

# Modeling and Forecasting Dynamic Factors of Pricing in E-commerce

Galyna Chornous, Yaroslava Horbunova

*Taras Shevchenko National University of Kyiv, Faculty of Economics, Department of Economic Cybernetics, Vasylkivska str. 90a, Kyiv, 03022, Ukraine*

## Abstract

The rapid development of information technologies contribute to the search for various methods to increase the e-commerce company's profits. Big Data technologies are conducive to developing varieties of personalized pricing strategies, in particular dynamic pricing meaning a process of setting and developing prices for goods and services, where prices change synchronously for all consumers depending on demand under current market conditions. The implementation of dynamic pricing approaches is directly related to the use of appropriate software capable of performing data mining. The important tasks are to study the successful experience of powerful e-commerce platforms in relation to the tools that they used for analysis, and to expand the toolkit of methods to support personalized pricing strategies with methods working with Big Data. The purpose of this research is to study the experience of modeling dynamic pricing and the proposal to involve the PLS regression both to confirm the importance of dynamic factors and to predict prices in the implementation of this strategy of personalized pricing in e-commerce. The approaches to pricing in e-commerce are studied in the research. The features of different types of personalized pricing (dynamic, customized, transactional) are identified too. The approaches to modeling of dynamic pricing are explored. Applying PLS regression for predicting the impact of dynamic factors on price is proposed. The study places special emphasis on identifying the dynamic factors that have become increasingly important in recent years. Results of modeling confirm that linear regression is not able to identify hidden predictors among a large number of collinear factors (dynamic factors are often hidden and collinear) whereas PLS regression does. The proposed approach is implemented by using data of the vacation rental e-platform Airbnb.

## Keywords

dynamic pricing, e-commerce, dynamic pricing modeling, dynamic factors of pricing, PLS regression, Big Data, information economy

## 1. Introduction

The rapid development of technology, as well as the growth of the information sector of the economy, contributes to the search for various methods to increase the company's profits or reduce its costs. A promising area of the information economy is e-commerce. Interest in the e-commerce market is increasing every year. There are objective reasons for this fact: the growing number of Internet audiences, the growing share of online sales in total sales, the development of social and mobile networks. The COVID-19 pandemic has also been a factor in the significant growth of this sector of the economy.

One of the most widely used modern types of pricing in e-commerce is personalized or customized pricing meaning dynamic price adjustment for consumers depending on the value that these customers

---

*IT&I-2020 Information Technology and Interactions, December 02–03, 2020, KNU Taras Shevchenko, Kyiv, Ukraine*

EMAIL: gach.2012@gmail.com (G.Chornous); yaroslava.gorbunova@gmail.com (Y.Horbunova)

ORCID: 0000-0003-4889-1247 (G. Chornous)

© 2020 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



CEUR Workshop Proceedings (CEUR-WS.org)

attribute to the commodities. Personalized pricing is a technology-based pricing system in which different prices are set for the same goods for different consumers. The development of Big Data technologies has contributed to the spread of personalized pricing strategy, as it provided companies with the opportunity to analyze in real-time many factors on the Internet environment, such as customer loyalty, purchase history on the site, consumer preferences, and more.

One of the types of personalized pricing is dynamic pricing. This type of pricing bases on a process of setting and developing prices for goods and services, where prices change synchronously for all consumers depending on demand under current market conditions.

The number of e-commerce retailers has grown over the past few years. With increasing competition, they face the difficult task of maximizing profits while maintaining price competitiveness. Dynamic pricing is a rational and effective solution to this problem because it takes into account changes in supply and demand and recommends an effective pricing structure. If this pricing strategy is implemented over a long period, it can significantly increase the overall revenue and profitability of the business.

The implementation of dynamic pricing approaches is directly related to the use of appropriate software capable of performing data mining. In recent years, arrays of information on goods, consumers, and market conditions have become increasingly large. Therefore, on the one hand, the task is to study the successful experience of powerful e-commerce platforms in relation to the tools that they used for analysis, and on the other hand - to expand the toolkit of methods to support personalized pricing strategies with methods working with Big Data.

The purpose of this research is to study the experience of modeling dynamic pricing and the proposal to involve the PLS model both to confirm the importance of dynamic factors and to predict prices in the implementation of this strategy of personalized pricing in e-commerce.

