The Analysis of the Potential Environmental Benefits by Investigating the Hedonistic Price

■

Giani Grădinaru
Ph.D. Lecturer
Raluca Ignat
Candidate Ph.D. Assistant
Academy of Economic Studies, Bucharest

Abstract. *Consumers get some usefulness from the attributes of heterogeneous products and they adjust their acquisitions as a response to the existing differences. Producers or vendors confront with varying prices depending on the scale of attributes provided. A plan for balancing the prices is developed as a consequence of the market interaction between the consumers and suppliers (by analogy it happens in the case of the interaction between employees and employers on the labour market). Taking into consideration the late concepts regarding hedonistic price, the article presents a way of analysis of potential benefits that environment may offer to human communities by hedonistic price investigation, using regression as instrument.*

Key words: regression; environmental benefits; variation explanation; hedonism.

■

JEL Codes: Q11, Q57.
REL Codes: 10D, 15C.

The model of establishing the hedonistic prices usually applies to the heterogeneous products and services: occupational opportunities, real estate components, computers, and machines etc. The models explain the price variations using information about their attributes. For example, salaries determination is expected to include the characteristics of an employee (educational level, age) and the job characteristics (risk of on-job accidents).

The hedonistic price helps in measuring the marginal function of damages. Such a functions measures the benefits to be received by an individual or group of individuals if a damaging phenomena is going to be reduced in a certain proportion.

Similar concepts: hedonistic salary and hedonistic property

The analysis of the hedonistic salary, known under the name of salary – risk analysis, starts from the pre-requisite that individuals choose jobs with a high occupational risk in exchange for some high salaries. Essentially, the jobs with a higher risk are paid with higher salaries. Such an analysis is made through the technique of statistical regression using the analysis model that takes into calculation the salary level as a dependent variable and the risk of the salary paid job as an independent variable.

The statistical model should allow the separation of the compensatory part associated with the risk of occupational health from other characteristics of the job including the management responsibility, security of job etc. The salary level is also influenced by the sector/domain in which the individual is employed, by the zonal characteristics and by the employee's personal characteristics (age, education, experience). All these data are necessary to disentangle the effects of the employee's characteristics from the ones of the job's attributes in determining the payable salaries.

In an analysis of the hedonistic salary the most difficult issues to solve are:

- the necessary data;
- implications;
- how to determine risk;
- model for estimation.

The necessary data: the analyses in this category require huge quantities of information on the behaviour of market labour. The employee data and the job characteristics are generally collected through the statistical inquiry and the risk info is frequently taken over from published sources, existing at the companies' level or at the level of the business domain.

Implications: the high risky jobs tend to attract the least holdouts when assuming risk in question and they do not always ask for compensation adequate to facing risk.

How to determine risk: certain analyses use actual calculation algorithms to determine the risk levels that an employee confronts with, that are not limited to the occupational risks including the morbidity and mortality risks the individual faces with at job and out of job.

Model for estimation

This must contain a sufficiently great number of variables so that the estimation would be clear. The analysis for determining the value of the hedonistic real estate property are based on the fact that individuals perceive dwelling as lots of attributes and obtain various levels of usefulness from different combinations of those attributes. When transactional decisions are made, individuals exchange money and attributes. The exchanges reveal the marginal values of the attributes representing the nucleus of the hedonistic real estate property value. In such analyses the technique of the statistical regression an real estate market information are used to examine increases in the values of properties associated with the different attributes.

Structural attributes (number of bedrooms and the age of the house), the attributes of the surroundings (demographic structure of population, number of crimes, quality of school) and the environmental attributes (air quality and nearby places to divert waste) could influence the values of properties.

When assessing an environmental improvement it is essential to separate the effect on a real estate property of the relevant environmental attributes from the effects of the other attributes. The applications lay stress on the relationships between the real estate property values and the environmental attributes such as: air quality, water quality, location nearby places to divert waste and the landscape characteristics. Based on the data collected for a cross-section of transactions, the price correlated with the observable attributes are analyzed with the help of the statistical regression and a hedonistic price function is estimated.

In such an analysis the most difficult issues to solve are:

- the necessary information;
- recording errors;
- measuring the environmental attributes;
- duration of effect on environments;
- model for estimation.

