3 Hedonic Regressions and Price Indices – Application to the Personal Computer Price Index in Croatia*

Valerija Botrić**

Abstract

Different quality adjustment methods have been used in the consumer price index compilation process in market economies. During the last few years the application of the hedonic regression method has increased markedly. This paper describes the method in short, overviews its use for statistical and other purposes, and discusses the possibility of its implementation in the price index compilation in Croatia. Personal computers were chosen for the empirical exercise. The main reason is that their share in the household consumption is increasing, while another reason lies in the fact that literature suggests the hedonic regression method should initially be applied to the personal computer sub-index.

Key words: hedonic regressions, consumer price indices, quality change
JEL classification: C43, E31

* *This paper was originally published in* Privredna kretanja i ekonomska politika *(Economic Trends and Economic Policy), No. 98, 2004, pp. 30-61.*

** *Valerija Botrić, Assistant, Institute of Economics, Zagreb.*

I would like to thank Dinko Dubravčić, Andrea Mervar, Sandra Švaljek, and an anonymous reviewer for very useful suggestions. Any remaining errors and omissions are solely those of the author.

1 Hedonic Regressions – Concept and Application

Hedonic regressions are regressions that use the prices of a given product (whether good or service) as a dependent variable and various characteristics of that product as independent variables. Hedonic regressions are a reduced form of the economic model, attempting to explain the influence that varying consumer tastes and preferences, different technologies and even companies' product differentiation strategies, may exert on market prices.[1] The main concept forming the basis on which hedonic regressions are applied is the assumption that consumers, when deciding to purchase a product, will check and compare the characteristics of similar products. Hedonic regressions basically represent an analytical method used to determine to what extent an improvement of a certain characteristic may go toward explaining the price difference between two similar products. In that way it is possible to determine what part of the price difference may be due to a quality improvement, and what part is a consequence of other factors. Thanks to this feature the method is used in a number of statistical offices around the world, when monitoring inflation, to calculate the so-called pure price movements.

However, the application of this method goes significantly beyond identifying the impact of quality changes when statistically monitoring price movements. Hedonic regressions may be found in literature, where they are frequently applied in the empirical assessment of the housing quality, as well as to assess the benefits of environmental protection or determine the quality of wine. The key motive to write this paper stems from the fact that the method is currently not used for statistical purposes in Croatia. The main intention of the paper, therefore, apart from presenting an overview of theoretical and empirical literature[2] on the subject, is to show what kind of influence possible application of this method might have on the statistical measurement of inflation in Croatia in the future.

The remainder of this paper is structured as follows: Part 2 discusses microeconomic assumptions that form the basis for application of hedonic

[1] Schultze and Mackie (2002, p. 149).

[2] For a more detailed overview of the manner in which hedonic regressions have been applied by the statistical offices of Finland, France, Canada and United States to assess the differences in the quality of personal computers and other products such as rents, cars and dish-washers, see Čizmić (2003).

regressions. Part 3 contains a brief history of the method's application, as well as an overview of the key issues and problems that have emerged in the course of its application, with a particular emphasis on the application for statistical purposes. Part 4 presents the empirical application of the method on the sample of the Croatian personal computer market, and interpretation of the results. Part 5 concludes.

2 Foundation of the Hedonic Regression Application in Economic Theory

The theoretical foundations of hedonic regressions, as is the case with all econometric models applied in practice, tend to be disregarded. Since theory is extremely important for the resolution of doubts that may occur in the course of its application or interpretation of empiric results, we shall begin with a brief overview of literature written in this field. According to Hulten (2003), there are two fundamental theoretical models that justify the application of hedonic regressions in practice: a model based on demand only, and another model that also includes supply.

a) Demand-based Model

Developed earlier, this model is founded on the microeconomic concept of utility. According to this concept, a consumer's readiness to buy a particular product, i.e. to accept its price, depends on the utility the consumer expects from a set of its characteristics. The term *hedonic* regressions is related to the utility that the consumer enjoys by consuming a product's extra or additional characteristic. An estimation of regression equations reveals consumer preferences for a particular set of useful product characteristics. Diewert (2003) has shown that it is possible to construct a set of conditions under which the hedonic equation may be based on the utility function. This restrictive set of conditions consists of the following:

- consumers share a common utility function, which depends on the product characteristics included in the hedonic function being estimated;
- total consumer utility is equal to the sum of the product utility in the hedonic function being estimated and the utility of other products;

- the consumer need not have a preference system for each individual model of the product on the market, but consumer preferences depend on the product characteristics;
- indifference curves have a negative slope and should be differentiable with respect to the set of product characteristics included in the hedonic function;
- consumers may buy non-negative whole numbers of a particular product only.

This approach includes the market demand aspect only, while completely disregarding supply. The final price of a product, which represents a dependent variable in the regression, depends on both the market supply and demand, meaning that neither aspect may be omitted from the analysis.

b) Supply- and Demand-based Model

Rosen (1974) developed an approach that confronts the consumer demand curve with heterogeneous preferences for the products of different characteristics and corresponding supply functions for each product that has different characteristics. This approach maintains that the hedonic function may be considered as an envelope of different equilibrium states. However, the equilibrium conditions established by such a model have their restrictions as well, and seem to exist solely under the conditions of perfect competition. When other market forms are considered, the model becomes more complicated. The estimated coefficients of desirable product characteristics included in the hedonic function in this model are supposed to be positive. Marginal propensity to buy a product that has a certain characteristic must equal the marginal production cost of that characteristic. In practice, it is often possible for a characteristic to have a negative coefficient. More recent models,[3] based on the same concept that includes both supply and demand, have built onto the Rosen model by making the price of a product characteristic dependent on the marginal cost and market position of the producer, where the market position depends on the elasticity of the demand for a particular product characteristic. This is an attempt at explaining the existence of negative coefficients in the hedonic function and at abandoning the conditions of perfect competition.

[3] *See Pakes (2002).*

Changes in the hedonic function resulting from changes in its components – product price or product quality – are significant for the statistical monitoring of price movements. In the event of inflation (general price level increase, and hence increase of the product price being monitored), production costs of one or several product characteristics increase compared to the reference period. This is reflected in the upward movement of the hedonic curve. Since there has been no parallel change in the product characteristics, consumer preferences will not have changed either. However – unless their income has also increased – consumers will be able to buy a smaller quantity of the same product.