The outlined goal requires the solution of the following research tasks:

- Investigate approaches to pricing in e-commerce and show the prospects of personalized pricing.
- Identify the features of dynamic pricing and explore approaches to its modeling.
- Propose prospective models for predicting the impact of dynamic factors on price.
- Develop a dynamic pricing model to identify the significant factors that determine price dependency.

The results of this study can be used in companies that apply or plan to use a dynamic pricing strategy. They also can be helpful for the researchers who are interested in understanding the dynamic factors that can significantly affect price changes.

## **2. Research review on pricing in e-commerce**

There are plenty of different approaches to pricing that are used in e-commerce. In the article [1] Tanir identifies such types of pricing as cost-based pricing (the company determines the unit costs for each of the products and sets a target margin for each of them); market pricing (the company follows the prices of competitors and offers the prices which are a little cheaper than their competitors'); dynamic pricing (the company sets the optimal price in a specific period of time for a specific request in real-time); consumer or behavioral pricing (the company segments the audience through the use of real-time data, purchase history to accurately identify segments for customers); complete pricing (the company combines products of the same nature as complementary goods, which increases the average cost of the order, but the amount per order is less than buying all units separately) and market capture (the company sets the price for a new product on the market below average and thus capture this market segment).

Amongst the first studies devoted to dynamic pricing was the work of Weatherford and Bodily [2]. In that paper, dynamic pricing was shown as a tool to determine the optimal price for a product. Further theoretical principles, methods, and approaches used for implementation are mentioned in the works of Den Boer [3], Dolgui and Proth [4], Faruqui [5], Kannan and Kopalle [6], Sahay [7], Tong, Wang and Zhou [8].

A broader term that is used for dynamic pricing is ‘customized’ or ‘personalized’. The theoretical foundations of personalized pricing are described in Obermiller, Arnesen, and Cohen [9]. This type of pricing are defined as a strategy when “identical products are delivered, regardless of time or situation, to different consumers at different prices” [9, p. 14]. Elmachtoub, Gupta, and Hamilton [10] in their work define customized pricing as pricing that is based on personal information about the customer. In this case, it is appropriate to use machine learning methods to determine personal qualities.

The Office of Fair Trading [11] and the Organization for Economic Co-operation and Development [12] consider that customized pricing is related to price discrimination.

According to Iarmolenko and Chornous [13], transactional pricing, as a type of customized pricing which focuses on direct and indirect transaction features, is used only in the e-commerce market and can minimize price discrimination.

The relationship between different types of pricing in the information economy is analyzed in [14]. This work presents the peculiar properties of the category ‘dynamic pricing’ in the papers of various researchers. According to the authors, the concept of customized pricing is broader than dynamic or transactional ones. It is not limited to a particular market and is defined by the widest variety of factors that affect pricing. Dynamic pricing does not take into account the subjective characteristics of customers. Transactional pricing narrows the category of customized pricing to the e-commerce market.

Bichler uses the term ‘flexible pricing’ in the research [15]. He distinguishes between differentiated pricing (when different buyers are assigned different prices based on expected price values) and dynamic pricing (prices are formed during trading between market participants). In the author’s opinion, flexible pricing covers both of these categories.

According to the research review, there is no consensus on the boundaries of the definition ‘dynamic pricing’ among researchers. There is no clearly defined list of parameters the variation of which will maximize profit [14]. So, Srivastava in the paper [16] showed that dynamic pricing includes two aspects: price variance and price discrimination. Price variance can be spatial or temporal. In the spatial variance of prices, several sellers offer this item at different prices. With a temporary dispersion of prices, the store changes its price for a given product over time, based on the time of sale and the situation of supply and demand. An aspect of dynamic pricing is differentiated pricing or price discrimination when the same products are charged different prices to different consumers.

Belyh in the work [17] identifies the following types of dynamic pricing to solve a specific problem in the market.

1. Segmented pricing: divide customers into different segments by certain factors (price, quality, etc.).
2. Time-based pricing: inherent in companies that earn on fast service and work around the clock.
3. Changing market conditions: frequent changes in market conditions which cause the need of the companies to adapt quickly.
4. Peak pricing: most effective for peak periods in all industries.
5. Penetration pricing.

