http://communications.science.ankara.edu.tr

Commun.Fac.Sci.Univ.Ank.Ser. A1 Math. Stat.
Volume 72, Number 2, Pages 551–562 (2023)
DOI:10.31801/cfsuasmas.1187009
ISSN 1303-5991 E-ISSN 2618-6470

Research Article; Received: October 10, 2022; Accepted: March 29, 2023

A STUDY ON USING ROBUST HEDONIC REGRESSION IMPLEMENTATION

Serdar Cihat GÖREN[1] and Olcay ARSLAN[2]

[1,2]Department of Statistics, Ankara University, 06100 Ankara, TÜRKİYE

Abstract. This article aims to determine the features affecting the price of a product with the hedonic regression model and to estimate the contribution of each feature to the price by using robust regression estimation methods. For the analysis, the price and feature information of the laptop product group were obtained from the big data source by using the web scraping method. Four alternatives of the hedonic regression model are used to determine the features affecting the price of the laptops. The contribution of each feature to the laptop price is estimated by using the robust (Huber M-estimator) estimation method and the Ordinary Least Squares (OLS) estimation method, and the resulting estimates are compared for both methods. In the framework of the data set used in the study, it is observed that the effective model is the Logarithmic Robust Hedonic Regression Model.

1. Introduction

Statistics producers aim to provide quality and accurate statistics to their users. As in many studies, the production of statistics requires resources in terms of money, human, and time. The data collection method with a questionnaire is one of the traditional and most widely used methods. In addition, data collection methods such as administrative data and big data are also widely used. Collecting data with administrative data and big data reduces the requirements considerably. They also reduce the burden on the respondents that occurs with the survey.

The market share of laptop computers in the technology sector is quite high. There are many brands of laptop manufacturers, which creates a serious competition environment in the market. Laptop manufacturers determine customer profiles, produce computers with features suitable for these customer profiles, and offer them

2020 *Mathematics Subject Classification.* 62J05, 62J07.

Keywords. Robust regression, hedonic regression, big data.

[1] scgoren@gmail.com-Corresponding author; 0000-0002-6253-6156

[2] oarslan@ankara.edu.tr; 0000-0002-7067-4997.

for sale. On the other hand, consumers aim to buy a laptop computer that is suitable for their use and budget. Campaigns, promotions, advertisements, and more than one brand model in technology stores make consumers' decisions complicated. This competitive environment also increases the risk of consumers purchasing laptops with the wrong choice. At this point, it will be useful for consumers to research computer brands, features, and prices in detail.

When consumers want to search for laptop computer prices and models, they prefer to look at technology market websites. However, they may not always have the necessary information about which computer is suitable for their use and what the right price should be for this computer. This article aims to determine the features that affect laptop computer prices by using big data from the internet prices of technological markets.

One of the methods used to determine the value of a good by breaking it down into its components is the hedonic regression method. The word "hedonic" means pleasure, satisfaction, or benefit after the consumption of goods and services (Bulut and Zaman [4]).

According to the study by Colwell and Dilmore [5], it is stated that the first user of hedonic regression was Haas [12]. Haas [12] estimated the price of agricultural land with the hedonic regression model using the variables of distance from the city center and the size of the city center.

According to the study by Sheppard [22], it is stated that Waugh [27] was the first study to measure the effect of quality on the price of products. On the other hand, It is stated that Court [6] was the first to use the term "hedonic" to characterize heterogeneous goods and determine demands for individual preferences.

In the literature, it has been seen that price estimation with hedonic regression has a wide usage area. In recent years, the hedonic regression model was frequently used to determine the effect of housing prices and housing characteristics on the price. In addition, analyses were conducted using the hedonic regression model for different products other than housing.

Diewert et al. [8]-[9], Hülagu et al. [17], Jiang et al. [18], and Selim S. [21], used the hedonic regression model for house prices to determine the features affecting house prices. They analysed feature contributes by using the Ordinary Least Squares (OLS) method.

Fixler et al. [11] used the hedonic regression model as a quality adjustment method in the US Consumer Price Indices.

Manoel et al. [19] determined the features of laptops with the hedonic regression model, and measured the contribution of each feature to the price with the OLS method.