In case the quality of one or more monitored product characteristics varies, the hedonic curve will change for two reasons: firstly, due to a variation in the characteristics of existing products and secondly, due to the appearance of new products. In the first case, the market will see the appearance of the products with improved characteristics, the production of which was previously not cost-effective. In the other, brand new products that did not exist before will appear on the market. These changes in the market occur as a consequence of changes both on the supply and on the demand side. On the supply side, new technological solutions are being introduced into the production process.. On the demand side, there may be changes that stem from demographic flows, shifts in consumer tastes, rise in the living standard. Both types of changes may result either in the shift along the same hedonic function, or in the shift of the entire hedonic function. Changes in the personal computer characteristics as a result of rapid technological development will lead to a shift of the entire hedonic function – downwards, to be specific.

3 History of the Hedonic Regression Method Application – Basic Functional Forms and Issues

The first instance of the hedonic regression method application was recorded back in 1930s.[4] The introduction of the term hedonic regressions is credited to Andrew T. Court, who wrote an article entitled "Hedonic Price Indexes with Automotive

[4] *See for instance, Pakko (2002) or Triplett (2001).*

Examples" in The *Dynamics of Automobile Demand* publication of the General Motors Corporation in 1939.

The application of hedonic regression methods in the statistical monitoring of prices or inflation began as late as 1980s, first of all in the United States. The rationale for their application in statistics is relatively easy to explain. In calculating the consumer price indices most statistical offices apply the Laspeyres formula – using constant quantities for the reference period, while registering price changes between the current and the reference period.[5] This poses the following problem – in order to apply the Laspeyres formula, it is necessary for the product sample used for the index compilation to be constant over the period of compilation. In other words, statistical offices compare *the same with the same* in order to make sure that the index will reflect a *pure* price change, rather than other factors. However, a rapid technological development or improvement of the products that make up the consumer basket of the reference population has made it increasingly difficult to identify similar products – products of similar characteristics – within the period that the index refers to. Therefore statistical offices began to resort to hedonic regressions in order to determine the extent to which new product characteristics, that were unavailable to consumers previously, influence the price difference. Naturally, initially the application was experimental only – in 1995, hedonic regressions were applied to no more than 0.2 percent of the products that constitute the sample used for price monitoring in the United States, which leads the way in the method's application.[6]

The greatest spur to the application of hedonic regression in the compilation of the consumer price index came from the Boskin Report.[7] It was a report produced in 1996 by an expert commission appointed by the U.S. Congress and led by M.

---

[5] *Different terminology related to the compilation of the consumer price index is a cause of major differences among the countries, methodologies and experts in this field. The International Labor Organization (ILO) has made a certain effort to unify the terminology used by different countries in their methodology, at least as far as the consumer price index compilation is concerned. As part of revising a manual for the compilation of the consumer price index it has proposed a harmonization of terminology. Details may be found on http://www.ilo.org/public/english/bureau/stat/download/gloss.pdf. That source also recommends the introduction of the terms* index reference period, weight reference period and price reference period *to replace the terms* base index, base weight period and base price period *that were used previously. This paper follows such recommendations.*

[6] *See Hulten (2003).*

[7] *Hulten (2003).*

Boskin. The main task of the commission was to determine a consumer price index bias, stemming from inappropriate treatment of the product quality issue when two products in the sample are switched. The commission found that the consumer price index, calculated according to the method used to date, was 0.55 percentage points higher than it would be if the *pure* price index calculation was applied. The Boskin Report estimated the total consumer price index bias at 1.1 percentage points. It meant that all other influences put together (such as the index calculation formula or representativeness of the product basket) contribute almost equally to the total index bias as the product quality differentiation. The findings of those studies led to an increase in the application of the quality differentiation method in the compilation of the consumer price index. According to Landefeld and Grimm,[8] a hedonic regression index is applied to approximately 18 percent of final consumption in the compilation of the U.S. Gross Domestic Product deflator. That figure points to a relatively large influence of the method on the statistical monitoring of economic indicators, while also testifying to a steep increase in its application over a relatively short period. Although the method has not been applied in Croatia to date, this or other[9] quality differentiation adjustment methods can be expected to come into use shortly as part of the statistical price monitoring. Hence, it is essential to get potential data users acquainted with the main characteristics of the method in a timely manner, so that at the time of the method's introduction they may be aware of the advantages and drawbacks of the statistical data they use. Following is a brief overview of the main functional forms used to estimate hedonic regressions.

As mentioned already, the product price is a dependent variable in hedonic regressions, while the other side of the equation contains product characteristics. The main functional forms used in practice are as follows.[10]

---

[8] *Cited in Benkard and Bajari (2003).*

[9] *Some other quality differentiation adjustment methods that are used in compiling the consumer price index include: estimation of additional option costs, production cost method, linking method, class mean imputation method, overlap method.*

[10] *Division according to Brachinger (2003). For a derivation of a functional form of the specific hedonic function from the underlying model assumptions, see Diewert (2003).*

a) Linear Model

In a linear model, a dependent variable is expressed as a linear combination of independent variables, or:

(1) $$p = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k$$

where:
p... product price
x_i... an i-th product characteristic
$\beta_i = \frac{\partial p}{\partial x_i}$ hedonic price of x_i characteristic

The hedonic price of a product characteristic, i.e. a coefficient in the estimated linear regression equation in this model, represents a marginal change in the product price in the event of development or improvement of the i-th product characteristic. Diewert (2003) says this functional form, while often applied in practice, should in fact not be applied since it is not derived from the theoretical model it is supposed to be based on.

b) Exponential Model

In the exponential model, a dependent variable is expressed as a product of multiplication of the exponential values of independent variables, or put into an equation:

(2) $$p = \beta_0 \prod_{k=1}^{K} e^{\beta_k x_k}$$

This model may be transformed into a linear one by converting the equation into logarithmic form:

(3) $$\ln p = \ln \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k$$

When applying that model, as well as in all other cases when a transformed original equation is estimated, one should pay attention to the interpretation of results. In that case, hedonic prices may be presented as follows:

(4) $$\frac{\partial p}{\partial x_i} = \beta_i p$$

The assessed coefficient appearing alongside the product characteristic x_i is therefore to be interpreted as the rate of price growth. In practice, this form of a hedonic function is often called semi-logarithmic since the dependent variable in the linear function has been turned into a logarithm, while independent variables have not.

c) Double-logarithmic Model

The hedonic function in this case has been specified in a way that the dependent variable equals a product of multiplication of the independent variables, raised to the power of accompanying coefficients, or expressed by means of an equation:

(5) $$p = \beta_0 \prod_{k=1}^{K} x_k^{\beta_k}$$

Applying a logarithm to equation (5) yields the following expression:

(6) $$\ln p = \ln \beta_0 + \beta_1 \ln x_1 + \beta_2 \ln x_2 + \ldots + \beta_k \ln x_k$$

Hedonic prices of product characteristics in this model may be calculated from the following equation:

(7) $$\frac{\partial p}{\partial x_i} = \frac{\beta_i}{x_i} p$$

Estimated coefficients appearing alongside independent variables may be interpreted as partial elasticities. In other words, that coefficient denotes the likely product price increase in percentage terms at a point of observation, provided that the i-th characteristic of that product changes by 1 percent.

d) Logarithmic Model

In this model the hedonic function is specified in the following manner:

(8) $$p = \beta_0 + \beta_1 \ln x_1 + \beta_2 \ln x_2 + \ldots + \beta_k \ln x_k$$

and hedonic prices of product characteristics are expressed as follows:

$$(9) \qquad \frac{\partial p}{\partial x_i} = \frac{\beta_i}{x_i}$$

The biggest problem in the application of hedonic regressions is determining a set of product characteristics that have an impact on the price of a given product. Theory says that it is necessary to choose a set of product characteristics that influence the consumer's decision to buy the product and the manufacturer's decision to produce it. However, to what extent any set of characteristics may reflect the wishes of both the manufacturers (suppliers) and the consumers, is questionable. Buyers of the products that undergo relatively rapid technological development, such as personal computers, very often do not take into account all the differences in the technology of two products but rather base their purchase decision on the basis of completely different characteristics – product design, for instance.[11] Such characteristics, hardly observable and generally excluded from regression equations, may often influence the bias of estimated coefficients. Furthermore, since not all consumers share the same preferences and tastes may change over time, these estimated coefficients may display considerable instability, particularly over a longer period.

Regardless of the selected form of the hedonic function, the method itself may be applied for the purpose of imputing prices or for index imputation. When imputing a price, the estimated coefficients are used for adjusting the price of a product either in the price reference period or in the current period. An adjusted product price is in that case comparable for both periods that a price index is compiled for. Alternatively, estimated coefficients in the hedonic function may be applied directly in the form of an estimated sub-index for the product that undergoes quality difference estimation. Basically, there are three different approaches to the application of hedonic regressions.

[11] *Similar examples may be found in the products where technological development is not so fast but probably faster than average. In cars the product color may have equal importance as some other technical characteristics that, for instance, increase driving safety.*

3.1 Application in Data Adjustment

An adjustment of the data that has been collected takes place when price analysts substitute products in the sample used for the compilation of the consumer price index with similar, although not identical products. Estimated coefficients of a regression equation are then applied to adjust the product prices, and to make two different products comparable.

This method does not enable a direct index calculation, but rather imputes the estimated quality adjustment of the product price. The regression equation coefficients serve to estimate how much extra consumers are prepared to pay for the quality difference that probably results from a certain product characteristic. The product price adjustment for the purpose of calculating the price of a comparable product is carried out by subtracting, from the price recorded in the market, an estimated value resulting from a new additional characteristic[12] of the product in question. The index calculation is performed subsequently according to a standard procedure, based on a set of adjusted prices and those recorded originally and requiring no quality change adjustment.

From the point of view of the statistical office, this method has an advantage in that no continual data gathering is necessary for the estimation of a regression equation. That is to say, rather than using the current month's data, it is possible to use the data collected during a previous month. Furthermore, once estimated, the coefficients may be used during several months to come.[13] However, this fact also gives rise to the main objection to this method. Namely, if the data does not refer to a recent period, it is impossible to maintain that the estimated coefficients really reflect the consumer willingness to pay for a certain additional unit of some product characteristic in the current period, compared to the price reference period. That is precisely why it is recommended to apply another method, one that will impose the use of adjusted data.

[12] *When a statistical office performs an adjustment of the reference period prices, the estimated value of a product characteristic is added to the reference period price.*

[13] *Triplett (2001) believes it is necessary to estimate the equation on the basis of new or current data at least 3 times a year in order to make the adjustment more credible.*

3.2 Hedonic Regressions Using a Binary Variable

Binary variables are applied primarily to denote a time period in hedonic regressions, although it is also possible for them to denote an existing product characteristic. To put it in concrete terms, when the data referring to several time periods are used in the equation estimation, it is possible to add a binary variable to each time period. An index of the price change of the selected product is then derived directly from the estimated coefficient. In the semi-logarithmic or double-logarithmic model the index is, according to Triplett (2001), calculated as a ratio of geometric price means in the two periods. Since the statistical offices have recently begun using the geometric mean to calculate the average product price values at lower aggregation levels, this approach is considered to be adjusted significantly to other methodological segments of the consumer price index compilation.

This method's advantages are largely the same as the deficiencies of the previous method, and vice versa. An additional advantage lies in the fact that no data matching is required before a regression equation is applied, i.e. the regression may be applied to a whole set of data rather than just the data used in the index compilation. This enables additional information from a large set of data to exert a positive influence on the quality of the total index obtained in the process. On the other hand, if an equation estimated data from several time periods simultaneously, adding data for a new period would change the estimated coefficients in previous periods, and that in turn would require a previously published price index to be revised. That is why some statistical offices avoid applying this method.[14]

---

[14] *See Ball and Allen (2003).*

3.3 Application in the Compilation of Superlative Indices[15]

This approach, founded most of all on the economic theory and on the index number theory, is applied to construct a superlative index.[16] Seeing that the price index, usually calculated according to the Laspeyres formula, does not correspond to an ideal[17] price index, applying the quality adjustment method need not necessarily lead to ideal theoretical solutions. Restrictions resulting from the need to possess the data on value weights for each product represent the main flaw of the application of hedonic regressions when attempting to construct superlative indices. In the consumer price index, value weights represent data showing the purchase value of each product – where the value is a product of multiplication of the quantity and the price – bought by the reference population. While some authors[18] are researching the conditions under which quantity weights may be applied to substitute value weights, in practice the method may only be applied if the so-called scanner data[19] is available. The appropriateness of scanner data is still the subject of research, and it is also difficult to ensure continuous collection of such large databases in practice.