In this study, while defining ‘dynamic pricing’, the authors emphasize the synchronous change of prices in order to clearly separate the concept from the category of ‘customized pricing’ and avoid the association of the category with price discrimination. Therefore, segmented pricing is not considered. The authors also decided not to limit the goals of dynamic pricing only to the financial goals of the company but also include reputational goals. Maximizing profits, especially for a limited period of time, is not a fairly correct goal: a wrong pricing strategy can have negative consequences in future periods. One of the most difficult aspects of using a dynamic pricing strategy is the difficulty in determining consumers' willingness to pay. In addition, the cost of setting such prices can be very high. In large measure, this problem has been overcome through the use of the global network and Big Data technologies.

### **3. Related work in modeling dynamic pricing**

The idea of dynamic prices is very transpicuous. The company should adjust prices to demand. However, its implementation is much more difficult. To achieve this idea, the company uses different pricing approaches. Dynamic pricing algorithms provide flexibility because e-commerce companies can set prices based on different customer groups. The main idea is to offer the best price based on market trends, fluctuations in demand, customer behavior, purchasing power, and many other factors.

Among the classical pricing models can be distinguished parameterization of the demand function. The essence of this model is that the form of demand function is known. It is only necessary to find the parameter for this function bases on the historical data that will satisfy the condition of the problem. The problem is that there are about a dozen different possible types of the demand function. These functions of demand are constantly influenced by macroeconomic, psychological factors, consumer preferences, changes in fashion, and so on. These non-price factors affect the form of the function. Bertsimas and Perakis [18], Boer [19], Carvalho and Puterman [20], Lobo and Boyd [21] studied the parametric model.

The nonparametric approach is inverted to the parametric. In this type of model, price optimization occurs without constructing the demand function. There are a lot of methods that are used for price optimization. Stochastic approximation, Bayesian inference, Poisson process as a basis for the model, etc. are among them. The application of nonparametric approaches is described in Parkhimenko's paper [22].

Now non-parametric approaches are very popular. Among non-parametric models are economic and mathematical models, data mining. Scientists have substantiated the effectiveness of statistical and econometric approaches in their researches. They have also described different models for modeling and forecasting dynamic and personalized pricing in their papers. Carta, Medda, Recupero, and Saia considered ARIMA models [23], Liu, Liu, and Chan investigated used stochastic differential equations [24], Tseng, Lin, Zhou, Kurniajaya, and Li suggest signal processing methods [25], and also Tyrallis, Papacharalampous presented the random forest method [26], etc.

In dynamic pricing, in addition to complex approaches based on mathematical modeling, there are other algorithms: the price changes depending on the offer of competitors based on price parsing (automatic monitoring of prices for competitors' products).

The model based on inventories was studied by Chan, Shen, Simchi-Levi, Swann [27], Elmaghraby, Keskinocak [28]. According to that approach, pricing decisions are based on the number of inventories of the firm. The problem of dynamic pricing can be defined as the problem of maximizing the company's total revenue except for production costs and inventory storage costs. The auction model assumes that the price is set by the auction participants based on the value of the good for each agent. This model was studied by Bichler [15], Elmaghraby [29].

Game models can also be used for dynamic pricing. Such models should be used in a highly competitive economy when sellers compete for the same group of buyers. Various ways of computer modeling of prices in such cases are considered. Models based on cooperation between e-market players are the most realistic and bring the most benefit to market agents. Federgruen, Bernstein [30], Cao, Shen, Milito, Wirth [31] studied the models of game theory. An e-commerce company can combine several models and use them for modeling and forecasting dynamic pricing.

### **4. Forecasting dynamic factors of pricing based on models of Big Data analysis**

The last decade has shown an unprecedented increase in data in all areas of economic activity, including e-commerce. The amount of data on the activities of customers collecting electronic platforms is constantly growing. It is important that the set of factors influencing the formation of the price is considerably expanded. That is why it is crucial to know what factors can be included in the models of dynamic pricing. Involving more factors helps to increase the accuracy of the forecasts.

The model based on the use of Big Data technology on the Internet environment is considered by Yang, Zhao, Xing [32]. The model involves the use of methods that analyze statistics or other

parameters to determine consumer preferences and optimal dynamic prices. Among the technologies used are data mining, necessary for decision making (search for associative rules, classification problems, cluster analysis, regression analysis, detection and analysis of deviations, etc.), machine learning, simulation, statistical analysis, time series analysis, etc. To implement these technologies, various software products are used, including R, Python, MatLab, NoSQL, Hadoop, etc.