McCormack [20] identified the features that affect the price for new cars by using the hedonic regression model, and measured the contribution of each feature to the price with the OLS method.

In many studies conducted with hedonic regression, estimations were generally carried out by using the OLS method. However, it is known that the OLS method does not give effective results if there are outliers in the data. In this case, robust statistical methods should be preferred to estimate the parameters. Since robust statistical methods are robust against outliers.

Bourassa et al. [2] used the robust hedonic regression model for house prices to determine the features affecting house prices. The efficiency of 3 robust statistical methods and the OLS method were compared. As a result of the study, it was observed that robust methods give more effective results in the case of outliers than the OLS method.

Bulut and Zaman [4] used the robust hedonic regression model for the car of "Beetle as Turtle" models to determine the features affecting the car. The efficiency of OLS, M-estimators (Huber, Tukey, Hampel), Mutli-stage Method (MM), Least Trimmed Squares (LTS), Least Median Square (LMS), and Least Absolute Deviations (LAD) methods were compared. It was observed that the LAD method gave effective results.

In this article, four alternatives for the hedonic regression model were used to determine the features affecting the price of the laptops. The big data is obtained from technological markets for laptops by using the web scraping method. The parameters of each model were estimated by using OLS and robust methods. It was analysed how much the features contributed to the laptop prices. Robust methods and the OLS method are compared in order to find and recommend the effective method.

This article is organized as follows. In Section 2, laptop computers, big data, hedonic regression, and robust statistical methods are introduced. In Section 3, the hedonic regression model alternatives are established within the framework of the obtained data set and the estimation results are given. In the last section, the findings obtained as a result of the study are evaluated.

2. Method

2.1. **Laptop Computers.** Nowadays, laptop computers are one of the most demanded and used technological products. The most important advantage of laptop computers is portability. Like many technological products, laptops also have many technical features. They are also called hardware features in market conditions. Hardware features are processor, ram, hard disk, graphic card, screen size, etc. While some of these features directly affect the laptop price, some of them have a very limited effect on the price. Statistical methods can be used to learn this distinction and to determine how much each feature contributes to the price of the laptop. As a result of the hedonic regression analysis carried out in this article and the main features affecting the laptop computer price were highlighted.

2.2. **Big Data.** The demand and need for statistical information are increasing day by day. Data collection methods such as surveys, administrative data, and big data have been developed for the collection of data. Big data has become very popular nowadays. Because statistics producers prefer to use existing data instead of collecting data from the field to achieve their goals. Thus, statistics producers provide significant benefits in terms of human, financial, and time resources.

Although big data has many definitions, it is generally defined as data that is too large for traditional users to store and process. Examples of big data are transportation, social media, and market data. It can not be said that big data is big only because of the volume of data (Doğan and Arslantekin [10]). To define the data as big data, some requirements must be met. These requirements are defined in the literature with the 5V (Volume, Velocity, Variety, Value, Veracity) approach. Here, "volume" refers to the volume and size of the data, "velocity" refers to the speed of the data, "variety" refers to the diversity of the data, "value" refers to the value of the data, and "veracity" refers to the validity of the data.

According to an article by Aktan [1]; it was stated that the concept of big data was used for the first time in the study of Cox and Ellsworth [7]. In this study, scientific data visualization study was carried out.

2.3. **Web Scraping.** Web scraping is the set of techniques used to automatically get some information from a website instead of manually copying it. The goal of a web scraper is to look for certain kinds of information, extract, and aggregate it into new Web pages. In particular, scrapers are focused on transforming unstructured data and save them in structured databases (Vargiu and Urru [26]).

In this article, data were collected from several technology websites for a certain period by web scraping method. Since the structure of the websites was different, the Spider code was used in Python specific to each site. Selenium was used as click technology, and Scrapy libraries were used as data collection technology. Since it was only used for a certain period and in a way not to increase the data traffic of the websites, no block was encountered by the websites.

2.4. **Hedonic Regression.** Hedonic regression is a method used in order to determine the value of a good or service by breaking it down into its components. The value of each component is determined separately by regression analysis (McCormack [20]). The hedonic regression model has the completely same structure as the classical linear regression. Since the concept of hedonic is based on consumer satisfaction, it is identified with the analysis of the relationship between the prices and properties of goods.

