HOUSING SALES IN URBAN BEIJING
Luis Alberiko Gil-Alana, Carlos Pestana Barros

To cite this version:
Luis Alberiko Gil-Alana, Carlos Pestana Barros. HOUSING SALES IN URBAN BEIJING. Applied Economics, Taylor & Francis (Routledge), 2011, pp.1. 10.1080/00036846.2011.591736 . hal-00719480
HOUSING SALES IN URBAN BEIJING
Carlos Pestana Barros¹, Zhongfei Chen² and Luis A. Gil-Alana³

¹. Instituto Superior de Economia e Gestão; Technical University of Lisbon. Rua Miguel Lupi, 20; 1249-078 - Lisbon, Portugal and UECE (Research Unit on Complexity and Economics).
Phone: +351 - 213 016115; fax: +351 - 213 925 912.
E-mail: Cbarros@iseg.utl.pt

2. Wuhan University of Technology, School of Economics, JianLake Campus, Luoshi Road 122, Wuhan, Hubei Province, China.
E-mail: hongyeczf@163.com

3. University of Navarra, Faculty of Economics and Business Administration, Edificio Biblioteca, Entrada Este, E-31080 Pamplona, Spain
Phone: +34 948 425 625 ; fax: +34 948 425 626.
E-mail: alana@unav.es

Abstract: In the housing market, new properties sometimes experience delays before they are sold. Such delays reflect the preferences of buyers in respect of the homes’ characteristics. Therefore, it is important for managerial purposes to identify the causes of housing sales delays. After analyzing the delays in sales of housing in Beijing City, China, the principal finding of this study is that delays are largely explained by the dwellings’ characteristics and location. Policy implications of the research findings, particularly those related to means of reducing the delays, are discussed.

Keywords: Duration models, survival models, Beijing, China.

Correspondence author: Prof. Dr. Luis A. Gil-Alana
University of Navarra
Faculty of Economics
Edificio Biblioteca, Entrada Este
E-31080 Pamplona, Spain
Phone: 34 948 425 625
Fax: 34 948 425 626
Email: alana@unav.es
1. Introduction

Delays in the selling of residential properties is a key issue in house demand management, since it increases the cost of the dwelling due to accruing interest and other costs while the dwelling remains unsold. Therefore, the question of housing sales delays is of importance to realtors because of the uncertainty that it involves (Cunningham, 2006). The research focus of this paper is to examine the determinants of housing sales delays in a city market - Beijing, China - using a sample of homes sold and adopting an innovative survival model. Survival models in housing have been adopted by Gronberg and Reed (1992), Mok, Chan and Cho (1995); Lacour-Little and Malpezzi (2003); Cunningham (2006); Bulan, Mayer and Somerville (2009); Feijten and Mulder (2010); Ström (2010) and Kau, Keenan and Li (2011). However, none of these authors analyse sales delay neither took sample selection into consideration in their survival model.

The present research was motivated by the following issues. First, survival models have proven to be particularly suited to the modeling and analysis of duration events. Although the usefulness and reliability of survival modeling for predicting duration events has been recognized in several urban contexts (Lacour-Little and Malpezzi, 2003; Cunningham, 2006; Bulan, Mayer and Somerville, 2009; Feijten and Mulder, 2010 and Ström, 2010, and Kau, Keenan and Li, 2011), their application to housing has not attracted much research interest. Second, since, as mentioned above, housing sales delays are of great importance in housing management, it would be of value to ascertain which covariates best explain the delays in selling housing, since these covariates may be specific to the market analyzed. Third, it is important for policy purposes to investigate the factors that influence purchasers’ home-buying
decisions. Finally, the consequence of sample selection is taken into account in the
present research.

The paper makes two important contributions to the relevant literature: first,
by analyzing the determinants of home sales delays in the case of the Beijing housing
market, an Asian city that has attracted little research so far; and second, by adopting
a survival model with sample selection, which is a novelty in the survival models
context. We are not aware of any research paper previously adopting survival models
with sample selection in housing.

This paper is organised as follows. After the introduction, the contextual
setting on Beijing house market is presented, followed by the literature survey in
Section 3. Next the theoretical framework background is presented in Section 4,
followed by the hypotheses to be tested in Section 5. The results are given in Section
6, while 7 contains some discussion and concluding comments.