3.4 Problems with the Application of Hedonic Regressions for Statistical Purposes

Although it is evident that the method of comparing same-quality products in different periods, applied by the statistical offices, is inappropriate for the monitoring of pure price changes, it is necessary to emphasize that the statistical offices are also developing other quality change estimation methods. The main objection to a rapid expansion in the application of the hedonic regression

---

[15] *Superlative indices are those that, according to the index number theory, meet all the characteristics. For more on desirable characteristics of price indices, see Diewert (1999).*

[16] *More details on the problem of linking hedonic regressions and the compilation of ideal indices are available in Diewert (2003).*

[17] *Where an ideal index is defined as the index with all the desirable characteristics specified in the index number theory.*

[18] *See, e.g. Okamoto (2003).*

[19] *Scanner data are those collected directly from outlets. Apart from the product name, characteristics and price, they include the number or value of transactions conducted at that particular outlet.*

method for statistical purposes stems precisely from the way it continues to be applied indiscriminately, particularly on large numbers of products. Identifying the product characteristics that influence the product price is hard work. Collecting appropriate data may put additional strain on and exhaust a statistical office. That set of characteristics need not be constant over time either and, if some important characteristics are not included in the regression equation, estimated coefficients may be biased. Overall, the cost and benefit ratio would not always suggest that hedonic regression methods ought to be introduced in the statistical monitoring of prices.

Another reason, emphasized particularly by Hulten (2003), lies in the old saying that "an old tax is a good tax". Applied to statistical data, it would go something like "data subject to the usual methodology is good data". Distrust of the statistical data, regardless of the economic or statistical system development level, is common to all countries. In transitional countries that distrust is amplified by the fact that a large number of statistical indicators may change over a short period, and that prevents comparison between longer series of statistical indicators – in the pre-transitional and transitional period. The introduction of previously unused methods may prompt additional suspicion among users about the credibility of statistical data. However, as pointed out by von der Lippe (2001), statistical offices in socialist countries also applied a kind of product quality adjustment. The main difference was that the socialist-era adjustment focused on the final result, or the total index. The “method” consisted, mainly, in the selection of sample products whose quality was less likely to change over a longer period. When a product still needed to be substituted, the choice of tactics was dictated by the final objective, i.e. lower inflation rate. The introduction of new methods such as the hedonic regression method, by means of which data collected on the ground is changed, may be experienced by the users as a return to the era when statistical data used to be manipulated.

Part 4 discusses the possibility of applying the hedonic regression method on the personal computer price data collected in Croatia. It will present a method based on the adjustment of the collected data, and a method based on the application of binary variables.

4 Applicability in Croatia – Research Case of the Personal Computer Market

Statistical offices most frequently apply hedonic regressions to carry out quality adjustments of the products that are subject to rapid technological advancement. At the time this paper was written, personal computer prices were not included in the retail price index or in the cost of living index published by the Central Bureau of Statistics. At the same time, results of a research carried out by the GfK Center for Market Research[20] in February 2002 showed 29 percent of the households having a personal computer, with that percentage on a constant increase. Hence, it is logical to expect that the statistical monitoring of prices will soon include the prices of personal computers, paving the way for the application of hedonic regressions. Schultze and Mackie (2002, p. 129) recommend that statistical offices begin applying hedonic regressions for the estimation of quality differences precisely on personal computers and to expand their application gradually to other products, in the event the method proves successful.

Data on the personal computer price movements on the Croatian market for this paper were collected by monitoring the websites of local computer dealers (companies). The monitored data included the prices of desktop personal computers only, while those on notebook computers were not taken into account. The companies in the sample were selected at random, the main criterion for their inclusion being the availability of a pricelist, updated regularly through the period of observation, on their websites. The sample included the following companies: Analogbit, Formel, HGspot, InfoGama, Infocor, Infoplus (PC Lab), King Računala, IBB Computer Shop and Ve-mil. Prices were first collected on November 19, 2002 and then again in December 2003. The week of 14-19 that month was chosen, and actual monitoring was done during workdays. The procedure tried to imitate the process of price monitoring at a statistical office. To illustrate the problems faced by a statistical office, the total sample of companies' pricelists collected in November 2002 was analyzed, and 57 desktop computers with available price data and various characteristics were selected. The same procedure was repeated using the 2003 data. In December 2003, these same companies had 54 computers on their website pricelists.

[20] *Results can be found on: www.gfk.hr/press/hrnet.htm.*

The basic method used by a statistical office when calculating the price index is to monitor the market price of a selected product. Therefore, the sample constructed this way served to analyze how many of the products that were on the market in November 2002 could still be found late in 2003. Matching was based solely on the main computer characteristics – computer processor speed and memory capacity. If their configuration included difference in components that required heavy expenditure – for instance, the configuration included a monitor in one period while in the other it did not – then the comparison was reduced to a comparable configuration by subtracting the published price of that extra component. However, if the differences referred to cheaper components, there was no reduction to common characteristics. No more than 10 personal computers out of the total sample, according to their main characteristics, could be considered to be comparable in both periods. This leads to the conclusion that the statistical office would have had to substitute as much as 83 percent of the original sample in a year. Results are presented in Table 1.