Necessary information: analyses require huge quantities of non-aggregated information. The prices of the market transactions for land surfaces or dwellings are preferred to the aggregated data (info obtained from the inventory of dwellings), as the aggregating problems can thus be avoided. Data on attributes can include: the characteristics of dwellings, seasonal sales, characteristics of surroundings (schools and parks) demographic characteristics and environmental quality.

Recording errors: there may appear errors in measuring the prices (aggregated data) and errors in measuring the statistical characteristics.

Measuring the environmental attributes: available information referring to the air or water quality is often used, and then is determined the way the info can be relevant also for the properties under study.

Duration of effect on environments: some effects on the environment manifest differently in time, and others can be understood differently in time depending on the available information.

Model for estimation: The selection of the function form, definition of the market size and identification the representative variables can create problems for the analysts.

Choosing the model of estimation: the regression models used in the statistical analysis of the environmental benefits through the technique of the hedonistic price are simple or multiple, linear or non-linear. The houses and land surfaces are the most commonly goods submitted to the statistical analysis through the technique of the hedonistic price.

Supposing there are two real estate properties with identical characteristics (built surface, land surface around the house, number of bedrooms etc.), except that one is near by a tip and the other farther away, the prices of the two louses are compared; and the price difference is assigned to the most expensive having the advantage of the environmental benefits.

If the real estate property nearby the tip is EUR 100,000 and the other is EUR 150,000, we could say that shutting down the tip would bring a benefit of EUR 50,000 to the owners of the cheaper house.

We will refer to a study that presented a mountain resort community. The existence of a tip at very close distance affects the quality of water from the lake, drinking water, air and soil and has different effects on community. The statistical observation period was: July-November 2007. The properties under study were those located at a distance of seven kilometres away from the tip at the most. The statistical characteristics are: the property price (C1), distance to the tip (C2), the house surface (C3) and the property surface around the house (C4). The collected data are included in table 1.

Distribution of properties depending on the statistical variables

Table 1

C_1 mii euro	C_2 km	C_3 m^2	C_4 m^2	C_1 mii euro	C_2 km	C_3 m^2	C_4 m^2
100	1	78	350	120	1	100	1543
110	2	78	370	100	1	100	654
110	2	78	500	115	2	100	384
140	3	80	1000	200	4	100	1783
140	4	80	100	300	7	100	2900
145	5	80	119	250	3	107	1000
145	6	81	200	150	3	108	100
150	7	81	700	150	3	109	67
110	1	85	250	190	1	110	589
120	2	85	300	260	3	110	500
120	3	85	200	300	5	110	500
200	4	85	700	300	6	110	500
150	5	85	200	200	1	110	476
115	6	85	160	125	1	120	200
100	1	85	200	200	3	127	700
130	2	85	312	180	3	128	300
190	7	85	207	180	1	141	300
170	6	85	226	230	5	142	399
100	1	93	167	300	7	142	2000
115	2	94	243	200	3	150	1000
119	1	95	156	200	1	150	1000
119	1	95	678	180	6	150	200
200	7	95	987	130	3	150	350
110	2	98	468	210	1	170	2000
250	6	98	1678	260	7	180	300

The distance to the tip should have an influence on the price of the house. Simple regression is used when there is a linear dependency between the two characteristics; it can be used when dependency is non-linear, but through various transformations it can be turned into a link for which parameters can be expressed. In this model, Y will be the price (dependent variable), and X the distance to the tip (explicative variable).

In the linear model, we will say that the dependent variable is expressed linear, depending on the independent variable. As variable Y is also influenced by a series of randomly, non-quantifiable factors, we are going to consider a probabilistic model of analyzing the dependency written in the form $Y = \alpha + \beta \times X + \varepsilon,$ where ε is the random variable, and $\alpha, \beta \in R$ are the parameters of the regression model.