2. The city of Beijing and its house market

Beijing, the capital of China, is one of the four municipalities directly under the
administration of the Central Government and located in the northwest of the country.
Nowadays, it is the hub of China’s political and cultural life. The city covers an area
of 16,410 square kilometers, with a current population of approximately 17.550
million. The nominal GDP was 1,186.6 billion yuan in 2009 (yuan/Dollar in

The real estate market of Beijing has developed in the last decades since the
Central Government launched its nationwide reform program with respect to housing
and the land-use system. Prior to the initiation of the reforms in 1978, there was no
real estate market in China, as trading in land and property was strictly prohibited.
During the decade from 1978 to 1988, the Chinese real estate market was dynamized as the country took its first bold steps from a planned economy towards a market economy. Deng Xiaoping, the chief architect of the reform and opening-up policy from 1978, proposed that the housing commercialization should be based on real estate reform. With the surge of real estate enterprises after the deregulation and stimulus, the real estate industry experienced excessive growth, the number of real estate agencies increasing from 3,140 in 1988 to 17,000 by 1992. To counter the subsequent overheating, the Central Government launched series of measures to intervene in the real estate market. In 1997, when the Asian financial crisis erupted, causing the strong downturn of its exports, China initiated a second wave of real estate reform, i.e. ending subsidized housing (Alona and Bian, 2005), encouraging the development of real estate finance and implementing the public accumulation funds program. These radical measures led to renewed rapid growth during the following decade, the real estate industry contributing 4.75% of China’s GDP in 2007, compared with the 2.20% registered in 1978. However, rising prosperity has brought potential bubbles as well as development. Today’s burgeoning real estate prices have incurred the inhabitants’ unanimous aversion to purchasing new homes and led to the Government’s harsh intervention. This has taken the form of limiting investment to reduce speculation-type demand for real estate, increasing the supply of economically affordable housing, forcing 78 state-owned companies whose core business is not real estate to withdraw from the real estate market, and so on.

Beijing is among those Chinese cities that have experienced soaring housing prices, due to brisk demand derived from rapidly expanding urbanization resulting from rural migration and location advantages with regard to employment. In addition, when the city was chosen in 2001 to host the 29th Olympic Games, the real estate
market of Beijing enjoyed another boom, as the amount of dwellings being constructed and traded was enhanced in successive years (See Table 1). On the other hand, as a political centre, Beijing is quite vulnerable to the Central Government’s policy decisions. The frequent national macroeconomic control occurring in recent times may obscure the short-term development of Beijing’s real estate market in the future.

Table 1: Beijing housing characteristics during 1999-2008 (yuan/Dollar in December 2009 = 6.8282).

<table>
<thead>
<tr>
<th>Years</th>
<th>Price of dwellings (yuan per square meter)</th>
<th>Sold space of dwellings (thousand square meters)</th>
<th>Dwellings unoccupied on sale (thousand square meters)</th>
<th>Dwellings space newly completed (thousand square meters)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1999</td>
<td>4786.8</td>
<td>4847.1</td>
<td>——</td>
<td>9082.6</td>
</tr>
<tr>
<td>2000</td>
<td>4557.2</td>
<td>8982.2</td>
<td>——</td>
<td>10136.6</td>
</tr>
<tr>
<td>2001</td>
<td>4716.0</td>
<td>11275.0</td>
<td>6340.6</td>
<td>13934.3</td>
</tr>
<tr>
<td>2002</td>
<td>4467.0</td>
<td>16044.0</td>
<td>6340.6</td>
<td>19261.7</td>
</tr>
<tr>
<td>2003</td>
<td>4456.0</td>
<td>17711.0</td>
<td>8969.2</td>
<td>20807.5</td>
</tr>
<tr>
<td>2004</td>
<td>4747.1</td>
<td>22858.2</td>
<td>7238.5</td>
<td>23439.5</td>
</tr>
<tr>
<td>2005</td>
<td>6162.1</td>
<td>28236.5</td>
<td>7997.3</td>
<td>28414.2</td>
</tr>
<tr>
<td>2006</td>
<td>7375.4</td>
<td>22050.3</td>
<td>4941.2</td>
<td>21933.2</td>
</tr>
<tr>
<td>2007</td>
<td>10661.2</td>
<td>17314.8</td>
<td>4117.7</td>
<td>18539.5</td>
</tr>
<tr>
<td>2008</td>
<td>11648.0</td>
<td>10314.3</td>
<td>5227.2</td>
<td>13993.0</td>
</tr>
</tbody>
</table>

Note: the dwellings include villas, high-end flats, economically affordable housing and ordinary apartments.

