Application of Hedonic Price Model in the Prague Property Market

Jan MELICHARa, Ondřej VOJÁČEKb, Pavel RIEGERc, Karel JEDLÍČKAd

a Charles University Environment Center, jan.melichar@czp.cuni.cz
b Department of Environmental Economics, University of Economics, ovojacek1@centrum.cz
c Department of Law, University of Economics, pavel.rieger@vse.cz
d Department of Mathematics, University of West Bohemia, smrcek@kma.fav.zcu.cz

Abstract

Hedonic price model was developed to quantify the influence of structural, accessibility and environmental attributes on the price of housing in the city of Prague. We have considered the size of flat as structural variable, the distance from the dwelling to the city center and to the nearest tube station as accessibility variables, and the proximity to the nearest urban forest as environmental variable. We have developed several regression models and tested their robustness in terms of the significance of parameters and the amount of variability explained. The sample is made up of 1,708 observations gathered from the city of Prague. Results show the size of flat is the most relevant variable explaining the price of housing. We have found also a significant inverse relationship between the selling price and its distance from a city center and urban forests.

Keywords: hedonic price, regression analysis, geographical information system, property market

Introduction

The paper presents hedonic price model (HPM) used to estimate the differential premium on property value derived from structural attributes and from proximity to some neighborhood and environmental attributes. This methodology is applied to estimate the implicit value of housing proximity to city center, tube station and urban forest in the city of Prague. We have tested various types of regression models to infer the significant determinants of housing market in Prague. Relevant information used in the regression analysis comes from a sample of 1,708 residential dwellings.

Hedonic price model is derived from Lancaster’s consumer theory [Lancaster, 1966] and it empirical models have been applied since the 1960s. HPM is based on the idea that properties are not homogenous and can differ in respect to a variety of characteristics. The method relies on the fact that house prices are affected by many factors such as number of rooms, size of flat or garden, or access to city center. We suppose that the price of housing is determined by the particular combination of characteristics it displays. We can describe any particular property by the quantities and the qualities or characteristics of its structure, location and environs.

To reveal the empirical relationship between the price of housing and its characteristics, the regression analysis is used as the primary statistical tool. Property prices are regressed against sets of explanatory variables, these are typically calculated using structural attributes of the

* Address of correspondence: Jan Melichar, U Křiže 8, 158 00 Prague, Czech Republic
housing (such as number of bedrooms), neighborhood variables (quality of local schools), accessibility variables (proximity to railway station) and environmental variables (ambient noise level). The results of regression analysis are used to derive a hedonic price function that indicates how much the price of housing will change for a small change in each characteristic, holding all other characteristics constant. The hedonic price function can be used to determine how much more must be paid for a property with each extra unit of a particular housing characteristic. This is known as the implicit price of a property characteristic.

The application of the hedonic price method to examine the impacts of environmental characteristics on the house prices has been widespread, ranging from air quality to noise exposure, landscape and urban amenities. In this study, our primary interest is to examine the impact of urban forest on the house prices. More than 30 studies have shown that people are willing to pay more for a housing located close to an urban open space than for a house that does not offer this amenity. With few exceptions [Luttik, 2000; Schroeder, 1982] studies find that homes adjacent to naturalistic parks and open spaces are typically valued at about 8 to 20 percent more than comparable properties [Crompton, 2001]. Values show a linear decline with distance from the edge of an open space, with a positive price effect declining to near zero at about a half mile away [Hammer et al., 1974; Tyrvainen et Miettinen, 2000; More et al., 1988]. Also, the study done by Morancho [2008] indicates that the price of the housing relates inversely with the distance that separates it from an urban green space. This result is in accordance with Bolitzer and Netusil [2000] who concluded that proximity to an open-space can have a statistically significant effect on home selling price. As well as with Tyrvainen and Miettinen [2000] whose results demonstrated that a 1 km increase in the distance from the nearest forested area leads to an average 5.9 % decrease in the market price of the dwelling.

In the following sections of the paper, the theoretical foundations of the method, the description of data set, the empirical models used and the results obtained are presented.

**Theoretical rinciples of hedonic price method**

Griliches [1971] and Rosen [1974] provided the theoretical support for the development of the hedonic price models. HPM relates the market price of the housing commodity (house, flat or rental housing) with the structural and accessibility characteristics that define it. It enables us to derive the monetary value of each characteristic that is calculated by observing the differences in the housing prices sharing the same attributes. The basic hypothesis is that the housing is formed by a heterogeneous set of characteristics. Therefore, when buying a house, we consider the price paid as the sum of the prices paid for each attribute (called implicit price). This could be expressed as follows:

\[
P = f(x_1, x_2, \ldots, x_n)
\]

where \( P \) is the market price of the house and \( x_1, x_2, \ldots, x_n \) are the housing characteristics it embodies. The partial derivatives of the price with respect to the previous variables, \( \frac{\partial P}{\partial x_i} \), provide information on the marginal implicit price for an additional unit of each characteristic.