A special place among the models of dynamic pricing is occupied by machine learning models, which provide for the analysis of large arrays of information. Usually, the main parameters of the market environment are subject to constant dynamic changes, so it is impossible to predict all options for the development of the market system. Computer technology and Big Data technology have provided an opportunity to predict the main trends in market development and design pricing models for them. The researches in this area were done by Gupta, Ravikumar, Kumar [33], Zagaynova [34].

The main purpose of dynamic pricing is to respond in real-time to changes in factors that affect the price. Based on the approach described in this paper, these factors can be divided into two groups: dynamic (factors values of which change over time and depend on market conditions) and others (characteristics of the product or service and the characteristics of customers). Importantly, the number of dynamic factors available for observation in the information economy, specifically in e-commerce, is growing as rapidly as the set of behavioral characteristics of customers.

The feature of dynamic factors is that often they are collinear and quite difficult to determine which of them should involve the model or be excluded. Dynamic factors contain latent variables, the impact of which on the price is difficult to investigate using classical econometric or statistical models.

Promising models that can effectively predict changes in model factors are, for example, ridge regression and neural networks but none of them include hidden factors that are most useful in modeling. This disadvantage is absent in the model of PLS regression (Partial Least Squares, or Projection into Latent Structure). The PLS regression uses the decomposition of the original predictors along the axes of the principal components. This model also allocates a subset of latent variables in which the relationship between the dependent variable and predictors reaches the maximum value. The research [35] provides a comprehensive overview of Partial Least Squares (PLS) methods with a discussion of the directions of current research and perspectives.

The general underlying model of multivariate PLS is

$$X = TP^T + E, \tag{1}$$

$$Y = UQ^T + F, \tag{2}$$

where  $X$  –  $n \times m$  matrix of predictors,  $Y$  –  $n \times p$  matrix of responses,  $T$  and  $U$  –  $n \times l$  matrices projections of  $X$  and projections of  $Y$ ,  $P$  –  $m \times l$  orthogonal loading matrix;  $Q$  –  $p \times l$  orthogonal loading matrix, matrices  $E$  and  $F$  are the error terms, assumed to be independent and identically distributed random normal variables. The decompositions of  $X$  and  $Y$  are made to maximize the covariance between  $T$  and  $U$ .

PLS regression is a useful method when the factors are many and highly collinear. The increase in the number of dynamic characteristics of price formation in e-commerce means the increase in the number of model factors that correlate with each other. There are many obvious factors. However, there may be only a few keys or latent factors that explain much of the variation in response. The main idea of PLS regression is to try to extract these latent factors. Thus, the PLS model should be used to determine the most influential dynamic factors in the modeling of dynamic pricing.

Until the advent of Big Data, this approach was not used to establish a set of important factors that affect dynamic pricing. The current volumes of information resources cause the possibility and prospects of using the PLS model for forecasting in the field of e-commerce.

In the next section, we will demonstrate the implementation of this approach in comparison with the classical linear regression. We use the booking service Airbnb to study the models of dynamic pricing. In a number of studies [36-37], this service has already been considered as an object of development of a flexible pricing system. The purpose of dynamic pricing was to obtain the maximum profit for each future booking date. Thus, the article [36] presents three pricing models. The reservation probability model, based on the binary classification model, shows the probabilistic forecast of the reservation of flat for each night. Based on this model and two indicators (price decrease recall and booking regret) a model of pricing strategy is built. The model is based on historical data on booking prices on a particular night.

Based on information from the booking probability model, pricing strategy model, and quantification, customers were offered personalized prices that satisfied them as consumers and allowed the company to maximize profits.

In this study, we are going to develop applied approaches through the use of the PLS regression to analyze the impact of dynamic factors on price change and to show the prospects for the widespread introduction of this method for pricing and price forecasting in e-commerce.

## 5. Modeling of dynamic pricing in booking accommodation

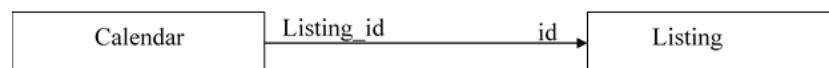
We examined the popular vacation rental e-platform Airbnb’s data [38] to model the dynamic pricing in order to determine the factors that most influence the price formation.