The real estate reforms have enabled more people to own their own homes and to enjoy more spacious housing than in the past. According to a survey conducted by the State Statistical Bureau in 2010, the households that possess their own home account for 72.8% of the total. Moreover, the average dwelling space per capita for urban residents has increased from 6.7 square meters in 1978 to 21.56 square meters in 2008 (See Graph 2). Meanwhile, the total income per household has increased by
15.14% on average since 1978, reaching 27,678 yuan per year, (US$ 4053.484), but when real estate prices are considered in relation to earnings, it can be seen that the cost of purchasing a home is extremely high for the average Beijing inhabitant.

3. Duration models in housing. A review

Duration models are applied in diverse research fields, with the aim of analyzing the duration of an effect. Vandell et al. (1993) analyse individual commercial mortgage default with data from a major multi-line USA insurance company with a Cox proportional duration model, concluding that the dominance of the equity effect, as proxied by contemporaneous market loan-to-value ratios, is a major covariance explaining defaults, alongside loan structure. Secondary effects were cash flow effects, borrower effects and property-type. Sternberg (1994) analyses the duration of rental unit vacancies with a gamma mixture survival model, adjusted for length, truncation and survivorship-bias, of exponential distributions with hazard rate linear in dwelling characteristics. The conclusion is that duration is directly related to dwelling age, atypicality, the degree to which the neighbourhood is "run down," and the number of rental units in the building. Gabriel and Nothaft (2001) analyse the duration of rental vacancies in major USA metropolitan areas with ordinary least square regressions. Ciochetti et al. (2003a) analyse the factors driving the borrower’s decision to terminate commercial mortgage contracts with the lender through either prepayment or default. Data on loans from large USA insurance companies are used. The Cox proportional duration model with competing risks is adopted, accounting for unobserved heterogeneity and identifying two distinctive borrower groups. It is found that implicit put and call options drive default and prepayment actions in a nonlinear and interactive fashion. High value of the put/call option is found to significantly
reduce the call/put risk since the borrower forfeits both options by exercising one. Cash flow and credit conditions as well as ex post bargaining powers are also found to have significant influence upon the borrower’s mortgage termination decision. Ciochetti et al. (2003b) analyse loans default with a proportional Cox model with competing risks hazard model using data on loans from large USA companies finding that probability of default is positively related to loans-to-value ratio, along the debt service coverage ratio.

Deng, Gabriel and Nothaft (2003) analyse the duration residence in rental housing in major USA metropolitan areas with a proportional Cox hazard model concluding that residence duration is positively explained by median housing costs, public housing share of rental stock, poverty rate and African-American and Hispanic share of tenant households. Negatively covariates are elevator buildings, unemployment rate, population growth and central city share.

Lacour-Little and Malpezzi (2003) analyze the quality of mortgages loans in Alaska with the Cox proportional duration model, concluding that decreasing appraisal quality is associated with an increase in the mortgage default hazards rate. Cunningham (2006) analyzes the delay in the timing of land development in Seattle with an exponential, a Weibull and a Cox proportional survival model, concluding that uncertainty over the future prices reduces the hazard of current-period development, and at the same time raises land prices. Another interesting paper is Foote, Gerardi and Willen (2008) where the authors relate negative equity and mortgage default in a dataset of Massachusetts homeowners during the early 1990s. Bulan, Mayer and Somerville (2009) analyze the extent to which uncertainty delays investment in condominium development in Vancouver, Canada with a Weibull survival model. These authors conclude that builders delay development during times
of greater idiosyncratic uncertainty in real estate prices and when the exposure to market risk is higher. Kau, Keenan and Li (2011) analyze residential mortgage defaults and prepayments, using a Cox survival model with frailty. They investigate whether mortgages originated in the same Metropolitan Statistical Area share any common factors. Feijten and Mulder (2010) analyze the timing of moving into long-stay housing based on postponed marriages and childbirth in the Netherlands with the Cox survival model. They conclude that the macro context explains the postponement decision. Ström (2010) analyzes the extent to which the housing-type decision is constrained by first births in Sweden. A Kaplan-Meier survival curve is used to analyze the time taken to move to a new dwelling after the first birth and piece-wise constant hazard survival models are used to analyze the dwelling choice. The author concludes that the size of dwelling seems to be the most important housing-factor decision related to the first birth. Archer, Ling and Smith (2010) analyze housing turnover rates in Chicago between 1992 and 2002 using a Cox proportional hazards model concluding that household characteristics are the most important determinant of housing turnover rate.