If, as a consequence of a statistical observation, for the two-dimensional random (X,Y) variable, the series of values $(x_i, y_i), i = \overline{1, n}$, has been obtained, then we could write $E(Y|X) = x + \beta \times X,$ and for each pair of values $Y_i = \alpha + \beta \times xi + \varepsilon_{i,}$, where $\varepsilon_i, i = \overline{1, n}$ are "achievements" of ε variable which satisfies the hypotheses, the regression model will write: $E(\varepsilon_i) = 0, V(\varepsilon_i) = \sigma_\varepsilon^2$ and $cov\ (\varepsilon_i, \varepsilon_j) = 0 (\forall) i, j = \overline{1, n}$ and $i \neq j$.

Observing these hypotheses, the regression model will write: $E(y_i|x_i) = \hat{y}_i = \hat{\alpha} + \hat{\beta} \times x_i$, where $\hat{\alpha}, \hat{\beta} \in R$ are the estimators of the two parameters.

More often the estimation of parameters is achieved through the method of the least squares based on determining the line of regression which should minimize the adjusting errors, $ei = yi - \hat{y}$. Thus, the estimators of the two parameters are given by the relations

$$\hat{\beta} = \frac{n\sum_i x_i \times y_i - \sum_i x_i \sum_i y_i}{n\sum_i x_i^2 - \left(\sum_i x^i\right)^2} \text{ and } \hat{\alpha} = \overline{y} - \hat{\beta} \times \overline{x}.$$

Through processing the parameters' values are obtained: $\hat{\alpha}$ =118.235 and $\hat{\beta}$ =15.335.

The parameter we are interested in for this analysis is the one corresponding to the distance between the real estate property and the tip. Therefore, if we were to move 1 km away from the tip, the value of our property would increase on average with EUR 15,335. Mention should be made that the results of the suggested model are conclusive only for certain distances up to the tip. The price difference between a real estate property located at 10 km and another one at 15 km away from the tip will not necessarily be of EUR 45,000. The results can only be conclusive just for the cases in which the environmental effects considered are felt.

In the case of the simple connection based on graphical representation, some hypothesis can be issued as regards the non-linear form of the dependency of price Y on the registered factor X. The testing of these two hypotheses may be achieved, for example, on the basis of the method of the least squares. In case the dependency is appreciated as a second degree parabola, the regression model has the form $Y_x = a_0 + a_1 \times x + a_2 \times x^2 + \varepsilon$. Considering from the test condition $\sum \left(y - \hat{a}_0 - \hat{a}_1 \times x - \hat{a}_2 \times x^2\right)^2 = 0$, the parameters of the function are determined. Through processing we get the values of the parameters $\hat{a}_0$=121.4; $\hat{a}_1$=12.9 and $\hat{a}_2$ =0.32.

In case the dependency is appreciated as a hyperbola, the regression model has the form $Y_x = a_0 + \frac{a_1}{x} + \varepsilon$, and in case of a relation of a logarithmic type we have the form $Y_x = a_0 + a_1 \times lg\, x$; in the case of a relation of an exponential type, the form of the model is $Y_x = a_0 \times a_1^x$.

The values of the parameters of the regression functions are presented in table 2.

The estimated values of the parameters of the regression functions

Table 2

Type of model	$\hat{a}_0$	$\hat{a}_1$
Hyperbolic model	207.76	-78.39
Logarithmic model	128.03	42.7
Exponential model	119.56	0.087

We go forward with the analysis in order to study the hedonistic price through the interaction between the distance of the real estate property up to the tip and the surface around the house, being obvious that the tip's environmental effects jeopardize the environment nearby. In this case, the regression model has the form $Y = a_0 + a_1 \times X_1 + a_2 \times X_2 + \varepsilon$. To determine a_0, a_1 and a_2 the method of the least squares is recommended by minimizing the function

$$\sum_i^n (y_i - a_0 - a_{1i} - a_2 \times x_{2i})^2$$

where $((y_i, x_{1i}, x_{2i}), i = \overline{1,n})$ represents the series of values registered. Though processing, the parameters of the regression function have the following values: a_0=99.5; a_1=13 and a_2=0.05.

So, if we were to leave 1 km away from the tip, the value of the real estate property would rise by an average amount of EUR 13,000 as long as the courtyard surface is attractive.