Airbnb is an American vacation rental online platform designed to view and book accommodation for rent. It is a marketplace where tenants and renters interact. The online platform connects owners and travelers and facilitates the rental process. In addition, this service develops the economy of shared use, allowing property owners to rent private apartments.

The main task of modeling is to find out the factors that determine the rental price dependency and their significance; identify the dynamic factors that are most influential in dynamic pricing when booking accommodation. Data analysis and modeling were performed in the software environment RStudio.

### 5.1. Data preparation

To build the model, we considered data on future bookings of vacation rent in Madrid (17.03.2020-17.03.2021), which contained the price for accommodation and information about available housing (table with data ‘Calendar’) and data with a description of housing (a type of housing, host rate, reviews, accommodation, and other characteristics) (table with data ‘Listing’). The dataset was formed by combining two tables (Figure 1).



**Figure 1:** Data connection scheme

The data needed to study pricing on the Airbnb platform contains a significant amount of categorical data. Some of the data can be transformed into Logical or Numeric views (for example, the presence/absence of certain rooms, facilities, and devices; proximity to the sights of the city), some important data (for example, such factor as feedback) may be involved in the study only after a sentimental analysis. In addition, the data also contained quite a few missing values. Based on the fact that the simulated building of missing values can distort the results, we got rid of rows and columns that contained a large number of spaces. The appropriate values of the categorical factors were converted to Logical (0/1) and Numeric types. In the next step, we formed several new factors. For example, among the non-dynamic factors based on the coordinates of the house location was calculated the distance to the city center. Because the important task of modeling is to determine the influence of dynamic factors, the set of such factors has been expanded. Variables were introduced to determine the number of days before settling: x1 - less than 290 days, x2 - more than 290 and less than 360 days. The threshold value of 290 days is justified by the fact that at this time there was a significant jump in price. This was the beginning of a new year, relative to the date of the study (March 17, 2020). We also singled out the day of the week when housing was ordered (0 - weekday, 1 - day off or holiday) and, on a similar principle, when housing was booked. A column was added with the popularity of housing, showing the percentage of booked housing per month of all housing offered by a particular owner. 23 columns (of 71) with 639,999 rows of data were formed after data processing. Table 1 describes the main factors of the processed dataset.

## 5.2. Data analysis

In order to initially understand the data, we have performed a visual analysis of the data on booking accommodation in Madrid. For the analysis, we used the RStudio software environment. For this purpose, we used such tools as bar and line charts, histograms, boxplots, scatter plots, correlation matrix and map, and others. The results of the analysis allowed us to obtain the set of conclusions.

The highest demand for housing was observed in the center of Madrid, slightly less in the north and north-east. The housing in the west and south areas are the less booked. In all areas of Madrid, a significant advantage is the booking of apartments (60% - 89%). The closer the area is to the center, the higher the percentage of apartment booking is. Private houses take the second place in terms of booking, both in the central and the outlying districts. The highest prices are in the central regions.

**Table 1**  
Data types

Data	Type	Description
listing_id	Numeric	Booking code
Date	Date	The date on which the accommodation is booked or available for booking
Available	Logical	Available or booked accommodation on a specific date
host_acceptance_rate	Logical	The host acceptance rate
is_location_exact	Logical	The correctness of the specified location
accommodates	Numeric	The capacity of housing
guests_included	Numeric	Number of guests
number_of_reviews	Numeric	Number of reviews
review_scores_rating	Numeric	Total rating by reviews
instant_bookable	Logical	The time rate of booking
TV	Logical	Availability of TV
Internet	Logical	Availability of the Internet
AirCondition	Logical	Availability of air conditioning
Pets	Logical	Opportunity to live with animals
Kitchen	Logical	Availability of kitchen
Breakfast	Logical	Breakfast included
Weekday1	Logical	The day of the week when the order is placed
Weekday2	Logical	The day of the week on which the booking is made
distance	Numeric	Distance to the center
x <sub>1</sub>	Logical	Less than 290 days before booking
x <sub>2</sub>	Logical	More than 290 and less than 360 days before booking
Price	Numeric	Price
i.available	Numeric	Percentage of booked housing per month

When analyzing prices considering popular days of the week for the rent, the highest demand is observed on Saturdays and Sundays. This tendency can be explained by the fact that people have days off. The cheapest prices are on Thursdays. Surprisingly, low prices are also observed on Fridays. The data also show a high spike in prices at the end of June. This can be explained by the fact that the famous national holiday of San Juan, which attracts tourists, occurs during this period in Spain. The rise in prices at the end of the year is due to the fact that tourists come to celebrate the New Year in Madrid.

