the mixed logit model.

The mixed logit model also allows us to calculate the probability when the coefficient for an attribute is positive (or negative) because the density function for the coefficient vector $\boldsymbol{\beta}_n$, $\phi(\boldsymbol{b}, \boldsymbol{W})$, is now known. For example, the probability when the coefficient for BLD_CTR is positive can be calculated as $\Pr(\beta_{BLD\_CTR}>0) = 0.9987$, suggesting almost all home buyers prefer to live in the centre of the complex than near major streets. We are particularly interested in the coefficient for FLR_H because it is widely accepted by real estate developers in China that home buyers do not like dwelling units on top floors and discounts should be offered (Liao, 2011). The mean of β_{FLR_H} is indeed negative, however, the probability when $\beta_{FLR\_H}>0$ is $\Pr(\beta_{FLR\_H}>0) = 0.2265$, indicating that there are about one-quarter of home buyers who prefer to live on the highest floor than on the sixth to the tenth floors, which is the base level for floor variables. To get a better understanding on home buyers' preference toward floor levels, we can also calculate the following probabilities: $\Pr(\beta_{FLR\_1}<0) = 1.0000$, $\Pr(\beta_{FLR\_2}<0) = 1.0000$, $\Pr(\beta_{FLR\_3}<0) = 0.9994$, $\Pr(\beta_{FLR\_4\_5}<0) = 0.9845$, and $\Pr(\beta_{FLR\_11\_SH}>0) = 0.9956$. Since these probabilities are all very close to 1.0000, we can draw the following conclusion:

FINDING 1 *Although home buyers have unanimous agreement on the preference for floor levels below the top floor, there are significantly different opinions on living on the top floor. Unlike the widely accepted belief that home buyers dislike dwelling units on the top floor, we find that a fair portion of home buyers, that is, about one-quarter of them, prefer to stay on the top floor.*

Substitution patterns

The multinomial logit model assumes that alternatives are independent of each other, which leads to equal cross elasticity among alternatives. That is, when a new alternative is introduced to a set of existing alternatives, the new alternative will draw proportionately from all existing ones. The mixed logit model, however, allows us to model complex substitution patterns (or correlations) among alternatives. In the mixed logit model, following our previous notation, coefficient β_n is decomposed into mean b and deviation ζ_n. We can re-write the utility function as:

Table 3. Examples that illustrate the complex substitution patterns among dwelling units.

Case	Attribute	Base		New
		Unit 1	Unit 2	Unit 3
1	BLD_CTR	1	0	1
	BLD_MNR	0	1	0
2	FLR_1	0	1	1
	FLR_2	1	0	0
3	BLD_CTR	1	0	1
	BLD_MNR	0	1	0
	FLR_1	0	1	1
	FLR_2	1	0	0

$$\begin{aligned} U_{in} &= (\boldsymbol{b}' + \boldsymbol{\zeta}'_n)\boldsymbol{x}_{in} + \in_{in} \\ &= \boldsymbol{b}' x_{in} + (\boldsymbol{\zeta}'_n \boldsymbol{x}_{in} + \in_{in}), \end{aligned} \tag{16}$$

where $\boldsymbol{b}'\boldsymbol{x}_{in}$ is the deterministic part of the total utility and $\boldsymbol{\zeta}'_n\boldsymbol{x}_{in} + \in_{in}$ is the random part of the total utility. $\boldsymbol{\zeta}'_n\boldsymbol{x}_{in} + \in_{in}$ introduces correlation among alternatives:

$$\begin{aligned} \mathrm{Cov}(U_{in}, U_{jn}) &= \mathrm{Cov}(\boldsymbol{\zeta}'_n\boldsymbol{x}_{in} \\ &\quad + \in_{in}, \boldsymbol{\zeta}'_n\boldsymbol{x}_{jn} + \in_{jn}) \\ &= \boldsymbol{x}'_{in}\boldsymbol{W}\boldsymbol{x}_{jn}. \end{aligned} \tag{17}$$

This interpretation of the mixed logit model allows us to examine the correlation or substitution patterns among the alternatives. In Table 2, rows (1), (3), $\cdots$, (29) are the mean values of the parameters, that is, vector $\boldsymbol{b}$ and rows (2), (4), $\cdots$, (28) are the estimated variances associated with $\boldsymbol{\zeta}_n$.