Table 1. **Comparison of some personal computer characteristics in the Croatian market, November 2002 – December 2003.**

	Nov. 2002	Dec. 2003
Average price, HRK	7885.62	4327.72
Average memory capacity, MB	206.60	263.11
Average processor speed, GHz	1.64	2.15
Average hard disk capacity, GB	32.97	60.19
Average monitor size, inch	16.22	17.00
Share of CDROM-equipped PCs, %	70.18	27.78
Share of DVD-equipped PCs, %	22.81	37.04
Share of CDRW-equipped PCs, %	10.53	55.56
Number of observations	57	54

Source: Author's calculation.

Based on the comparison of the computers in this sample, one may observe a substantial decline in the average prices of personal computers. Simultaneously, a substantial improvement of all the main characteristics was also recorded. Such movements are typical of the personal computer markets in all countries.

The most frequently applied technique in the statistical monitoring of prices is finding the product with the same characteristics. It requires a search in the market for a comparable product, or a product that will display the same characteristics in two consecutive periods. The application of this method on personal computers might mean that a price collector would very soon be monitoring a product which is either no longer available in the market, or may not be representative. In other words, the statistics may monitor a product that would actually be purchased by very few buyers. The problem of representativeness is present particularly in the products that undergo frequent technological changes. Considering that manufacturers compete not only in quality (new technologies) but also in prices, new product models are very often cheaper or come at the same price as the existing products, while displaying some quality improvements. This makes existing products less competitive and even if they remain on the market, their market share will shrink until it becomes negligible. In the goods that undergo rapid technological advancement, in most cases the price of the products that have lost their competitive edge will not be allowed to fall, they will simply be withdrawn from the market. Still, it means that the prices included in the index monitored by the statistical office will stagnate, although the quality adjusted average prices of those products will actually fall.

The chosen set of computer characteristics presented in Table 1 is fairly standard in literature when comparing computer similarities. Based on the research conducted so far in other countries (or groups of countries), those computer characteristics proved to be the most significant for explaining price movements. Thus, for instance, Benkard and Bajari (2003) used a total of 26 characteristics in their research, including: binary variable for the operating system, CPU type, RAM capacity, hard disk (HDD) capacity, SCSI, CDROM, DVD, modem speed, binary variable for the monitor, monitor size, zip drive, etc. Ball and Allen (2003) analyzed the notebook market separately from the desktop computer market. They used the following set of characteristics for the desktop computers: processor type and speed, memory type and capacity, hard disk, monitor size, DVD, CD-RW, DVD-RW, Combo, graphic card type, audio card type, operating system and others (printer, scanner, digital camera, etc.). The characteristics used in notebook computers do not differ significantly from those listed above.

What follows is a description of the hedonic method's application on the set of characteristics described above. The possibility of applying two methods – one based on data for a single one-month period, and the other using data for both periods of observation – was analyzed. Firstly, we shall explain the application of the method based on data for a single period.

A) Regression Equation Estimated on the Basis of Data for a Single Period

For the sake of comparison, the regression equation was estimated for both periods included in Table 1. All the functional forms were estimated, and results are presented in Table 2.

While reference literature suggests the application of the double-logarithmic model as the most appropriate in hedonic regressions when it comes to personal computers, other forms are often used too. For instance, the exponential form is applied in the research by Allen and Ball (2003), while Brachinger (2003) believes it is worth testing several functional forms before deciding which one fits the collected data best. This part of the paper presents the estimated regression equations for all four functional models described above.

The fact that the equations estimated here contained a relatively small number of characteristics should be explained. All the functional forms estimated the equation consisting of four main product characteristics – HDD capacity, CPU speed, RAM capacity and monitor size. Although data on other characteristics was also available in the sample (see Table 1), a comparison with the research results in other countries[21] showed it would be better to focus on the characteristics that had the largest impact on the formation of personal computer prices in other studies. The inclusion of a greater number of explanatory variables would require a very large database for the sample.

---

[21] *Also, see a research conducted by Konjin, Moch and Dalen (2003). In their study, which compares hedonic regressions on personal computer prices in Germany, Great Britain and France, CPU speed, HDD capacity and RAM capacity proved to be the best explanatory variables in all these countries.*

Table 2. **Regression estimate, OLS method, single period**

Period	Model	Dependent variable	Independent variables					R^2	Sample size
			RAM	HDD	Speed	Monitor	Constant		
Nov. 2002	Linear	price	8.15 (0.96)	126.73** (2.17)	3398.41 (1.39)	-1550.50 (-1.63)	21920.78 (1.68)	0.33	46
			13.28* (1.97)	88.40* (1.96)	1275.86 (0.82)		-76.28 (-0.03)	0.31	56
			14.58** (2.05)	95.47** (2.06)			1726.29 (1.26)	0.29	57
	Exponential	ln(price)	0.00 (1.15)	0.01** (2.67)	0.40 (1.61)	-0.13 (-1.36)	9.58*** (7.17)	0.43	46
				0.02*** (5.09)	0.43* (1.73)	-0.15 (-1.57)	9.88*** (7.52)	0.43	46
				0.02*** (4.84)	0.25 (1.54)		7.75*** (29.43)	0.38	56
	Double logarithmic	ln(price)	0.17 (0.66)	0.71** (2.70)	0.48 (1.32)	-1.83 (-1.21)	10.39 (2.38)	0.48	46
			0.22 (1.03)	0.59*** (2.75)	0.27 (1.09)		5.51*** (7.89)	0.42	56
				0.76*** (5.58)	0.28 (1.15)		6.07*** (13.95)	0.43	56
	Logarithmic	price	1254.86 (0.48)	5996.60** (2.30)	4146.23 (1.15)	-22894.61 (-1.53)	43291.06 (1.01)	0.39	46
				7048.91*** (4.96)	4394.64 (1.24)	-24314.93 (-1.67)	50088.61 (1.24)	0.40	46
				6171.22*** (4.70)	1224.86 (0.52)		-13761.16*** (-3.26)	0.32	56