The set of structural and accessibility variables such as size of flat could not be the only determinants that explain the price of housing commodity. There are also environmental aspects such as proximity to a green area, noise level that could also explain the differences in
its market price. Then, the price function specified in equation (1) can be formulated as follows:

\[ P = f(x_1, x_2, ..., x_n, z) \]  

(2)

where \( P \) is again the house market price; \( x_1, x_2, ... , x_n \) are structural and accessibility characteristics and \( z \) is the environmental variable without a market price (the hedonic variable). The essence of the method consists of finding what portion of the price is determined by the hedonic variable. This information is obtained again by calculating the partial derivative of the price with respect to the variable \( z \), \( \frac{\partial P}{\partial z} \), which gives us the marginal implicit value for an additional unit of the environmental asset, and thus allows us to obtain an estimate of its monetary value. The theoretical model specified in equation (2) has been used to obtain the implicit price of the distance to urban forests as environmental variable considered in this study.

**Study area and data**

In order to apply the hedonic price model we had to obtain relevant information about sales of dwellings realized on the Prague’s real estate market. Individual housing characteristics including selling price were obtained from the real estate catalogue operating by the Czech company reality.cz (see [http://www.reality.cz](http://www.reality.cz)). The data set contains information on 1,708 apartments and flats sold to personal ownership from 2005 to spring 2008. See the geographical position of data set in Figure 1 as was detected by ArcExplorer.

**Figure 1: Delineation of geographical position of the data set using ArcExplorer**

In this study, the distance from the dwelling to the city center, the nearest tube station (accessibility variables) and the nearest urban forest (environmental variable) were measured by Geographical Information System (GIS), software ArcGIS respectively. GIS also served as
the important tool for detecting wrong observations (duplicity in observations, wrong positions of dwellings etc.). Altogether, 279 records have been detected as wrong, hereafter not used in the further analysis. All variables used in the regression analysis, including their definitions and expected influence on the dependent variable are summarized in Table 1.

Table 1: Dependent, structural, accessibility and environmental variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Expected sign</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price</td>
<td>Price of the property (in CZK)</td>
<td>DV</td>
</tr>
<tr>
<td>Flat_area</td>
<td>The area of the flat (in the square meters)</td>
<td>+</td>
</tr>
<tr>
<td>Center</td>
<td>Distance to the city center (in meters)</td>
<td>-</td>
</tr>
<tr>
<td>Tube</td>
<td>Distance to the nearest underground station (in meters)</td>
<td>-</td>
</tr>
<tr>
<td>Forest</td>
<td>Distance to the edge of the nearest urban forest (in meters)</td>
<td>-</td>
</tr>
</tbody>
</table>

Note: DV marks dependent variable
* + increasing / - reducing effect on the selling price

Variable Price is the selling price of flat or apartment expressed in CZK and represents the dependent variable. In the proposed regression models the variable price is explained by the following explanatory variables:

- Flat_area: the size of flat (structural variable),
- Center: distance to the city center represented by the Saint Wenceslas Statue (accessibility variable),
- Tube: the proximity to the nearest tube station (accessibility variable),
- Forest: distance to the nearest urban forest in meters (environmental variable).

Altogether, we have considered 21 of the main urban forests and forest parks in Prague.

Table 2 gives maximum, minimum and average values for the dependent variable and the explanatory variables.

Table 2: Descriptive statistics (n = 1,708)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price</td>
<td>4,008,013</td>
<td>2,987,803</td>
<td>400,000</td>
<td>35,000,000</td>
</tr>
<tr>
<td>Flat_area</td>
<td>74</td>
<td>36</td>
<td>15</td>
<td>455</td>
</tr>
<tr>
<td>Center</td>
<td>5,130</td>
<td>3,355</td>
<td>86</td>
<td>20,435</td>
</tr>
<tr>
<td>Tube</td>
<td>1,432</td>
<td>1,564</td>
<td>17</td>
<td>9,774</td>
</tr>
<tr>
<td>Forest</td>
<td>1,560</td>
<td>1,333</td>
<td>1</td>
<td>7,074</td>
</tr>
</tbody>
</table>

The average price of flat is 4 Mil. CZK and ranges from 400 thousands CZK to 35 Mil. CZK. The size of flat is in average 74 m². The average distance from dwelling to the city center is slightly above 5 km, on the other hand the distance to the nearest tube station is in average around 1.5 km, and the proximity to the nearest urban forest is slightly above 1.5 km. These results indicate that, in general, people prefer to live in the suburb parts of Prague, quite far from the city center, but with an easy accessibility to the tube station and the green area.