Let x_{in}^k be the k-th element in $\boldsymbol{x}_{in}$ and the corresponding coefficient is $\beta_n^k = b^k + \zeta_n^k$. If x_{in}^k is a dummy variable, it suggests that alternatives having the same value for x_{in}^k are correlated and they are substitutable for each other to a certain extent. This can be illustrated through the three examples shown in Table 3. In Case 1, a dwelling unit located in the centre of the complex (Unit 3) is introduced to a base situation consisting of Unit 1 located in the centre of the complex and Unit 2 located near the minor street, and all the other attributes are the same across the three units. We know that the substitution effect between Units 3 and 1 is stronger than that between Units 3 and 2, because Cov $(U_{unit\,1}, U_{unit\,3})$ − Cov $(U_{unit\,2}, U_{unit\,3}) = \mathrm{var}(\zeta_{BLD\_CTR}) = 2.5051^2 > 0$. Similarly, in Case 2, a dwelling unit on the first floor (Unit 3) is introduced to a base situation consisting of dwelling units on the second and first floors, respectively (Units 1 and 2), and all the other attributes are the same across the three units. We know that the substitution effect between Unit 3 and Unit 2 is stronger than that between Unit 3 and Unit 1. This is because Cov $(U_{unit\,1}, U_{unit\,3})$–Cov $(U_{unit\,2}, U_{unit\,3}) = -\mathrm{var}$ $\zeta_{FLR\_1}) = -9.0330^2 < 0$.

Case 3 shows a more complex situation: Unit 3 on the first floor of a residential building located in the centre of the complex is introduced to a base scenario consisting of Unit 1 on the second floor of a residential building located in the centre of the complex and Unit 2 on the first floor of a residential building located near the minor street, and all the other attributes are the same across the three units. We know that the substitution effect between Unit 3 and Unit 2 is stronger than that between Unit 3 and Unit 1, because Cov $(U_{unit\,1}, U_{unit\,3})$ − Cov $(U_{unit\,2}, U_{unit\,3}) = \mathrm{var}(\zeta_{BLD\_CTR}) - \mathrm{var}(\zeta_{FLR_1}) = 2.5051^2 - 9.0330^2 < 0$.

Since the estimated variances associated with $\boldsymbol{\zeta}_n$ in Table 2 are all significant and they are roughly in the same order of magnitude,

Table 4. WTP estimates under the mixed logit model. Absolute WTP estimates are reported in CNY, while relative WTP estimates are measured as percentages of the average unit price.

	Attribute	Mean of WTP		Standard deviation of WTP	
		Absolute value (CNY)	Relative pct	Absolute value (CNY)	Relative pct
(1)	BLD_CTR	24,546	0.555%	4388	0.099%
(2)	BLD_MNR	4669	0.106%	6598	0.149%
(3)	FLR_1	−156,864	−3.545%	21,301	0.481%
(4)	FLR_2	−98,707	−2.231%	15,768	0.356%
(5)	FLR_3	−69,751	−1.576%	15,763	0.356%
(6)	FLR_4_5	−45,422	−1.026%	15,624	0.353%
(7)	FLR_11_SH	27,866	0.630%	7311	0.165%
(8)	FLR_H	−26,968	−0.609%	26,563	0.600%
(9)	UL_END	−13,644	−0.308%	9134	0.206%
(10)	ORNT_S	−4279	−0.967%	11,656	0.263%
(11)	ORNT_EW	−93,941	−2.123%	12,383	0.280%
(12)	BATH_MW	3093	0.070%	5216	0.118%
(13)	BAYWND	3908	0.088%	2276	0.051%

it clearly indicates that complex correlation exists among dwelling units along multiple dimensions, including location in the complex, floor levels, location on the floor, orientation and so on. Hence, the mixed logit model is more appropriate for the modelling of the dwelling unit choice problem.