Table 2. **continued**

Period	Model	Dependent variable	Independent variables					R^2	Sample size
			RAM	HDD	Speed	Monitor	Constant		
Dec. 2003	Linear	price	8.28* (2.11)	43.45*** (3.24)	1482.77** (2.51)	303.48 (1.60)	-7903.75** (-2.39)	0.91	18
			9.04*** (3.60)	28.28** (2.52)	854.19 (1.36)		-1623.52 (-1.61)	0.70	37
			9.53*** (4.51)	27.02*** (3.45)			192.79 (0.43)	0.66	54
	Exponential	Ln(price)	0.00 (0.79)	0.01** (2.72)	0.34** (2.73)	0.01 (0.15)	7.04*** (10.21)	0.84	18
			0.00*** (2.78)	0.01*** (2.75)	0.22* (1.91)		7.15*** (38.90)	0.70	37
			0.00*** (3.59)	0.01*** (4.63)			7.48*** (87.73)	0.67	54
	Double logarithmic	Ln(price)	-0.04 (-0.15)	0.66*** (3.93)	0.81*** (3.26)	0.41 (0.61)	4.36* (1.83)	0.83	18
				0.64*** (5.27)	0.79*** (4.12)	0.44 (0.72)	4.13** (2.32)	0.84	18
				0.56*** (4.10)	0.52** (2.14)		5.69*** (12.41)	0.61	37
	Logarithmic	price	1433.88 (1.05)	3762.10*** (4.12)	3633.04** (2.69)	7434.44* (2.03)	-41219.32*** (-3.18)	0.88	18
				4390.55*** (6.36)	4485.36*** (4.16)	6179.11* (1.78)	-32611.37*** (-3.23)	0.88	18
				3006.93*** (3.71)	2532.56* (1.75)		-9408.86*** (-3.47)	0.55	37

Source: Author's calculation.

*Notes: Independent variables: RAM – memory capacity; HDD – hard disk capacity; Speed – computer processor speed; Monitor – monitor size. Independent variables may appear in logarithmic form in certain models (see Chapter 3 for details). Coefficients marked *** are significant at a level of 1%, ** at a level of 5%, * at a level of 10%, while t-values are presented in brackets below the regression coefficients.*

All the regression forms were first run on an equation that included all four computer characteristics, while the two remaining equations were obtained by eliminating the independent variables that proved less significant. All these equations included a constant. Even though the application of some functional forms on such a data set might produce satisfactory regression function estimates without intercept, it is not logical. That is because we cannot be absolutely certain that a different set of product characteristics, excluded from the equation, would not have a greater impact on price movements. Hence, there is no justification for the attempts at explaining price variations by developments in the selected product characteristics only.

After the product characteristics whose estimated coefficients proved insignificant were eliminated, HDD capacity and CPU speed emerged as the most significant pair of characteristics in 62.5 percent of the cases, while the remaining 37.5 percent of cases highlighted RAM capacity and HDD capacity as the most significant.

We can use an example to illustrate the application of estimated regression equations in the compilation of the price index. An overview of the estimated equations for 2002 shows the best results were obtained by running the double-logarithmic form. At the same time, all the functional forms estimated on the basis of 2003 data fitted the data in that concrete sample better than 2002 data. One may also observe that the coefficients obtained in the estimated equations for the two periods are not robust, and that is the main argument used against possible application of the method in the compilation of the consumer price index. Namely, the application of the estimated equations based on the reference period and the current period data may obtain quite different estimates for the quality differences. Statistical offices have dealt with that problem by applying an equation that has been estimated on several time periods. In order to illustrate the differences that may result, in the following section we shall apply two estimated equations – a double-logarithmic model based on 2002 data and the same functional form estimated on the basis of 2003 data. In both cases the equation using HDD capacity and CPU speed as significant characteristics.

On the basis of coefficients from the regression equation estimated in this manner it is necessary to perform an adjustment of the prices that were actually collected and refer to the products that were previously found to be comparable. This can be done in two ways – by adjusting current and reference period prices. The reference period prices were adjusted by applying an estimated equation that refers to November 2002 data. The adjustment of the current period prices was performed according to an estimated equation that refers to December 2003 data. The following table presents the indices calculated by adjusting the prices for the quality difference according to the estimated equation, and without the quality adjustment. The indices were calculated by applying a non-weighted arithmetic mean.

Table 3. **Indices calculated without and with quality adjustment**

No.		Index $\frac{XII\,2003}{XI\,2002}$
1	All sample prices	54.8
2	Matched products	34.5
3	Reference period adjusted acc. to the 2002 equation	62.3
4	Current period adjusted acc. to the 2003 equation	56.7
5	Reference period adjusted acc. to the 2002 equation, matched products	105.6
6	Current period adjusted acc. to the 2003 equation, matched products	52.0

Source: Author's calculation.

Table 3 contains the results of six possible methods of calculating the index. The first index was calculated as a ratio of average prices of all products in the sample. The application of an index calculated in such a manner is relatively rare in practice since the statistical offices usually choose a representative product for the purpose of its price monitoring, and only substitute that product for a new one if it disappears from the market. In that case the prices may or may not be adjusted for the difference in quality between the two products.

The next case identifies the products that can be found in both periods of observation. The second index (No. 2 in Table 3) refers to their price change. One may observe that the price decline identified in that case is even larger than it would be if the index was calculated from all the collected data. Taking into account that statistical offices sometimes estimate the quality differences by assuming that the prices of unmatched products move identically to those of matched products, it is evident that the application of that method in this case would result in a relatively steeper price decline than without estimating the quality price differences.

The next two indices (No. 3 and 4 in Table 3) were obtained by adjusting the prices of the entire sample in one period of observation, before calculating the index from the entire sample in both periods. More concretely, in the first of these two cases (index value 62.3) reference period prices were adjusted, and the index calculated as a ratio of the average unadjusted current period price and the average adjusted reference period price. In the second case (index value 56.7) current period prices were adjusted, and the index calculated as a ratio of the average adjusted current period price and the average unadjusted reference period price. Both indices show a smaller price decline than if the total price index is observed. Part of the explanation may be found in a relatively narrow sample used as the basis for the index calculation. It was actually even smaller since in 2003, no data on CPU speed was available in as many as 17 products. The average quality adjusted price was therefore also calculated on a considerably smaller sample.