We include the house surface as an independent variable in the analysis. Generally, the regression model is written under the matrix form $Y = X \times \beta + \varepsilon$, where $Y \in M(T,1), X \in M(T,p), \beta \in M(p,1)$. In order to estimate the parameters, we introduce the following hypotheses:

H1: The exogenous variables are not co-linear. Thus, we can say there are not any non-null real numbers $\lambda_1, \lambda_2, ..., \lambda_p$, for which $\sum_{t=1}^{p} \lambda_i \times x_{it} = 0, t = \overline{1,T}.$ If the variables $X_1, X_2, ... X_p$ are co-linear, then $\lfloor X \times X^t \rfloor = 0$. In this case, matrix X is not reversible and the parameters of the model cannot be estimated.

H2: The random variable e satisfies the hypotheses and $V(\varepsilon) = E(\varepsilon \times \varepsilon^1) = \sigma^2 \times I$. The variables $\varepsilon_t, t = \overline{1,T}$ have the same variance and are non-correlated. We are going to say there is homoscedasticity and the phenomenon of error self-correlation does not manifest itself. Moreover there is considered that e follows a normal T dimensional distribution. If these hypotheses are satisfied, $\hat{Y} = X \times a,$ applying the method of the least squares, we get $a = (X^T \times X)^{-1} X^T \times Y.$ The matrix $(X \times X1)^{-1}$ exists, because $X_1, X_2, ..., X_p$ are linearly independent. Thus,

$$a = (X^T \times X^{-1}) X^T (X \times \beta + \varepsilon) =$$
$$= \beta + (X^T \times X)^{-1} X^T \times \varepsilon$$

The regression model may also have a multiplying form $Y_{\overline{x_1, x_n}} = a_0 \times a_1^{x_1} \times a_2^{x_2} \times ... \times a_n^{x_n}$ in which case through logarithm the additive form can be obtained.

Through processing, based on an additive model, the parameter of the interesting variable is 13.245, its interpretation being similar to the previous case.

The measurement of the direction of the established connection begins through the analysis of co-variation. It is used as a tool indicator in measuring the intensity of the linear connection between the two variables. It is given by the relation

$$cov(x,y)=\frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{p}(x_i-\bar{x})(y_i-\bar{y})\times n_{ij}$$

In case the majority of deviations ($x-\bar{x}$) and ($y-\bar{y}$) are of the same sign (direction), the co-variation is positive and it indicates a direct statistical connection, otherwise the co-variation is negative and it indicates an inverse statistical connection.

The analysis of co-variation has a series of insufficiencies as it is not a normalized indicator for which reason to measure the intensity of the connection, the correlation coefficient (r) is worked out. The bigger $|r|$ is, the stronger the intensity of the connection between the two variables is. If the two variables are independent, then r = 0, the reverse not being generally true. So, if r = 0, it does not implicitly result that the two variables are independent. For a series of the form $\left((x_i y_i), i=\overline{1,n}\right)$ the linear correlation coefficient is determined with the relation

$$r_{y/x}=\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\left[\sum(x_i-\bar{x})^2\right]\left[\sum(y_i-\bar{y})^2\right]}}.$$

The ratio of determination is worked out through comparing the price variation (Y) as a function of the (X) factorial variable modification, with the total variance of the price. Based on the contents of the two types of variances, we can write the relationship of measuring the intensity of connection between the two $R^2=1-\frac{\sum_{i-1}^{n}(y_i-\hat{Y}_{xy})^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}$,

where $\bar{y}$ is the average price and Y_{xy} is the theoretical measure of the price determined on the basis of the regression equality. The indicator has the advantage of being also used in the cases where multiple regressions is resorted to. Extracting the square root, we get the correlation ratio, which is one of the most adequately measuring indicators for the intensity of the statistical connection.

$$R_{y/x}=\sqrt{R^2}=\sqrt{1-\frac{\sum_{i=1}^{n}(y_i-\hat{Y}_{xy})^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

The correlation ratio is analyzed similarly with to the coefficient of correlation. The sign of the correlation relationship is given by the sign of the coefficient of regression (Table 3)

Values of determination and correlation ratio in the case of simple connections

Table 3

Type of model	Value of determination ratio	Value of correlation ratio	Standard error of extimation
Linear model	0.303	0.551	50.377
Parabolic model	0.304	0.551	50.897
Hyperbolic model	0.202	0.449	53.918
Logarithmic model	0.269	0.519	51.602
Exponential model	0.312	0.558	0.281

The simple relations established between the price of the real estate property and its distance from the tip, irrespective of the model used, backed up the initial presupposition, the relationship not being very strong. The exponential model explains the most of the variations in the house prices (31%) depending on their distance from the tip. Actually, this proves to be the best model for estimating the simple connections analyzed, having the smaller estimating error.