Chinese housing-related variables have been examined by authors such as Peng and Wheaton (1994), Fu, Tse and Zhou (2000) and Yan (2006) though none of them use duration models in their research.

From this literature survey, it is worth noting that almost all survival models have focused on USA mortgage market (Vandell et al., 1993; 2001; Ciochetti et al., 2003a,b; Lacour-Little and Malpezzi, 2003; Foote, Gerardi and Willen, 2008; Kau, Keenan and Li, 2011) and duration of rental vacancies in USA (Sternberg, 1994; Gabriel and Nothaft, 2001; Deng, Gabriel and Nothaft, 2003; Archer, Ling and Smith, 2010); timing of land development in USA (Cunningham, 2006). Papers outside USA
include Bulan, Mayer and Somerville (2009) in Canada, Feijten and Mulder (2010) in the Netherlands, and Ström (2010) in Sweden. Therefore the focus on a main city and the analysis of sales duration and adoption of a sample selection survival model have not yet been used in urban research. Therefore, the present research innovates in this context.

4. Theoretical Framework

The survival model estimated in this study is based on the theory of consumer behavior developed by Lancaster (1966) and the concept of hedonic prices (Rosen, 1974). The economic theory of consumer behavior assumes that a consumption decision faced by an individual aiming to maximize utility, subject to budgetary restrictions, is taken on the basis of prices and income (Varian, 1987). This traditional framework, however, does not allow for circumstantial conditions and product characteristics which are known to play an important role in shaping housing demand, given the composite or differentiated nature of the services that comprise a housing market (Garcia and Raya, 2010). Therefore, the demand for houses is derived from the demand for the various goods and services offered by a house market (Ben-Akiva and Lerman, 1985), adopting the Lancaster approach.

4.1 The Method

In this study, the housing sales delay is analyzed by means of a survival modeling approach (Cox and Oakes, 1984; Allison, 1984; Yamaguchi, 1991; Hosmer and Lemeshow, 1999; Kalbfleisch and Prentice, 2002; Cleeves, Gould and Gutierrez, 2002). Survival analysis, also known as the duration model, measures the duration of
an event, which is the time elapsed until a certain event occurs, or is completed. The length of a vacation is an example of a duration event (Lacour-Little and Malpezzi, 2003; Cunningham, 2006; Bulan, Mayer and Somerville, 2009; Feijten and Mulder, 2010; Ström, 2010; Kau, Keenan and Li, 2011).

The use of survival models in order to model duration is based on the fact that the error distribution in this context, by necessity, must be skewed to the right (Hosmer and Lemeshow, 1999). The survival model regresses the duration of an activity on covariates. Such models are generally adopted in economics on the basis of two facts: (i) time, as a dependent variable, is strictly positive, and therefore the use of the traditional Gaussian distribution is not adequate to capture the characteristics of the time variable; and (ii) in clinical trials, censoring occurs when an individual participant in the initial phase of the study subsequently dies. Survival analysis can adequately accommodate the loss of observations when censoring occurs.

Traditional regression models are unable to resolve these issues and therefore, survival models, such as the Cox model and the Weibull model, have appeared (Hosmer and Lemeshow, 1999). The dependent variable of interest is the number of days between the property being put on the market and its sale. Three issues must be addressed when analyzing survival models: 1) identification of the data set (i.e., cross-section vs. time series); 2) censoring of the data; and 3) sample selection. With regard to the first issue, the present study opted for a time series approach. Therefore, time-variant modeling, known as non-proportional hazard models, was adopted (Wooldridge, 2002).

A survival time is described as censored when there is a follow-up time, despite the fact that the event has not yet occurred, or is not known to have occurred. For example, if the length of time being studied concerns a home that is on the
market, but remains unsold at the end of the data-gathering phase, then the start of 
selling is observed, and the end-time would be censored. With regard to the third 
issue, ignoring sample selection may bias the results (Cameron and Triverdi, 2005).

4.2 Sample Selection

Sample selection occurs when unobserved factors that influence the process of 
interest help to determine whether that process is observed. When this happens, values 
of the dependent variable for uncensored observations are systematically 
unrepresentative of the population studied. If high-quality dwellings that sell well are 
registered, while low-quality homes that are not easily sold are not registered, we 
have a selection problem.