Hedonic price functions are used for two main purposes: to create general price indices that take into account changes in the quality of manufactured goods and to analyse consumer demands for the characteristics of heterogeneous goods (Sheppard [22]). The second of these aims that Sheppard [22] has defined is one of the aims of our article. Measuring the contribution of laptop features to the price with the hedonic regression model will show us how the consumer demand for the features of this product.

By using the hedonic price model, a relationship is established between the features of a good and the price of the relevant good. In other words, the hedonic price model is a method that evaluates the price of a particular good as the sum of the values of the features it has and estimates the value of each feature using regression analysis (Shimizu et al. [23]).

$$p = h(c_i) \tag{1}$$

The hedonic price model is defined at (1). In this model, p is called the price of a good, and $h(c_i)$ is the hedonic function of the properties of that good. A hedonic function is estimated by regression analysis (McCormack [20]).

2.4.1. *Hedonic regression model alternatives.* When the hedonic regression studies in the literature are investigated, it is shown that hedonic regression model alternatives vary. The purpose of hedonic price modelling is to determine the model that will identfy the functional relationship between price and features. In the literature, we come across 4 different model structures which are Linear Model (LinLin), Logarithmic Model (LogLog), Linear Logarithmic Model (LinLog), and Logarithmic Linear Model (LogLin).

Model alternatives vary according to the structure of dependent and independent variables. The linear and logarithmic structure of the dependent and independent variables allows the hedonic regression model to be diversified and the efficiency of parameter estimations to be compared. Established hedonic regression model types show different results according to the price and features of the estimated product.

In the hedonic regression model, the dependent variable is the price of the product, while the independent variables are the features of the product. Some of the independent variables consisting of the features of the product can be quantitative and some of them can be qualitative variables. While quantitative variables can be directly included in the model, qualitative variables should be defined as dummy variables. In this case, if the qualitative variables defined as dummy variables are in the relevant product as a feature, it takes the value of 1, if not, it takes the value 0.

In this study we define the variables;
Y (dependent variable): Price of the laptop
X_1 (independent variable): Processor speed of the laptop
X_2 (independent variable): Ram size of the laptop
X_3 (independent variable): Hard disk size of the laptop
X_4 (independent variable): Processor (Intel i7) of the laptop
X_5 (independent variable): Processor (Intel i5) of the laptop
X_6 (independent variable): Processor (Intel i3) of the laptop
X_7 (independent variable): Graphic Card of the laptop

Four model structures are given at (2), (3), (4), and (5). Here, $Y_i \in R$ the price of the features is the price of the product to be estimated. In other words, it is defined as the response variable in the regression model. $\alpha \in R$ is the constant term of the model. $\beta_i \in R$ are unknown parameters. ε_i are independent and $E(\varepsilon_i) = 0$, $Var(\varepsilon_i) = \sigma^2$

- Linear Model (LinLin)

$$Y_i = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} + \beta_6 X_{i6} + \beta_7 X_{i7} + \varepsilon_i \quad (2)$$

- Logarithmic Model (LogLog)

$$lnY_i = \alpha + \beta_1 lnX_{i1} + \beta_2 lnX_{i2} + \beta_3 lnX_{i3} + \beta_4 lnX_{i4} + \beta_5 lnX_{i5} + \beta_6 lnX_{i6} + \beta_7 lnX_{i7} + \varepsilon_i \quad (3)$$

- Linear Logarithmic Model (LinLog)

$$Y_i = \alpha + \beta_1 lnX_{i1} + \beta_2 lnX_{i2} + \beta_3 lnX_{i3} + \beta_4 lnX_{i4} + \beta_5 lnX_{i5} + \beta_6 lnX_{i6} + \beta_7 lnX_{i7} + \varepsilon_i \quad (4)$$

- Logarithmic Linear Model (LogLin)

$$lnY_i = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} + \beta_6 X_{i6} + \beta_7 X_{i7} + \varepsilon_i \quad (5)$$

McCormack (2013) used the Logarithmic Linear (LogLin) hedonic regression model in order to determine the features that affect new car prices.

Bulut and Zaman (2018) analysed the factors affecting the price with (2), (3), (4), and (5). As a result of the study, it was stated that the most effective results were obtained with the Logarithmic Linear Model (LogLin).