In the last two cases (No. 5 and 6 in Table 3) the very same method was applied to matched products. In view of the relatively small number of matched products, the difference resulting from whether the reference or the current period is applied grows even larger. However, large index differences before and after the quality adjustment may be expected even when a larger set of data is available. White, Abel, Berndt and Monroe (2004) applied the hedonic regression method for the compilation of the price index of the computer operating systems. Using a sample of 340 products, they found that the average price fell 1.16 percent in the 1987-2000 period, matching obtained a price decline of 7.02 percent, while prices were found to have fallen 16.2 percent by applying the hedonic regression method.

B) Application of Binary Variables

The following step analyzed the application of various forms of the hedonic function by using binary time variables. As mentioned above, the main difference with respect to the previous example is that the regression equation here is estimated by using both data sets (for the reference and current price period), with a binary time variable inserted in the equation to denote the data belonging to either of the two periods of observation. Results of the analysis in which the binary variable is associated to the current period is presented in Table 4.

Table 4. **Application of binary time variables in OLS hedonic regressions**

Model	Dependent variable	Independent variables						R^2	Sample size
		RAM	HDD	Speed	Monitor	Binary	Constant		
Linear	Price	11.23* (1.70)	80.15** (2.18)	902.29 (0.65)	-411.27 (-0.77)	-5869.28*** (-4.60)	8264.79 (1.02)	0.36	64
		12.13* (1.89)	83.49** (2.31)		-248.38 (-0.53)	-5809.69*** (-4.59)	6792.76 (0.88)	0.37	64
		14.68*** (3.76)	35.38* (1.96)			-5350.67*** (-7.08)	3685.17*** (4.83)	0.41	111
Exponential	ln(price)	0.00* (1.89)	0.01*** (2.88)	0.22 (1.52)	-0.04 (-0.80)	-0.79*** (-6.03)	8.55*** (10.27)	0.50	64
		0.00*** (3.42)	0.01*** (3.12)	0.20* (1.83)		-0.89*** (-8.99)	7.87*** (48.08)	0.54	93
		0.00*** (3.94)	0.01*** (4.15)			-0.82*** (-9.94)	8.17*** (98.08)	0.56	111
Double logarithmic	ln(price)	0.18 (0.91)	0.70*** (3.64)	0.37* (1.68)	-0.83 (-0.95)	-0.89 (-7.02)	7.65*** (2.91)	0.60	64
			0.84*** (7.37)	0.42* (1.96)	-1.05 (-1.25)	-0.89*** (-7.00)	8.72*** (3.71)	0.56	64
			0.71*** (7.16)	0.33* (1.84)		-0.98 (-9.74)	6.23*** (20.19)	0.54	93
Logarithmic	Price	1845.78 (0.94)	5327.00*** (2.86)	1542.10 (0.71)	-7261.96 (-0.86)	-6838.18*** (-5.51)	86.42 (0.00)	0.44	64
		2183.71 (1.15)	5339.68*** (2.88)		-4097.10 (-0.57)	-6815.29*** (-5.52)	-9826.35 (-0.46)	0.44	64
		3172.41*** (2.76)	2278.52** (2.31)			-5763.36*** (-7.64)	-16371.91*** (-3.96)	0.42	111

Source: Author's calculation.

*Notes: Independent variables: RAM – memory capacity; HDD – hard disk capacity; Speed – computer processor speed; Monitor – monitor size. Independent variables may be in logarithmic form in certain models (see Chapter 3 for details). Coefficients marked *** are significant at a level of 1%, ** at a level of 5%, * at a level of 10%, while t-values are presented in brackets below the regression coefficients.*

Of all the equations estimated here, the equation estimated according to the exponential functional form is the most suitable for further use. All the variables included in the last version of that form were substantially different from zero at a 1% level of significance, and introducing variations in the independent variables may explain 58 percent of the total variations in prices. One detail needs to be pointed out – the estimated coefficients appearing next to product characteristics (HDD and RAM) are very low. This is due to the fact that coefficients are interpreted as growth rates.

To obtain a value for the quality adjusted price change, it is necessary to calculate the index according to the estimated coefficient with a binary variable. The -0.82 estimated coefficient gives an index value of 44.[22] That index value is lower than any of the adjusted indices presented in Table 3 (listed under 3-6).

This paper discusses the hedonic regression method by testing its applicability for statistical purposes. However, as in all other empirical studies, a wider interpretation of the results is also worth considering. Naturally, the question is to what extent the results obtained here reflect actual price movements on the Croatian computer market, and whether a wider result interpretation may be subject to some restrictions. First of all, the very method of product sampling makes it impossible to claim that the sample is representative. However, since there is no alternative – data on the total personal computer population for sale, it is impossible to monitor the entire population or identify a sample that would be representative. Identifying a representative sample of companies that sell these products is also problematic. Hence, the research results showing a change in price movements or in the quality of personal computers available on the Croatian market may be indicative only.

Another question that arises concerns is product homogeneity. The hedonic regression method is often applied in practice to a relatively extensive set of data on a product whose homogeneity is questionable. Theory has it that, in order for a method's application to be meaningful, it is necessary to estimate the function for a specific type of products with truly similar characteristics, or else it would not be clear what the coefficients obtained from the estimated regression equations

[22] *exp(-0.82)*100.*

actually refer to. In the empirical studies, due to problems with the availability of data, hedonic functions are often estimated on an extended set of data. In the application of hedonic regressions to the consumer price index, the homogeneity presents less of a problem if tight product specifications are applied.[23]

Without conducting additional research into the sales of certain computer types on the Croatian market it is difficult to draw more concrete conclusions, besides a very general one that the market has seen a decline in the average price and a quality increase. It is also necessary once again to draw attention to the fact that the estimated equations were based on data that included various computer configurations offered by dealers. The general impression is that computers are more often sold in separate components than preconfigured. Nevertheless, in any concrete case of statistical monitoring a "typical" computer configuration should be chosen by the statistical office, reducing the problem of comparing the products of different quality in two periods of observation to the case described here. However, for the purpose of market research it is possible to analyze all the products available in the market and apply the hedonic regression method to estimate the valuation of each particular computer characteristic by the buyers. To enable that kind of analysis, it is necessary to collect the sales data for each product, or to have so-called scanner data available.