Selection bias can be avoided by estimating a model that accounts for the 
satisfaction while simultaneously estimating the factors that influence the duration. 
Relative to the length of sale, let us assume that some high-quality houses sell well 
and therefore they are reported for statistical purposes. The statistics are biased only 
towards those high-quality houses that sell well. Under such conditions, ‘quality’ will 
be endogenous. Controlling for repetitions will enable the researcher to correct the 
results.

The sample selection model used in this paper was proposed by Boehmke, 
Morey and Shannon (2006). The objective is to model a continuous duration variable 
$y_i$ with an associate density $f(y_i)$. The selection depends on a set of covariates by 
letting $y_i = \exp(-X_i \beta) \varepsilon_i$. The selection process leads us to observe $y_i$ only for 
observations for which the binary censoring variable $c_i$ takes the value 1 (repeat) and 
d below is an indicator of right censoring. The Weibull likelihood with right censoring 
is the following:
\[ \ln L(\beta, \gamma, \alpha, p | X, W, y, c, d) = \]
\[ \sum_{i=1}^{n} c_i (1 - d_i) \left[ -\lambda_{2i} + \ln(1 + \alpha(2\exp(-\lambda_{2i} y_i) - 1)\exp(\lambda_{1i} - 1)) + \right. \]
\[ + \ln(p) \ln(\lambda_{2i}) + (p - 1) \ln(\lambda_{2i} - (\lambda_{2i} y_i)^p) + \]
\[ \left. + c_i d_i \left[ -\lambda_{2i} - (\lambda_{2i} y_i)^p + \ln(1 + \alpha(1 - \exp(-\lambda_{2i} y_i)^p)(1 - \exp(-\lambda_{1i}))) \right] \right] \]
\[ + (1 - c_i) \left[ \ln(1 - \exp(-\lambda_{1i})) \right] \]

where \( \beta, \gamma, \) and \( \alpha \) are parameters to be estimated; \( p \) is the Weibull distribution shape parameter when the error term follows an exponential distribution; \( X \) is a vector of explanatory variables; \( W \) is the Weibull distribution; \( y \) is the length of stay; \( c \) is a censoring binary variable; and \( d \) is an indicator for right-censored variables. \( \lambda_{1i} = \exp(-w_{i\gamma}) \) and \( \lambda_{2i} = \exp(x_i \beta) \) (Boehmke, Morey and Shannon, 2006).

5. Research Design: The Hypotheses

As noted above, the delays of sales can be explained by several factors. Based on the available data set, the following characteristics were used: 1) price of dwelling; 2) area of the dwelling; 3) building type, 15-24 floors; 4) property management fee; 5) location, and 6) quality. Using the data on these characteristics, we tested the following hypotheses.

*Hypothesis 1 (Price):* the delay on a dwelling’s sale is a negative function of its price, signifying that the higher the price, the shorter the time to sell, because rich people who need a house can buy it. The credit rationing exists for lower middle class. Based on the fact that the Chinese present development middle class is estimated as 37% in 2009 compared with the South Korean 67% (Amadeus, 2010), this hypothesis makes sense in this context. Furthermore, there is an abundant literature on house prices, using various methods and analyzing distinct regional markets. However, Strange
(1992) notes that these studies have had minimal success in finding significant effects that explain dwelling prices. Sirmans et al. (2005) review the hedonic pricing models of many empirical studies, recognizing that there is some parameter of uncertainty, even for key housing characteristics. This research justifies the importance of the housing price in any market decision, supporting the present hypothesis.

**Hypothesis 2 (Area of dwelling):** the delay on a dwelling’s sale is a negative function of the dimensional area of the dwelling. The importance of a property’s area is well recognized in housing markets, (Forrest et al., 1996), which therefore justifies the inclusion of this variable in the present research.

**Hypothesis 3 (Building type):** the delay on a dwelling’s sale is a positive function of the dwelling type. Given China’s vast population, the most common form of housing construction is the apartment tower block. According to the classification criteria applied by Chinese real estate companies, if the number of the floors is above 25, the building is classified as a ‘super high-rise’ building, 15-24 floors as a ‘high-rise’ building, 8-14 floors as a ‘middle-rise’ building, and 4-7 floors as a ‘multi-rise’ building. If the number of floors does not exceed 4, it is a low-rise building, which is the category in which homes are priced beyond the means of most households. This is a traditional hypothesis in housing demand studies (Raya, Montolio and Garcia, 2010; Garcia and Raya, 2010).
Hypothesis 4 (Management fee). The delay on a dwelling’s sale is a negative function of the dwelling management fee, Anderson (2005). The service of property management fee is a traditional cost, but as explained in the contextual setting is new for Chinese citizens and so they may react negatively to this price as it happens to any price effect.