Selim (2008) determined the house characteristics by using hedonic regression model. The Logarithmic Linear Model (LogLin) was used in the study and the estimations were made with the OLS method.

Jiang *et al.* (2014) analysed housing features by using the OLS method with establishing a Linear Model (LinLin) hedonic regression model.

This article aims to settle model alternatives and decide which model is effective according to the model standard error criterion. In the implementation phase of the study, the parameter estimation of the four alternative hedonic regression models was carried out by OLS and robust (Huber M-estimator) estimators. The effective model is highlighted from the estimation results obtained.

2.4.2. *Ordinary Least Square (OLS).* The Ordinary Least Square (OLS) method is widely used in regression analyses. Stigler [24] mentioned that the first users of OLS were Carl Friedrich Gauss and Adrien Marie Legendre, but which one was the first user is a matter of debate. It was stated that while Legendre had a publication on the subject in 1805, Gauss had a publication in 1809. However, it has been concluded that there is serious evidence that Gauss used the OLS method for the first time in 1795 and that the inventor of the OLS was Gauss. In addition, Stigler [24] defined the OLS method was the automobile of modern statistics, and the person who first discovered the OLS method was identified as the inventor of the automobile, Henry Ford.

The linear regression model can be written with the help of matrices at (6).

$$Y_i = X_i^T \beta + \varepsilon_i \tag{6}$$

for $\mathrm{Y}_i \in R, X_i \in R^T, \beta \in R^T, i = 1, 2..., n$.

The OLS estimators for the regression parameter vector are obtained at (7):

$$\widehat{\beta} = (XX')^{-1} X'Y \tag{7}$$

In the literature, as in linear regression, parameter estimations in the hedonic regression model are generally carried out by the OLS method. Although OLS is one of the most common statistical methods, it does not give effective results in cases there are outliers in the data. In case of outliers in the data, robust statistical methods give more effective results than the OLS method.

Conventional hedonic regression models are highly sensitive to these outliers because they are estimated by minimizing the sums of the squared residuals. This gives outliers, and particularly large outliers, disproportionately large influence. Even a small number of outliers can have a large effect. Robust methods generally down-weight observations automatically based on the size of their residuals. (Bourassa et al. [2]).

In this article, parameter estimations in the hedonic regression model were carried out with OLS and robust methods and their effectiveness was compared with each other. In the next section, robust methods are introduced.

2.4.3. *Robust Statistical Methods.* In case of outliers in the data set, it is recommended to use robust methods that are less affected by outliers compared to the OLS method when estimating parameters. As mentioned in the previous section, robust statistical methods, which reduce the effect of outliers, make the estimation more reliable.

Problematic sales prices, such as sudden discounts and foreclosure sales, seriously reduce the price. Some types of problematic transactions may be flagged in some hedonic data sets. In other cases, it may be possible to identify these transactions, but only with considerable investment of time and effort. Robust methods provide a means for responding to data problems when it is difficult or impossible to identify all of the transactions with contaminated data. (Bourassa et al. [2]).

The term robust was introduced to the statistical literature by Box [3]. The field of modern robust statistics emerged with the pioneering work of Tukey [25], Huber [15], and Hampel [13] and has been extensively developed over time (Heritier et al. [14]). In this article, parameter estimations in hedonic regression are carried out via Huber's M-estimator.

2.4.3.1. *Huber M-estimator*

The widely used M estimation method for robust regression was introduced by Huber [15]-[16]. This class of estimators has been accepted as a generalization of the maximum likelihood estimation. The M estimator method is designed to minimize an objective function that increases less rapidly than the OLS objective function.

In contrast to OLS, M-estimators minimize some function that gives decreasing weights to observations as the size of the standardized residual increases (Huber [15]).

$$\frac{1}{n}\sum_{i=1}^{n}\rho\left(Y_i - X_i^T\beta\right) \tag{8}$$

Huber M-estimation function is given at (8). Here, ρ is non-negative and non-decreasing. M estimators aim to minimize (8). If ρ is differentiable, the 1st derivative concerning β is taken and set to 0, and the following equation is obtained.

$$\frac{1}{n}\sum_{i=1}^{n}\psi\left(Y_i - X_i^T\beta\right)X_i = 0 \tag{9}$$

where psi $=\rho'$.