Willingness to pay analysis

The WTP estimates based on the mixed logit model are summarised in Table 4. Column 1 shows the names of the attributes. Column 2 shows the absolute WTP estimates measured in CNY. Column 3 shows the relative WTP estimates, which is expressed as percentages of the average unit price. That is, Column 3 is obtained by dividing Column 2 by the average unit price in this residential complex. Columns 4 and 5 show the standard deviations of WTP estimates in both absolute and relative terms. The WTP for FLR_1 is negative and its absolute value is the largest among all attributes, indicating that home buyers have strong objections to living on the first floor. They have to be compensated by 156,864 CNY if they are asked to move from a unit on the sixth floor (the base level for floor variables) to the first floor, which is about 3.545% of the average price of the dwelling units. Besides floor levels, home buyers place heavy weight on the orientation of a dwelling unit. For example, compared with a dwelling unit whose orientation is south- and north-facing, the price of a dwelling unit whose orientation is east- and west-facing needs to be about 93,941 CNY (or, 2.123% of the average unit price) lower. Since preferable attributes have greater WTP estimates, by examining the mean values of the WTP estimates (that is, Columns 2 and 3 in Table 4), we can characterise what an ideal dwelling unit looks like.

FINDING 2 *In Chinese home buyers' minds, an ideal dwelling unit meets the following criteria: (1) located on or above the eleventh floor but below the highest floor; (2) south- and north-facing; (3) away from streets; (4) not at either end of the floor level; and (5) both bathrooms are MingWei.*

More importantly, the WTP estimates reported in Table 4 can be used as guidelines

Table 5. Ranges of the WTP measures along key dimensions of a dwelling unit (in CNY).

Attribute group	Mean WTP		
	Min	Max	Max-Min
Location of the dwelling unit in the complex	0	24,546	24,546
Floor level of the dwelling unit in a building	−156,864	27,866	184,730
Location of the dwelling unit on a floor	−13,644	0	13,644
Orientation of the dwelling unit	−93,941	0	93,941
Bathroom	0	3093	3093

to price dwelling units in a condominium complex:

FINDING 3 *The relative percentages of the WTP estimates, that is, Column 3 reported in Table 4, can be used to assist real estate developers in adjusting the prices of dwelling units according to the differences in their attributes, for example, floor levels, orientation and so on.*

We use two examples to illustrate how Finding 3 can be applied. Suppose we have a dwelling unit located on the first floor of a building, then it should be priced 3.545% lower than a dwelling unit located on the sixth floor of the same building when all else are held equal. Another example could be that, all else held equal, a dwelling unit on the first floor of a building located in the centre of the complex should be priced 3.545%–0.555% = 2.99% lower than a dwelling unit on the sixth floor of a building located next to major streets.

In Table 5, we summarise the ranges of the mean WTP estimates according to the attribute groups defined in Table 1. Column 1 shows the names of the attribute groups. For each attribute group, we show the minimum and maximum mean WTP estimates in Columns 2 and 3, and calculate their differences in Column 4. For example, the first attribute group describes the location of the dwelling unit in the complex and the variables are BLD_CTR, BLD_MNR, and BLD_MJR (the base). Attribute BLD_MJR has the minimum mean WTP estimate, which is 0 CNY; while attribute BLD_CTR has the maximum mean WTP estimate, which is 24,546 CNY. The difference between the maximum and minimum mean WTP estimates is 24,526 CNY. Since attribute groups having large values in Column 4 have greater influence on home buyers, we can immediately obtain the following result.

FINDING 4 *The top two factors that affect home buyers' decisions are the floor level of the dwelling unit and its orientation, followed by its location in the complex, its location on a floor level, and finally, whether both of its bathrooms are MingWei.*

The above observation is consistent with the findings reported by Heinzle et al. (2013) where various characteristics of the condominium unit and the condominium complex are analysed to quantify real estate investors' preferences toward the Green Building Certification. Their results show that among attributes associated with a dwelling unit, the main aspect of living room (a concept similar to the orientation of a dwelling unit) and the floor level are the two most important ones.