Hypothesis 5 (Green): The environmental location of a dwelling as characterized by the proportion of green space in the building’s immediate vicinity affects negatively the delays on a dwelling’s sale. Reasons for this negative effect are unclear in the literature (Conway et al., 2010), but may be related to the country origin of most city dwelling Chinese.

Hypothesis 6 (Location): the delay on a dwelling’s sale is a positive function of its location (Li and Brown, 1980; Kiel and Zabel, 2008). This is also a traditional hypothesis in house demand, which justifies the inclusion of a location variable in the present research (Bonnet, Gobillon and LaFerrère, 2010).

Hypothesis 7 (Quality): the delay on a dwelling’s sale is a positive function of the quality of its decoration and fittings (Zabel, 1999). Quality is a main characteristic in housing and the price-quality relationship is a main issue on buying a house (Cheung, Ihlanfeldt and Mayock, 2009)

5.1 Study Context
The research study conducted on the Beijing housing market is based on data gathered from the China Real Estate Statistics Yearbook 2001-2009, in relation to dwellings sold in the period 2001-2008. Table 2 summarizes the characteristics of the data.

### Table 2: Characterization of the Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Min(^a)</th>
<th>Max(^b)</th>
<th>Mean</th>
<th>Std. Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delay</td>
<td>Number of days between home’s appearance on the market and its sale</td>
<td>11</td>
<td>1057</td>
<td>433.10</td>
<td>348.45</td>
</tr>
<tr>
<td>Price</td>
<td>The price of the dwelling (yuan/square meter)</td>
<td>2580</td>
<td>75000</td>
<td>11613.08</td>
<td>7880.6</td>
</tr>
<tr>
<td>Area</td>
<td>The area of the dwelling (floor space in square meters)</td>
<td>829</td>
<td>8000000</td>
<td>101212.6</td>
<td>346793.5</td>
</tr>
<tr>
<td>Type 1</td>
<td>Building type $15-24=1$, otherwise $=0$</td>
<td>0</td>
<td>1</td>
<td>0.421</td>
<td>0.494</td>
</tr>
<tr>
<td>Type 2</td>
<td>Building type $8-14$, otherwise $=1$</td>
<td>0</td>
<td>1</td>
<td>0.501</td>
<td>0.500</td>
</tr>
<tr>
<td>Fee</td>
<td>The building’s management fee (yuan/square meter/month)</td>
<td>0.4</td>
<td>28</td>
<td>2.943</td>
<td>1.989</td>
</tr>
<tr>
<td>Green</td>
<td>Green area proportion</td>
<td>5</td>
<td>80</td>
<td>34.557</td>
<td>7.971</td>
</tr>
<tr>
<td>Location 1</td>
<td>City location, defined by the 3rd to 4th Ring Roads $=1$, otherwise $=0$</td>
<td>0</td>
<td>1</td>
<td>0.258</td>
<td>0.437</td>
</tr>
<tr>
<td>Location 2</td>
<td>City location, defined by the 2nd to 3rd Ring Roads $=1$, otherwise $=0$</td>
<td>0</td>
<td>1</td>
<td>0.228</td>
<td>0.419</td>
</tr>
<tr>
<td>Quality</td>
<td>Quality index, defined as: poor decoration $=1$, blank/adequate decoration/undecorated $=2$, good decoration $=3$, high-quality finishings $=4$</td>
<td>1</td>
<td>4</td>
<td>2.623</td>
<td>0.990</td>
</tr>
</tbody>
</table>

\(^a\) Min – Minimum; \(^b\) Max – Maximum

In order to evaluate the degree of correlation among the explanatory variables, which can cause multicollinearity, the correlation matrix is estimated in Table 3.

### Table 3: Correlation between the variables

<table>
<thead>
<tr>
<th></th>
<th>Price</th>
<th>Area</th>
<th>Type 1</th>
<th>Type 2</th>
<th>Fee</th>
<th>Green</th>
<th>Loc 1</th>
<th>Loc 2</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area</td>
<td>-0.0398</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type 1</td>
<td>-0.0622</td>
<td>-0.0324</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type 2</td>
<td>-0.0831</td>
<td>0.005</td>
<td>-0.1465</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fee</td>
<td>0.1316</td>
<td>-0.0767</td>
<td>0.0475</td>
<td>-0.1021</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Green</td>
<td>-0.0355</td>
<td>0.0822</td>
<td>-0.0147</td>
<td>0.0474</td>
<td>-0.0364</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loc 1</td>
<td>0.1055</td>
<td>-0.1078</td>
<td>0.1193</td>
<td>-0.0340</td>
<td>0.1148</td>
<td>-0.1752</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
It is observed that the degree of correlation among the variables is small.