(9) is solved by iterative methods and $\widehat{\beta}$ is obtained.

3. Implementation

In this article, technological market data was used in order to determine the model that gives effective estimations about laptop prices. The data set was obtained by web scraping method for the period covering the last quarter of 2020. Since web scraping method was only used for a certain period and in a way not to increase the data traffic of the websites, no block was encountered by the websites. The size of data set is about 5 thousand rows.

In the data set, it was decided to use the processor, processor speed, ram, hard disk, and graphic card features that directly affect the price of a laptop computer in the hedonic regression model. In the hedonic regression models established, the dependent variable is price, and the independent variables are processor, processor speed, ram, hard disk, and graphic card. At this point, price, processor speed, ram, and hard disk variables are defined as numerical variables, while other variables are defined as categorical (no exist (1,0)) dummy variables. The definition and properties of the variables used in the study are given in the Table 1.

3.1. **Results of Analysis.** Hedonic regression model analyses were performed in the R Package. In the regression analysis, 4 model structures were analysed, namely Linear Model (LinLin), Logarithmic Model (LogLog), Linear Logarithmic Model (LinLog), and Logarithmic Linear Model (LogLin). Coefficient estimates were carried out for each model structure with both OLS and robust M-estimators. Four different hedonic regression models were applied to the data set. They are "OLS-Logarithmic Model (O-LogLog)", "Robust (M-estimator)- Logarithmic Model (R-LogLog)", "OLS-Linear Model (O-LogLin)" and "Robust (M-estimator) - Logarithmic Linear Model (R-LogLin)". The coefficient estimates obtained for each model are given in Table 2. It is seen how much the features that affect the price of the laptop contribute to the price of the laptop. In addition, model standard errors

TABLE 1. Definitions of variables.

Variable	Name	Type	Measurement
Y	Price	Numeric	Turkish Lira
X_1	Processor speed	Numeric	Gigahertz (GHz)
X_2	Ram	Numeric	Gigabyte (Gb)
X_3	Hard disk	Numeric	Gigabyte (Gb)
X_4	Processor- Intel i7	Dummy	i7=1, other=0
X_5	Processor- Intel i5	Dummy	i5=1, other=0
X_6	Processor- Intel i3	Dummy	i3=1, other=0
X_7	Graphic card	Dummy	EXT =1, INT =0

obtained for each model are given in Table 3. According to Table 3, it is observed that the model with the lowest standard error is the "Robust Logarithmic Model (R-LogLog)".

TABLE 2. Estimations of coefficients.

Variable	O-LogLog	R-LogLog	O-LogLin	R-LogLin
μ	$6,91$	$6,90$	$8,32$	$8,33$
X_1	$0,06$	$0,05$	$0,02$	$0,03$
X_2	$0,48$	$0,45$	$0,04$	$0,04$
X_3	$0,15$	$0,16$	$0,00$	$0,00$
X_4	$0,48$	$0,46$	$0,65$	$0,62$
X_5	$0,24$	$0,23$	$0,32$	$0,30$
X_6	$0,04$	$0,00$	$-0,09$	$-0,12$
X_7	$0,01$	$0,02$	$-0,02$	$-0,02$

TABLE 3. Model standard errors.

Model	Standard errors
O-LogLog	$0,14$
R-LogLog	$0,11$
O-LogLin	$0,15$
R-LogLin	$0,15$

3.2. **Rankings of Contributions.** The coefficient estimates are obtained by using the hedonic regression models to rank the contribution of laptop features to the price for each model are shown in Table 4. When the coefficient estimates are examined, we see that the i7 processor made the biggest contribution. It is known that Intel's processors, i3, i5, and i7, are in the form of $i7 > i5 > i3$ in terms of performance.

The fact that the contribution rankings of the processors that contribute to the laptop price in all 4 models are observed as $i7 > i5 > i3$ also shows that the estimates coincide with reality.

TABLE 4. Rankings of contributions.