Explaining the behavioural heterogeneity

An advantage of the mixed logit model is that it offers considerable insights into the taste variation among home buyers. The last two columns in Table 4 show the standard deviations of the corresponding WTP estimates in both absolute and relative terms under the mixed logit model. A large standard deviation indicates that there exists significant taste variation for this attribute. Row (1) shows the standard deviation of the WTP estimate for dwelling units located in the centre of the complex, and its value is relatively small (4388 CNY for BLD_CTR), indicating that there is a unanimous preference among home buyers to live in the centre of the complex, so they can get a quieter environment at home. Now let us look at the values in Rows (3) through (8) in Table 4, which quantify the taste variation toward floor level variables. Although in section 'Willingness to pay analysis' we show that home buyers prefer units on higher floors except the highest floor, we can now see that this preference is accompanied by significant behavioural heterogeneity, which is reflected by the relatively large standard deviations. We also notice that the standard deviations for variable FLR_H is the largest (which is 26,563 CNY, or 0.600% of the average unit price) among all variables, indicating that there are significantly different opinions toward living on the highest floor.

Real estate developers are interested in the socioeconomic sources that cause the heterogeneity in preference, so that they can better match home buyers with their preferred units. For example, if real estate developers know the kind of home buyers that prefer dwelling units on the highest floor, they would then specifically target those buyers with units on the highest floors. To accomplish this goal, we follow the steps outlined in section 'Willingness to pay analysis' to relate each individual's WTP estimate to his or her socioeconomic variables. We first estimate each home buyer's individual WTP for the attribute we are interested in conditional upon his or her actual choice. We then fit a regression model in which the dependent variable is a home buyer's WTP for this attribute and the explanatory variables are his or her socioeconomic attributes presented in the last four rows in Table 1. The estimation results are reported in Table 6 and the details are as follows.

- The first group of results are related to the floor variables. We can see that the WTP regression for variable FLR_1 has an r^2 of 0.248, which suggests that about 25% of the heterogeneity in WTP can be explained by the explanatory variables. The coefficient for AGE is negative and significant at the 0.01 level, which signifies that older home buyers dislike dwelling units on the first floor. The coefficient for INC is also negative and significant at the 0.05 level, which indicates that high income home buyers are less willing to buy units on the first floor. The coefficient for FTHB is positive and significant at the 0.05 level, which implies that first-time home buyers are more prone to purchase dwelling units on the first floor. A possible explanation is that since they are novice home buyers, they probably have not fully recognised the disadvantages associated with living on the first floor. Similar results are observed for the other floor variables: Home buyers with older ages and higher income are less likely to purchase dwelling units located on poor floor levels, while first-time home buyers are more willing to stay on those floors. It is worth mentioning that in the regression model for variable FLR_H, the coefficient of GNDR is positive and significant at the 0.05 level, indicating that male home buyers are more interested in living on the top floors.
- The second group of results are related to the orientation of the dwelling unit. In the WTP regression result for attribute ORNT_S, the coefficients of AGE and INCOME are negative and significant at the 0.01 level, which suggests that home buyers with older ages and higher annual

Table 6. Parameter estimates for the WTP regression models

Attribute group	Dependent variable		Intercept	AGE	GNDR	INC	FTHB	r^2
Floor level	FLR_1	Coefficient	−12.6724‡	−0.0668‡	−0.0407	−0.0272*	0.3008*	0.248
		S. E.	0.8773	0.0153	0.0872	0.0132	0.1766	
	FLR_2	Coefficient	−8.0062‡	−0.0627‡	−0.0250	−0.0229‡	0.1906	0.182
		S. E.	0.6516	0.0114	0.0391	0.0201	0.1312	
	FLR_3	Coefficient	−5.2580‡	−0.0635‡	−0.0180	−0.0188*	0.0942	0.152
		S. E.	0.6524	0.0132	0.0392	0.0144	0.1313	
	FLR_4_5	Coefficient	−2.8582‡	−0.0317‡	−0.0268	−0.0094	0.0118	0.069
		S. E.	0.6467	0.0127	0.0380	0.0133	0.1302	
	FLR_11_SH	Coefficient	3.1784‡	−0.0141‡	0.0166	−0.0051	−0.0557	0.030
		S. E.	0.3044	0.0041	0.0150	0.0053	0.0613	
	FLR_H	Coefficient	0.1457	−0.1077‡	0.0263*	−0.0343*	0.8377‡	0.127
		S. E.	1.0982	0.0287	0.0144	0.0191	0.2210	
Orientation	ORNT_S	Coefficient	−3.1717‡	−0.0369‡	−0.0355	−0.0199‡	0.0291	0.153
		S. E.	0.2744	0.0032	0.0586	0.0048	0.0552	
	ORNT_EW	Coefficient	−8.2902‡	−0.0312‡	0.0513	−0.0141‡	−0.0476	0.116
		S. E.	0.2620	0.0046	0.0559	0.0041	0.0527	
Location in the complex	BLD_CTR	Coefficient	1.7901‡	0.0155‡	−0.0301	0.0124‡	−0.0373	0.043
		S. E.	0.1826	0.0022	0.0390	0.0012	0.0368	
	BLD_MNR	Coefficient	0.2229	0.0197‡	0.0351	0.0119†	−0.0580	0.047
		S. E.	0.2686	0.0055	0.0573	0.0047	0.0541	
Location on a floor	UL_END	Coefficient	−0.8690‡	−0.0149	0.0577	−0.0112*	0.0312	0.011
		S. E.	0.3789	0.0167	0.0808	0.0066	0.0763	
Bathroom	BATH_MW	Coefficient	0.0751	0.0068*	0.0362	−0.0026	0.0046	0.005
		S. E.	0.2169	0.0047	0.0463	0.0038	0.0437	