6. Results

Table 3 presents the results of the estimated duration models. The Weibull model is presented, both without and with sample selection, for comparative purposes. The dependent variable is the number of days between the date of the property first appearing on the market and its sale. Positive covariates increase the hazard, e.g./i.e. the number of days unsold on the market. Negative covariates decrease the hazard.

Table 4: Survival Model Results

<table>
<thead>
<tr>
<th></th>
<th>Weibull model without sample selection</th>
<th>Weibull model with Sample Selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality</td>
<td>-0.026</td>
<td>-0.103</td>
</tr>
<tr>
<td></td>
<td>(-0.57)</td>
<td>(-3.21)</td>
</tr>
<tr>
<td>Constant</td>
<td>—</td>
<td>2.60</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(7.55)*</td>
</tr>
<tr>
<td>Duration</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log price</td>
<td>-0.149</td>
<td>-0.448</td>
</tr>
<tr>
<td></td>
<td>(-2.44)</td>
<td>(-4.55)</td>
</tr>
<tr>
<td>Log area</td>
<td>-0.090</td>
<td>-0.087</td>
</tr>
<tr>
<td></td>
<td>(-1.40)</td>
<td>(-2.49)</td>
</tr>
<tr>
<td>Type 1</td>
<td>-0.149</td>
<td>0.128</td>
</tr>
<tr>
<td></td>
<td>(-0.95)</td>
<td>(1.94)</td>
</tr>
<tr>
<td>Type 2</td>
<td>-0.177</td>
<td>0.109</td>
</tr>
<tr>
<td></td>
<td>(-1.99)</td>
<td>(2.16)</td>
</tr>
<tr>
<td>Fee</td>
<td>-0.011</td>
<td>0.030</td>
</tr>
<tr>
<td></td>
<td>(-0.43)</td>
<td>(3.22)</td>
</tr>
<tr>
<td>Green</td>
<td>-0.005</td>
<td>-0.003</td>
</tr>
<tr>
<td></td>
<td>(-1.03)</td>
<td>(-3.21)</td>
</tr>
<tr>
<td>Location 1</td>
<td>0.033</td>
<td>0.174</td>
</tr>
<tr>
<td></td>
<td>(0.34)</td>
<td>(2.02)</td>
</tr>
<tr>
<td>Location 2</td>
<td>-0.199</td>
<td>-0.094</td>
</tr>
<tr>
<td></td>
<td>(-1.90)</td>
<td>(-1.12)</td>
</tr>
<tr>
<td>Constant</td>
<td>-4.464</td>
<td>-4.125</td>
</tr>
<tr>
<td></td>
<td>(-4.75)</td>
<td>(-4.98)</td>
</tr>
<tr>
<td>Ln p</td>
<td>-0.016</td>
<td>0.061</td>
</tr>
<tr>
<td></td>
<td>(-0.51)</td>
<td>(1.96)</td>
</tr>
<tr>
<td>P (duration dependence)</td>
<td>0.983</td>
<td>1.063</td>
</tr>
</tbody>
</table>
The results obtained from the two models are clearly not similar in their main effects. Given the model specification, positive values for the parameters imply that the sale delay increases with increasing values in the respective variable. A negative value for the parameters implies a negative relationship. The results across the two models demonstrate that the parameters have different signs for endogenous variables, and the same signs but different values for exogenous variables, underlining the importance of controlling sample selection in survival models.

On the basis of the log likelihood statistic, the Weibull model with sample selection provides the superior fit to the data. The rationale for this result is that sample selection represents unobserved characteristics that influence the conditional probability of the sales delay which are not measured and therefore, not taken into account in the measurement errors of the variables; this can cause bias (Boehmke, Morey and Shannon, 2006). The rho (p) duration dependence parameter is higher than one in the sample selection model, implying that the sales delay increases with the events, but it is lower than one for the model without sample selection.