Variable	O-LogLog	R-LogLog	O-LogLin	R-LogLin	Average Rank
i7	1	1	1	1	1
Ram	2	2	3	3	$2,5$
i5	3	3	2	2	$2,5$
Hard disk	4	4	5	5	$4,5$
Processor speed	5	5	4	4	$4,5$
Graphic card	7	6	6	7	$6,5$
i3	6	7	7	6	$6,5$

3.3. **Case of outliers in the data.** In data sets, errors may occur due to data entry, system, and unit of measurement. These errors are sometimes difficult to detect in large data sets. In addition, there is a possibility that they may be overlooked. To address this situation, few and large amounts of incorrect data entries were added to the obtained data set. In Table 5, there are model standard errors obtained from 4 models in case of a few (Model standard error-2) or a large amount of incorrect data entry (Model standard error-3) into the data set. In addition, the model standard errors before the incorrect data entry are included in the table as "Model standard error-1" to be able to compare.

TABLE 5. Case of outliers in the data.

	O-LogLog	R-LogLog	O-LogLin	R-LogLin
Model standard error-1	$0,14$	$0,11$	$0,15$	$0,14$
Model standard error-2	$0,21$	$0,12$	$0,23$	$0,15$
Model standard error-3	$0,68$	$0,14$	$0,66$	$0,18$

According to the Table 5, it is seen that all model standard errors increase in case of incorrect data entries. In both cases (few and large amounts of incorrect data), R-LogLog appears to give the lowest standard error. It is observed that both models (R-LogLog and R-LogLin) obtained with a robust estimator from 4 models give lower standard errors than models estimated with OLS (E-LogLog and O-LogLin). It is seen that the model standard errors of robust estimators are considerably lower than the model standard errors of OLS, especially, in case of a large amount of incorrect data entry. In addition, the fact that the model standard errors of robust estimators do not increase much in the data set despite a few or a large amount of incorrect data entries shows that these estimators can achieve effective estimations despite incorrect entries in the data.

4. Conclusion

The article focuses on price prediction with hedonic regression models. An implementation study was carried out on the technological market data obtained for the last quarter of 2020 for laptop computers by the web scraping method. Within the scope of the implementation study, the contribution of each of the features affecting the price of the laptop to the price was predicted with four different hedonic regression models. As a result of the analysis, the efficiency of the four models was compared and it was observed that the Robust- Logarithmic Model (R-LogLog) gave the most effective estimations. According to the results obtained with this model, it has been concluded that the processor (i7, i5, i3), processor speed, ram, hard disk, and video card have an increasing effect on the price of the laptop. In addition, it was seen that the i7 processor made the most contribution to the price of the laptop.

In the outlier and residual analysis, it was determined that there was no problem in the data set. However, in case of potential errors in such data sets, which model would be more effective was also examined. It has been observed that the Robust-Logarithmic Model (R-LogLog) is the most effective estimator of the four models in case of errors in the data set by entering the data set with incorrect data.

Within the scope of the data set used in this article, the price estimation of each of the features affecting the price of the laptop and its contribution to the price was measured with the R-LogLog model. Robust methods, which are robust against outliers, are recommended to be used in such studies.

Author Contribution Statements The authors contributed equally to this paper. All authors read and approved the final copy of this paper.

Declaration of Competing Interests The authors declare that they have no competing interest.

References

[1] Aktan, E., Büyük veri: uygulama alanları, analitiği ve güvenlik boyutu, *Bilgi Yönetimi Dergisi*, (2018). https://doi.org/10.33721/by.403010

[2] Bourassa, S. C., Cantoni, E., Hoesli, M., Robust hedonic price indexes, *International Journal of Housing Markets and Analysis*, (2013), 47-65. https://doi.org/10.1108/IJHMA-11-2014-0050

[3] Box, G. E. P., Non-normality and tests on variances, *Biometrika*, 40 (1953), 318-335. https://doi.org/10.2307/2333350

[4] Bulut, H., Zaman, T., Robust hedonik modellerin karşılaştırılması: Beetle Türkiye piyasa fiyatını etkileyen faktörlerin incelenmesi, *C.Ü. İktisadi ve İdari Bilimler Dergisi*, 19 (2018), 24-43.

[5] Colwell, P. F., Dilmore, G., Who was first? An examination of an early hedonic study, *Land Economics*, 75 (1999), 620-626. https://doi.org/10.2307/3147070

[6] Court, A. T., Hedonic price indexes. The dynamics of automobile demand, *General Motors Corporation*, 1939.