Empirical models
The relationship between the selling price and the characteristics of housing can take several functional forms. In this study, we have specified four functional forms: linear, log-linear, linear-log, and log-log.

If the linking price relationship with the characteristics of housing is assumed to be linear (both the dependent and explanatory variables enter the regression in their linear form), the equation (2) becomes:

\[ P_i = b_1 x_{1i} + b_2 x_{2i} + ... + b_n x_{ni} + b_z z_i + \epsilon_i \]

where \( x_{1i}, x_{2i}, ..., x_{ni}, z_i \) are variables describing the attributes of housing \( i \), parameters \( b_1, b_2, ..., b_n, b_z \) represent the marginal implicit price of each attribute and \( \epsilon_i \) is the error term. The marginal implicit price for an additional unit of the environmental good \( z \) is \( b_z \). Under the linear specification, the willingness to pay for an additional unit remains constant, i.e. it does not depend on the starting level of \( z \). This assumption is a strong restriction since, as Rosen [1974] points out, there are many reasons to suppose the relationship between the price and the environmental variable to be non-linear. Thus, logarithmic specifications are frequently formulated, as follows.

In the case of log-linear specification, the log of the dependent variable is regressed against linear explanatory variables:

\[ \ln(P_i) = b_1 x_{1i} + b_2 x_{2i} + ... + b_n x_{ni} + b_z z_i + \epsilon_i \]

In equation (4), both its slope and elasticity change at each point are the same sign as \( b_n \).

On the other hand, linear-log specification represents the relationship where a linear dependent variable is regressed against the log of the explanatory variables:

\[ P_i = b_1 \ln(x_{1i}) + b_2 \ln(x_{2i}) + ... + b_n \ln(x_{ni}) + b_z \ln(z_i) + \epsilon_i \]

The slope of the function is \( b_n/x \) and elasticity of the variable is represented by \( b_n/y \).

Log-log is the specification form where both the dependent and explanatory variables enter the regression in their log form, as follows:

\[ \ln(P_i) = b_1 \ln(x_{1i}) + 2 \ln(b x_{2i}) + ... + b_n \ln(x_{ni}) + b_z \ln(z_i) + \epsilon_i \]

A logarithmic model allows us to measure the impact that changes in explanatory variables cause in the dependent variable in relative terms. Then, the parameter \( b_n \) represents the value of elasticity of corresponding variable.

## Results

Table 3 presents the results of the ordinary least squares (OLS) regressions for the four specified models. The explanatory variables all are statistically significant at the 0.01 significance level, except Tube variable for linear, linear-log and log-log model that is not significant even at the 0.05 level. All significant variables have the expected signs. Log-linear
model however fail to confirm the researchers expectation, that the distance to the tube station would have negative impact on the price, i.e. the higher the distance to the tube, the lower the price.

The distance to the tube station is almost in all regression models insignificant, also it seems to be very unstable in the terms of the estimated sign. The reason is the high correlation between the distance to the city center and the distance to the nearest tube station. The Pearson correlation coefficient of these two variables is 0.54. That is why, Tube variable is excluded in the final model.

Table 3: Estimated coefficients for the specified models (t-statistic in parenthesis)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Linear</th>
<th>Linear-log</th>
<th>Log-linear</th>
<th>Log-log</th>
<th>Log-linear</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flat_area</td>
<td>54,000</td>
<td>(39,03)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Center</td>
<td>-264.76</td>
<td>(-13.54)</td>
<td>-0.000054</td>
<td>-0.25</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(-15.91)</td>
<td>(-21.17)</td>
<td></td>
</tr>
<tr>
<td>Tube</td>
<td>76.80(^\text{ns})</td>
<td>(1.80)</td>
<td>0.000023</td>
<td>0.003803(^\text{ns})</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(3.14)</td>
<td>(0.39)</td>
<td></td>
</tr>
<tr>
<td>Forest</td>
<td>-157.14</td>
<td>(-4.17)</td>
<td>-0.000028</td>
<td>-0.03</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(-4.21)</td>
<td>(-5.00)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>1,481,644</td>
<td>(-11,281(^\text{ns}))</td>
<td>15</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(9.09)</td>
<td>(-0.01)</td>
<td>(514.69)</td>
<td>(102.26)</td>
<td></td>
</tr>
</tbody>
</table>

All coefficients are significant at the 0.01 level.
\(^\text{ns}\)Not significant at the 0.05 level.
\(^1\)Explanatory variables are in their log form.
\(^2\)Dependent variable is in logarithmic scale.
\(^3\)All variables are in their log form.

The variability of the housing prices measured by \(R^2\) adjusted is accounted differently in the estimated models. Both models, linear and linear-log, explain 55% of the variation in the housing prices, log-linear accounts 62%. However, log-log model accounts for 75% of the price variance, thus provides the best fit from all these models.