The trend is also similar considering the dependency of the price on the day of placing the order. You will have to pay more ordering on weekends. However, low prices are observed when looking for housing on Mondays.

The increase in the price of accommodation is directly proportional to the number of bathrooms up to 5 rooms, while after that amount the price decreases. There is a similar dependence on the number

of rooms in an apartment but prices increase slowly. Considering a housing capacity, the cheapest prices are for the housing which can accommodate 1 person or 9 people, and the most expensive prices are for the housing which can accommodate 6 or 12 people.

The correlation analysis of the dependence of the variables demonstrates a high correlation (value 0.7 and higher) for almost half of the factors. Among collinear factors, almost 70% are dynamic factors.

### 5.3. Modeling of dynamic pricing and interpretation of results

Among the main dynamic factors for developing models, we included the day of the week when the order was made, the booked day of the week, the number of days before booking, the number of reviews, the total rating for reviews, the time rate of accommodation booking, the percentage of accommodation booked with the same owner during the month. And non-dynamic factors include the distance to the center, the capacity of housing, the availability of a kitchen, the Internet, air conditioning, TV, the inclusion of breakfasts, and so on.

We determined the significance of the factors using linear regression and the PLS model. Their use provides an answer to the question of whether dynamic factors significantly affect price formation.

Linear regression is a standard and reliable tool for finding linear relationships between factors but it has many limitations and disadvantages, especially in Big Data environment. The PLS model should help in finding latent pricing factors, especially this model is suitable for forecasting based on large information arrays in conditions of high collinearity of factors.

To construct a linear regression, we selected factors that showed a low level of correlation in the previous analysis. This limitation significantly narrowed down many factors, especially dynamic, because, for example, the days of the week or the number of days before booking had a high level of collinearity. Several linear regression models were built based on the selected factors. The best model is:

$$Price = 3.67 + 0.18accommodates - 0.06x_1 - 0.06distance + 0.15i.available \quad (3)$$

where *accommodates* – capacity of housing,  $x_1$  – less than 290 days before booking, *distance* – distance to the center, *i.available* – the percentage of accommodation booked with the same owner during the month.

To determine the quality of the model, we need to check several conditions: the absence of multicollinearity, absence of heteroscedasticity, and absence of autocorrelation.

So, in this case we explore panel data as the basis for modeling, and not time series, therefore it is sufficient that the first two conditions are met.

To check for multicollinearity, we realized two tests: VIF and CI. Both tests confirmed the absence of multicollinearity (CI = 11.14 < 30; VIF = 1.05 < 10).

To verify the absence of heteroscedasticity, we used criteria Goldfeld-Quandt and White. Both tests verified heteroskedasticity. We got rid of heteroskedasticity using a covariance matrix without heteroskedasticity.

All the factors of the regression are significant but  $R^2$  has a value of 0.5059, which indicates that the value is not enough to justify the forecasts. At the same time, we can confirm the presence of dynamic factors in the model. According to the simulation results, it turned out that the price is affected by the housing capacity, distance to the center, accommodation booking for 290 days, and the percentage of accommodation booked with the same owner during the month. As you can see, among the dynamic factors, these are the number of days before booking and the percentage of booked accommodation for the exact date. The presence of dynamic factors suggests that they affect the price and confirms that this online platform is used the dynamic pricing strategy.