Notes: ‡ : $p - value < 0.01$, † : $p - value < 0.05$, * : $p - value < 0.10$. The dependent variables, that is, the individual WTP estimates, are measured in 10,000 CNY.

household income levels are reluctant to stay in dwelling units whose orientation is south-facing. Similar results are observed for attribute ORNT_EW.

- The third group of results are related to the location of the dwelling unit in the residential complex. The coefficients for AGE and INC are positive and significant in both regressions, which suggests that home buyers with older ages and higher income prefer dwelling units located far from major streets so as to stay away from urban noise.
- The fourth group of results are related to the location of the dwelling unit on a floor. The coefficient for INC is negative and significant at the 0.05 level, which suggests that high income home buyers do not like dwelling units at either end of the floor. The coefficients for AGE and FTHB are not significant, indicating that they are somehow indifferent to the location of the dwelling unit on a floor.
- The last group of results are related to the type of bathroom in a dwelling unit. The positive coefficient for AGE indicates that older home buyers prefer to have two *MingWei* bathrooms.

Overall, the relationships between the coefficients and the dependent variables are reasonable and we can summarise the above analysis into the following two findings:

FINDING 5 *Home buyers with older ages and higher annual household income have stronger preferences for dwelling units with good quality, that is, dwelling units that satisfy the criteria specified in Finding 2; while first-time home buyers are more flexible to accept dwelling units with one or more less-desirable attributes, for example, those on low floor levels, or those that are not south- and north-facing.*

FINDING 6 *The gender of the home buyer appears to have little impact on the taste heterogeneity in dwelling unit choice; we do, however, find evidence that male home buyers are more interested in dwelling units on the top floors.*

It is worth noting that the explanatory power of the independent variables differs across models. For example, in the regression model for FLR_1, we have $r^2 = 0.248$. However, in the regression models for FLR_11_SH, the r^2 is rather small, which suggests that only a very small portion of the observed preference heterogeneity is explained. Such phenomena are also reported in Walker and Li (2007) and Hoshino (2011). A possible explanation is that the purchase of a dwelling unit is a collective decision process involving not only the home buyer him- or herself, but also members in the home buyer's family as well as members of the extended family, that is, the home buyer's spouse, parents and parents-in-law. Therefore, the behavioural heterogeneity may need to be interpreted by collecting data about the home buyer's family members as well as his or her extended family members.

Conclusions

The main purpose of this paper is to quantify home buyers' preference toward dwelling units in a condominium complex, a problem that has not been thoroughly investigated in the housing research literature. We estimate a multinomial logit model and a mixed logit model using actual sales data from a condominium complex in the city of Beijing. Results from the mixed logit model provide strong evidence that considerable behavioural heterogeneity exists among home buyers. This also suggests that complex substitution patterns exist among dwelling units and the mixed logit model is more appropriate for the dwelling unit choice problem.