[7] Cox, M., Ellsworth, D., Application-controlled demand paging for out-of-core visualization, *8th IEEE Visualization'97 Conference*, 1997. https://doi.org/10.1109/VISUAL.1997.663888
[8] Diewert, W. E., Haan, J., Hendriks, R, The decomposition of a house price index into land and structures components: A hedonic regression approach, *Economic Measurement Group Workshop*, 2010.
[9] Diewert, W. E., Haan, J., Hendriks, R., Hedonic regressions and the decomposition of a house price index into land and structure components, *Statistics Netherlands*, 2014.
[10] Doğan, K., Arslantekin, S., Big data: Its importance, structure and current status, *DTCF Journal*, 56 (2016), 15-36. https://doi.org/10.1501/Dtcfder-0000001461
[11] Fixler, D., Fortuna, C., Greenlees, J., Lane, W., The use of hedonic regressions to handle quality change: The experience in the U.S. CPI, *U.S. Bureau of Labor Statistic*, 1999.
[12] Haas, R., Transportation conditions in Europe, *The Annals of the American Academy of Political and Social Science*, 104 (1922), 157-167.
[13] Hampel, F. R., A general qualitative definition of robustness, *Ann. Math. Stat.*, 42 (1971), 1179-1186.
[14] Heritier, S., Cantoni, E., Copt, S., Victoria-Feser, M. P., Robust methods in biostatistics, *Journal of Biopharmaceutical Statistics*, 20 (2010), 482-484. https://doi.org/10.1002/bimj.201000194
[15] Huber, P. J., Robust estimation of a location parameter, *Annals of Mathematical Statistics*, 35 (1964), 73-101. https://doi.org/10.1214/aoms/1177703732
[16] Huber, P. J., Robust regression: Asymptotics, conjectures and Monte Carlo, *The Annals of Statistics*, 5 (1973), 799–821. https://doi.org/10.1214/aos/1176342503
[17] Hülagu, T., Kızılkaya, E., Gencay A., A hedonic house price index for Turkey, *Central Bank of the Republic of Turkey*, 2016.
[18] Jiang, L., Peter, C., Phillips, B., Jun, Y., A new hedonic regression for real estate prices applied to the Singapore residential market, *Cowles Foundation For Research In Economics Yale University*, 2014. https://doi.org/10.2139/ssrn.2533017
[19] Manoel, N., Felisoni, C., Roberto, M., A five-year hedonic price breakdown for desktop personal computer attributes in Brazil, *Brazilian Administration Review*, 6 (2009), 173-186. https://doi.org/10.1590/S1807-76922009000300002
[20] McCormack, K., Diagnostics for hedonic models using an example for cars, *Biometrics and Biostatistics International Journal*, 2 (2015), 23-37. https://doi.org/10.15406/bbij.2015.02.00022
[21] Selim, S., Determinants of house prices in Turkey: A hedonic regression model, *Journal of Dogus University*, 9 (2008), 65-76. https://doi.org/10.31671/dogus.2019.223
[22] Sheppard, S., Hedonic analysis of housing markets, Oberlin College, Ohio, USA, 1997.
[23] Shimizu, C., Takatsuji, H., Ono, H., Structural and temporal changes in the housing market and hedonic housing price indices: A case of the previously owned condominium market in the Tokyo metropolitan area, *International Journal of Housing Markets and Analysis*, 3 (2010), 351-368.
[24] Stigler, M. S., Gauss and the invention of least squares, *The Annals of Statistics*, 9 (1981), 465-474.
[25] Tukey, J. W., A survey of sampling from contaminated distributions. In: Oklin, I., Ed., Contributions to Probability and Statistics, Stanford University Press, 1960.
[26] Vargiu, E., Urru, M., Exploiting web scraping in a collaborative filtering based approach to web advertising, *Artificial Intelligence Research*, 2 (2013), 44-54. https://doi.org/10.5430/air.v2n1p44
[27] Waugh, F. V., Quality as a Determinant of Vegetable Prices, Columbia University Press, New York, 1929. https://doi.org/10.7312/waug92494