5 Conclusions

The application of hedonic regression methods for statistical purposes has begun relatively recently and even though it is becoming increasingly present in practice, at economic conferences and in research papers, one still cannot claim that all has been said about its advantages and deficiencies. The fact that the method's application is founded on the economic models with restrictive sets of

[23] *Product specifications monitored for the consumer price index compilation may be tight and loose. Tight specifications describe the product to be monitored more closely so, apart from a detailed list of product characteristics, such a specification may also include a product brand. For instance, a tight specification of a washing machine may be manufacturer Gorenje, model AX, 1800 rpm, loading capacity etc. With loose specifications, the number of product characteristics is set by the statistical office while a price collector may choose a concrete product at the sales outlet. It does not mean that in doing so, the price collector does not need to record other product characteristics but only that the choice of them may be wider, depending on what products are available at that particular outlet.*

assumptions should be sufficient to spur further search for the models that would be less so.

Until recently, the price monitoring statistics in Croatia used no method to estimate the quality differences, including the hedonic regression method. However, the liberalization of the local product market has brought about major changes in the household spending structure and a shift toward the products that experience rapid technological development. Due to this fact statistical offices in other countries are also facing the need to introduce some changes in the statistical monitoring of price movements. An additional incentive for the application of the hedonic regression method in statistical price monitoring in the EU member countries and transition countries that are candidates for accession stems from the need to compile a harmonized consumer price index. Although Eurostat does not explicitly require the application of the hedonic regression method, a number of research papers as well as studies into the application experience in various countries and conferences devoted to that topic in the past few years serve as an indication that the method might be applied at least experimentally in all EU member states. On the other hand, it needs to be emphasized that the compilation of a harmonized consumer price index continues to be based largely on recommendations, while the only methods to be banned are the ones considered to be inappropriate. The level of harmonization is not such to impose the application of concrete methods, but rather provides the statistical offices of individual countries with a set of methods that are believed able to produce comparable results. Therefore, the Croatian statistical system has plenty of time to decide which methods are feasible and most reliable.

The research performed as part of this paper demonstrated that the hedonic method's application to estimate the quality price differences is subjective. While the statistical offices can be expected to have larger databases available for the application of hedonic regressions than was the case in this research, the actual choice of the functional form of an equation, product characteristics and the model itself could still lead to substantial differences in the total price index. A decision on the most appropriate model lies largely with the price statistician performing the regression. Hence, it is recommended to introduce new methods gradually and test their applicability very carefully. If this recommendation is heeded, a gradual introduction of targeted products will have a smaller impact on the total consumer price index.

Literature

Ball, Adrian and Andrew Allen, 2003, "The Introduction of Hedonic Regression Techniques for the Quality Adjustment of Computing Equipment in the Producer Prices Index (PPI) and Harmonised Index of Consumer Prices (HICP) ", *Economic Trends*, 592, London: ONS.

Benkard, C. Lanier and Patrick Bajari, 2003, "Hedonic Price Indexes with Unobserver Product Characteristics, and Application to PC's", NBER Working Paper, 9980, Cambridge, Mass.: National Bureau of Economic Research.

Brachinger, Hans Wolfgang, 2003, "True Hedonic Price Indices: Concepts and Estimation Problems", University of Fribourg, Faculty of Economics, presented at the "National and International Monetary Policy under New Conditions" conference organized by Verein für Socialpolitik, Zürich, Sept 30-Oct 3, 2003.

Čizmić, Draženka, 2003, "Primjena hedonističke regresije u kvalitativnom prilagođivanju indeksa potrošačkih cijena", *Ekonomija*, 10(3), pp. 493-516.

Diewert, Erwin, 1999, "The Consumer Price Index and Index Number Purpose", paper presented at the "Joint ECE/ILO Meeting on Consumer Price Index" conference organized by ILO and UNECE, Geneva, Nov 3-5.

Diewert, Erwin, 2003, "Hedonic Regressions: A Consumer Theory Approach", in R.C. Feenstra and M. D. Shapiro, ed., *Scanner Data and Price Indexes*, NBER and University of Chicago Press: Studies in Income and Wealth, 64, pp. 317-348.

Hulten, Charles R., 2003, "Price Hedonics: A Critical Review", *Economic Policy Review*, September, Federal Reserve Bank of New York, 9(3), pp.5-15.

Konijn, Paul, Dietmar Moch and Jörgen Dalén, 2003, "Comparison of Hedonic Functions for PCs across EU countries", paper presented at the "54th ISI Session" conference, organized by the Irving Fisher Committee, Berlin, Aug 13-20.

Okamoto, Masato, 2003, "Comparison of Hedonic Indices Compiled using Different Types of Weights", National Statistics Center Japan, Research Department, presented at the "7th Ottawa Group Meeting on Price Indices" conference organized by INSEE, Paris, May 27-29.

Pakes, Ariel, 2002, "A Reconsideration of Hedonic Price Indices with an Application to PCs", NBER Working Paper, 8715, Cambridge, Mass.: National Bureau of Economic Research.

Pakko, Michael R., 2002, "Comparing Apples and Oranges", *The Regional Economist*, October, The Federal Reserve Bank of St. Louis.

Rosen, Sherwin, 1974, "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition", *Journal of Political Economy*, 82(1), pp. 34-55.

Schultze, Charles L. and Christopher Mackie, 2002, ed., *At What Price? Conceptualizing and Measuring Cost-of-Living and Price Indexes*, Washington, DC: National Academy Press.

Triplett, Jack E., 2001, "Hedonic Indexes and Statistical Agencies, Revisited", *Journal of Economic and Social Measurement*, 27(3-4), pp. 131-153.

Turvey, Ralph et al., 1998, *Indeksi potrošačkih cijena: priručnik MOR-a*, Zagreb: Državni zavod za statistiku.

Van der Lippe, Peter, 2001, "Some Conservative Comments on Hedonic Methods", University of Essen, paper presented at the "Symposium on Hedonic Methods", organized by Deutches Bundesbank and German Federal Statistical Office, Wiesbaden, June.

White, Alan G., Jaison R. Abel, Ernst R. Berndt and Cory W. Monroe, 2004, "Hedonic Price Indexes for Personal Computer Operating Systems and Productivity Suites", NBER Working Paper, 10427, Cambridge, Mass.: National Bureau of Economic Research.