Taking as reference the Weibull model with sample selection, as hypothesized, the price has a positive effect on the hazard, which means that more expensive homes tend to take longer to sell. Moreover, the area has a positive impact on the hazard, signifying that larger houses have a longer delay. Furthermore, the building type also has a positive impact on hazards, which means that super high-rise buildings have an
increased hazard of delaying the selling of homes in this category. Moreover, the building management fee has a positive impact on the hazard, signifying that this attribute increases the sales delay. In addition, green space has a positive but statistically insignificant effect. Finally, quality decreases the sales delay, having a negative effect on the hazard.

7. Discussion and conclusions

This paper presents a sample selection duration model to measure housing sales delays in Beijing, China (Boehmke, Morey and Shannon, 2006). The sample selection accounts for unobserved characteristics that influence the conditional probability of the sales delay. Sample selection in market data is a common technique which has not previously been applied in housing research.

Conclusions drawn without taking sample selection into account may be erroneous (Heckman, 1979). In the present case, the sample selection variables change the sign and the value when sample selection is taken into account, signifying that traditional survival models may distort results when they do not consider the sample selection problem. Quality, a variable commonly used in housing market studies, is found to be sample-selection in the present research. This is an intuitive result, since quality is a major factor among dwelling’s characteristics.

The general conclusion is that most of the hypotheses presented in our study are rejected. Firstly, price is negative signifying that the higher the price, the shorter the delay, which implies that higher priced homes sell easier than lower priced homes. This is a situation found in markets with high growth rates and high income inequality, in which poorer households cannot afford to buy there homes, whereas the more affluent can do so with ease. Second, the greater the dimensions of the property,
the shorter the delay, reinforcing the price effect, which is explained by owners born on countryside and now buying city flats. Third, the higher the building type, the longer the sale delay, signifying that Chinese house owners do not like tall buildings. Fourth, the higher the management fee, the higher the sale delay, due to the price effect the management fee has on selling. Fifth, the presence of gardens will decrease the sales delay, which signify a preference for gardens near the house. Sixth, location 1 (defined by the 3rd to 4th Ring Roads) increases the sales delay, but location 2 (defined by the 2nd to 3rd Ring Roads) decreases it, which is due to their location. Seventh, quality decreases the length of stay, invalidating Hypothesis 7. Therefore, more expensive, higher-quality homes have the shortest sales delays in the Beijing housing market.

With regard to the policy implications, while recognizing that it is not possible to control all the elements contributing to sales delays, it is possible to concentrate on those identified in the present research, to eliminate it. Therefore, the results suggest that home realtors seeking to manage housing sales in Beijing should target the more expensive, high-quality homes, since they have shorter sales delays.

The research found that a survival model that accounts for sample selection is superior to models that do not do so. (Boehmke, Morey and Shannon, 2006). Therefore, sample selection should be taken into consideration in researching housing markets. With regard to the managerial implications, this research provides a deeper insight into the driving motives behind housing sales decisions and can thus aid the development of the marketing strategies of home realtors. The statistically negative significant variables should be used in promotional strategies to reduce sales delays.

The sample employed in this research was restricted to a single market during a limited number of months, thus the conclusions cannot be generalized to other
housing contexts. However, since this research was an exploratory study designed to investigate and demonstrate the suitability of survival modeling as a means of developing a deeper insight into the factors which drive housing sales, these limitations are not critical. Survival modeling has been shown to be a useful technique for this purpose.

With regard to comparisons of these findings with previous research, this paper is innovative, in that statistically significant variables were chosen to explain the sales delay duration. Consequently, our findings are not directly comparable with those of previous studies.

Finally, in the present research, a parsimonious procedure identified the small group of statistically significant variables among the many insignificant variables. There is, thus, a need for further research to apply survival modeling in analyzing the determinants of sales delay duration in a variety of other contexts. As the body of research into the determinants of housing sales delays grows, at some point in time it may be possible to investigate the differences in results across studies using meta-analytical techniques. More research is needed to confirm the present findings.
Acknowledgments

Comments from the Editor and from two anonymous referees are gratefully acknowledged.

Carlos P. Barros: UECE (Research Unit on Complexity and Economics) is financially supported by FCT (Fundação para a Ciência e a Tecnologia), Portugal. This article is part of the Multi-annual Funding Project (POCI/U0436/2006).

Luis A. Gil-Alana gratefully acknowledges financial support from the Ministerio de Ciencia y Tecnología (ECO2008-03035 ECON Y FINANZAS, Spain) and from a PIUNA Project from the University of Navarra.
References