Based on the WTP estimates from the mixed logit model, we produce the appropriate price adjustments real estate developers shall make for dwelling units with different attributes, for example, different floor levels, orientation and so on. We characterise what an ideal dwelling unit looks like in Chinese home buyers' minds and quantify the importance of factors related to dwelling unit choice. We find that the top two factors that affect dwelling unit choice are the floor level and the orientation of the unit, followed by the location in the complex, the location on a floor level and finally whether both bathrooms are *MingWei*.

To further investigate the causes for the behavioural heterogeneity among home buyers, we estimate each home buyer's WTP for attributes where significant behaviour heterogeneity is observed. We then conduct regression analysis to explain the differences in WTP across individuals by their socioeconomic characteristics. We find that the relationships between the explanatory variables and the dependent variables are reasonable and that home buyers with older ages, higher annual household income and previous home ownership experience tend to focus more on the quality of the dwelling units, that is, they prefer to stay in units on higher floor levels (but not on the top floor) with good orientation. Nonetheless, we also notice that the explanatory variables we examine in the regression models only account for a limited portion of the preference heterogeneity. This signifies that more factors need to be considered to account for the preference heterogeneity in dwelling unit choice behaviour.

A possible extension to this research is to administer SP surveys to include broader socioeconomic characteristics of the home buyers as well as his or her family and extended families in the analysis. These data can then be combined with the sales data to leverage the strengths from each data source. In this way, we can develop more effective ways to segment home buyers according to their preferences, which will assist real estate developers in designing and pricing dwelling units tailored toward each segment and improve profitability.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Note

1. Note that the word 'condominium' can have different meanings in different countries. In China, a condominium (or a condo) refers to the form of housing tenure and other real property where a specified part of a piece of real estate (usually of an apartment house) is individually owned. It is an economical way to own a home. This definition also applies to the USA and most provinces of Canada. In other countries such as Singapore, however, a condominium is used for housing buildings with luxury features such as security guards, swimming pools and tennis courts.

References

Barrios Garcia JA and Rodriguez Hernandez JE (2007) Housing and urban location decisions in Spain: An econometric analysis with unobserved heterogeneity. *Urban Studies* 44(9): 1657–1676.

Ben-Akiva M and Lerman S (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand*. Volume 9. Cambridge, MA: MIT Press.

Campbell D (2007) Willingness to pay for rural landscape improvements: Combining mixed logit and random-effects models. *Journal of Agricultural Economics* 58(3): 467–483.

Habib KMN and Kockelman KM (2008) Modeling the choice of residential location and home type: Recent movers in Austin, Texas. Paper presented at the *87th Annual Meeting of the Transportation Research Board*, Washington, DC., 2008. Available at: http://www.caee.utexas.edu/prof/kockelman/public_html/TRB08HomeType%26Location.pdf.

Heinzle SL, Boey Ying Yip A and Low Yu Xing M (2013) The influence of green building certification schemes on real estate investor behaviour: Evidence from Singapore. *Urban Studies* 50(10): 1970–1987.

Hoshino T (2011) Estimation and analysis of preference heterogeneity in residential choice behaviour. *Urban Studies* 48(2): 363–382.

Ishikawa T, Nakata S and Asami Y (2011) Perception and conceptualization of house floor plans: An experimental analysis. *Environment and Behavior* 43(2): 233–251.

Liao Z (2011) *A Practical Guide to Real Estate Pricing*. Beijing: China Architecture & Building Press.

Quigley J (1976) Housing demand in the short run: An analysis of polytomous choice. In: NBER *Explorations in Economic Research*. Volume 3(1). Cambridge, MA: NMBER, pp. 76–102.

Skaburskis A (1999) Modelling the choice of tenure and building type. *Urban Studies* 36(13): 2199–2215.

Train KE (2003) *Discrete Choice Methods with Simulation*. Cambridge, UK: Cambridge University Press.

Tu Y and Goldfinch J (1996) A two-stage housing choice forecasting model. *Urban Studies* 33(3): 517–537.

Walker J and Li J (2007) Latent lifestyle preferences and household location decisions. *Journal of Geographical Systems* 9(1): 77–101.

Wu W, Zhang W and Dong G (2013) Determinant of residential location choice in a transitional housing market: Evidence based on micro survey from Beijing. *Habitat International* 39: 16–24.

Yates J and Mackay D (2006) Discrete choice modelling of urban housing markets: A critical review and an application. *Urban Studies* 43(3): 559.