In the case of the multiple models (table 4), the issue of measuring the intensity between variables can be approached as follows:

- dependence on all the independent variables;
- partial dependence.

The dependence on all the independent variables:

In order to determine the dependence degree of the price Y upon all the independent variables, we will calculate the multiple determination and correlation ratio based on

$$R^2 = 1 - \frac{\sum_{t=1}^{T}(y_t - \hat{y}\times\overline{x_1, x_2})^2}{\sum_{t=1}^{T}(y_t - \bar{y})^2}$$

Values for the determination and correlation ratio in the case of connections

Table 4

Model type	Values for the determination ratio	Values for the correlation ratio	Standard error of estimation
Bi-factorial model	0.486	0.697	43.739
Tri-factorial model	0.628	0.792	37.62

This implies the building up of the correlation matrix of the multiple model of regression:

$$R = \begin{pmatrix} \Gamma_{y/y} & \Gamma_{y/x_1} \ldots \Gamma_{y/x_p} \\ \Gamma_{x_1/y'} & \Gamma x_{1/x_1} \ldots \Gamma_{x_1/x_p} \\ & \\ \Gamma_{x_p/y} & \Gamma_{x_p/x_1} \ldots \Gamma_{x_p/x_p} \end{pmatrix}$$

The main characteristics of the matrix are that each element is calculated following the formula of the linear coefficient of the variables correlation specified as the case may be and $\Gamma_{ij} \in [-1,1]$.

Partial dependence

Determining the partial correlation ratios, respectively the determinant coefficients for measuring the dependency degree of the resulting variable depending on each of the factorials posted in (table 5):

Values for the relationship of determination and correlation

Table 5

Model type	Dependence on	Values for the partial correlation ratio
Bi-factorial model	distance to the tip	0.551
	land surface around the house	0.523
Tri-factorial model	distance to the tip	0.551
	land surface around the house	0.523
	House surface	0.480

The distance from the real estate property to the tip in interaction with the land surface around the house explains almost half of the price variations for the real estate properties analyzed, each of the two factors directly influencing the price increase with a weighted intensity.

The degree of explaining the increase in the real estate properties prices as we go away from the tip increases considering the house surface. That explains over 60% of the price variations for real estate properties. The rest is accounted for by the other characteristics of the real estate properties: the number of rooms and bathrooms, garages, utilities, the interior arrangements (tiles, parquet, central heating, fireplace, air conditioning) and those in the courtyard (terraces, trees, flowers, swimming pool, alleys). The house surface would have been expected to be the most influencing factor of the real estate price. The values of the partial correlation ratio show that the distance from the tip represents the determining factor in the price variations of the real estate properties analyzed. Thus, the environmental benefits provided by shutting down the tip and neutralizing its effects are measured in monetary terms for each of the individuals who might benefit from the qualitative improvements of the surrounding environment nearby.

References

Droesbeke, J.J., Fichet, B. (1992). *Modeles pour l'analyse des donnes multidimensionnelles*, Editura Economica, Paris

Droesbeke, J.J., Fichet, B., Tassi, P. (1989). *Analyse statistique des donnees de survie*, ASU, Editura Economica, Paris, 1989

Isaic-Maniu, Al., Grădinaru, A., Voineagu, V., Mitruț, C. (1994). *Statistică teoretică și economică*, Editura Tehnică, Chișinău

Isaic-Maniu, Al., Korka, M., Voineagu, V., Mitruț, C. (1998). *Statistică*, Editura Independența Economică, Pitești

Rojanschi, V., Bran, F., Diaconu, G. (2002). *Politici și strategii de mediu*, Editura Economică, București

Vișan, S., Crețu, S., Alpopi, C. (1998). *Mediu înconjurător: poluare și protecție*, Editura Economică, București