To see how the predictability of log-log model is changed when the variable of the distance to the nearest tube station is excluded we have run OLS regression for log-log model once more. The results are presented in the following Table 4.

Table 4: Log-log model estimated by OLS, Tube variable excluded

<table>
<thead>
<tr>
<th>Variable</th>
<th>Unstandardized B coefficient</th>
<th>Standard error</th>
<th>Standardized B coefficient</th>
<th>t-ratio</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>13.56</td>
<td>0.13</td>
<td></td>
<td>103.40</td>
<td>0.00</td>
</tr>
<tr>
<td>Ln_Flat_area</td>
<td>0.88</td>
<td>0.02</td>
<td>0.69</td>
<td>51.53</td>
<td>0.00</td>
</tr>
<tr>
<td>Ln_Center</td>
<td>-0.25</td>
<td>0.01</td>
<td>-0.35</td>
<td>-25.62</td>
<td>0.00</td>
</tr>
</tbody>
</table>
The estimated coefficients have similar values as those derived in log-log model with Tube variable. Regarding the partial significance of each variable, the Student’s t-statistic reveals the housing living area as the variable with the greatest explanatory power. It is followed by the variable referring to the distance to the city center and to the nearest urban forest. The explanatory variables all are statistically significant at the 0.01 significance level.

The set of explanatory variables accounts for 71% of the price variance and the F-ratio test indicates that the model fits better (F-ratio = 1,363) than the log-log where the variable of the distance to the tube station is included (F-ratio = 1,022).

Using the estimated coefficients presented in Table 3, the relationship between the distance to the nearest urban forest and the price of housing with a given set of characteristics can be approximated by a hedonic price function. The hedonic price function derived for each model is presented in Figure 2. The hedonic price functions show the decrease in the housing price as the distance from urban forest increases while holding the other variables constant.

![Figure 2: Hedonic price function for each specified model](image)

However, the price reduction is quite moderate. This indicates low elasticity, i.e. the percentage change in the price of flat associated with a 1% change in the proximity to urban forest. The elasticity ranges in average from -0.06% for linear model to -0.03% for log-log model. Nevertheless, considering the hedonic price function derived from log-log model, it is evident that the price of flat falls down significantly up to 400 meters from dwelling, then decreases slightly.

According to the obtained coefficients in the final log-log model and having the variables at their mean values it can be said that:

<table>
<thead>
<tr>
<th>Ln_Forest</th>
<th>-0.03</th>
<th>0.01</th>
<th>-0.07</th>
<th>-5.24</th>
<th>0.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adjusted R2</td>
<td>0.71</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F-statistic</td>
<td>1,363</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>1,708</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Dependent variable as Ln of price
Each additional square meter in the flat size increases the price by 2,263 CZK, an increase in the distance from the city center about 1 meter decreases the price by 630 CZK, and the same increase in the distance from urban forest decline the price by 79 CZK. This means that a 1 km increase in the distance from the nearest forested area leads to an average 1.96 % decrease in the market price of the dwelling in Prague.

**Conclusions**

As preceding hedonic pricing studies have shown, the effects of urban environmental amenities such as urban forest on the price of housing are measurable and significant. Using a hedonic price model and ordinary least squares regression, we have estimated a hedonic price function of housing (flats and apartments) in which the sale price is related with the proximity to urban forests in the city of Prague. Together with a set of conventional explanatory variables such as the size of flat, the distance to the center city and the proximity to the nearest tube station, the distance to the nearest urban forest or forested park as environmental variable was included in the right-hand side of the regression.

We have tested four types of regression models. Compared to linear, linear-log and log-linear, log-log model accounts for 75 % of the price variance, thus provides the best fit from all these models. Regression analysis revealed the size of flat as the variable with the greatest explanatory power. Other statistical variables are the distance to the city center and to the nearest urban forest. As expected, the size of housing affects the price of flat positively, on the other hand, the distance to the city center and to the nearest urban forest decrease the price of housing. According to the estimates derived by log-log model, every 1 m further away from urban forest means a drop of 79 CZK in the housing price. It means that a 1 km increase in the distance from urban forest leads to an average 1.96 % decrease in the housing price. Compared to other studies that had assessed this effect at about 6 to 20 %, the distance effect in this study can be considered as modest. Although, the influence is moderate, the proximity to urban forest is significant determinant influencing people’s preferences where to live. The policy implication for urban planning could be to establish and maintain the green areas, even in the suburb parts of Prague where, at present, people prefer to live.

**Acknowledgement**

The research on this study was supported by the grant of the Internal Grant Agency, the University of Economics: IG 50 90 37 – “Application of the hedonic price theory at the labour and real estate market in the Czech Republic”. This support is gratefully acknowledged.

**References**