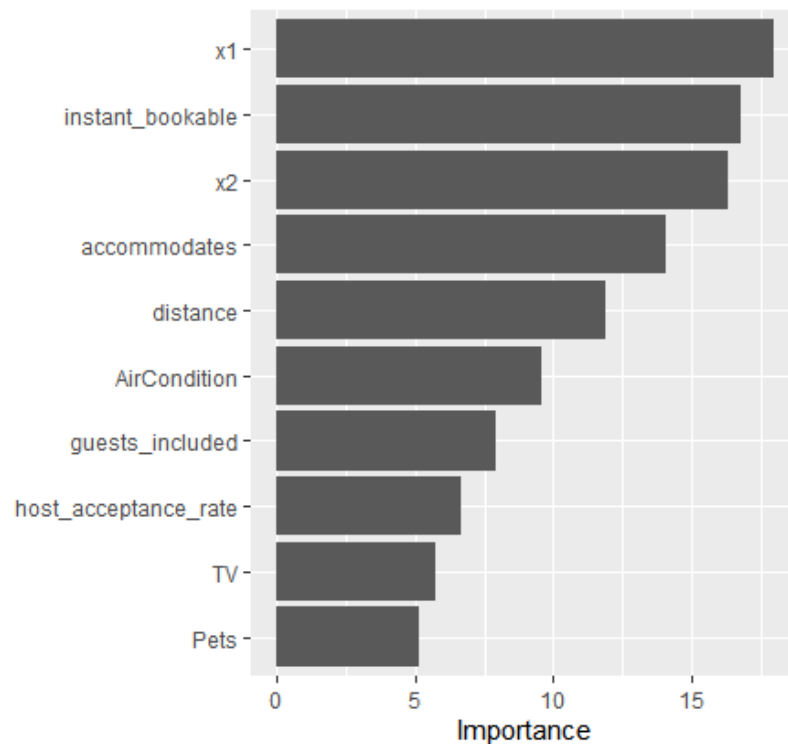
The next step was the implementation of the PLS regression. Among the advantages of using this model in comparison with the linear regression is that there is no need to narrow down factors set because this model effectively works with multicollinear data and with a large number of predictors.

PLS model finds the useful information that is contained in independent variables and in the relationship between dependent and independent variables. The goal of PLS is to maximize the covariance between variables X and Y in order to uncover latent factors. The main assessment of the quality of the model is the minimization of standard errors.



Based on the results of PLS modeling, we see that one component can explain 17.7% of the information, and 3 components can explain 51.2%.

The resulting model was the model with the lowest standard deviation and the highest  $R^2$ . The  $R^2$  estimate for this model is a non-standard quality assessment, so it was calculated using an econometric formula. The VIF projection approach was used to randomize the significance of the factors. The results of comparing the importance of predictors are presented in Figure 2.



**Figure 2:** Influence of importance of factors on pricing

By estimating the magnitude of the impact of each of the factors on the price, we see that the most influential are such dynamic factors as booking accommodation 290 days before the trip, the time rate of booking, and booking more than 290 days but less than 360. Slightly less impact on price has accommodation capacity and the distance to the center.

The PLS regression demonstrates a high dependence of dynamic factors on the price. As the importance of the dynamic factors is 51.2% for determining the price.

Estimates of the quality of the main resulting models in the research, as well as the names of significant dynamic factors, are summarized in Table 2.

Thus, both models, which were developed based on the booking database of the Airbnb platform in the city of Madrid, demonstrated the dependence of the rental price on dynamic factors. In addition, we can make sure about the high level of significance of dynamic factors in both models.

At the same time, we see that the results of applying the linear regression model give a rather rough estimate on a large data set (low value of  $R^2$ ). The PLS regression delves more subtly into the hidden relationships between factors in the context of Big Data and demonstrates a range of significant dynamic factors. This fact once again confirms the feasibility of using such an approach to modeling dynamic pricing in the presence of huge arrays of observations.

A significant amount of effort to prepare the data for modeling is one of the main problems that occurred during modeling. This is due to both the need for data integration and the low quality of the original data. A large amount of categorical data was needed to be transformed. Some problems concerned the expansion of the range of dynamic factors.

Some problems were caused by the fact that the initial data were uploaded on a specific date, so it is difficult to see the dynamics of change. The dataset is based on the offered prices by the service for home owners and this affects the accuracy of forecasting (for example, prices were presented in the

format: 70, 75, 80, 85, etc, rounded to integer multiples of 5). Adjusted host prices would have shown a greater dynamic relationship.

**Table 2**  
Estimation of model quality

Indicators	Linear regression	PLS regression
R <sup>2</sup>	0.5059	0.6842
RMSE	-	27.4188
Dynamic factors	x <sub>1</sub> , i.available	x <sub>1</sub> , instant_bookable, x <sub>2</sub>

It should also be noted that the analysis of large amounts of data requires significant computing capacity and requires more time for modeling.

## 6. Conclusion

In today's competitive market environment, e-commerce has a number of advantages over traditional trading. Among these advantages are the absence of maintenance costs, low market entry costs, low 'menu' costs, flexibility, as well as the ability to use significant amounts of information on the specifics of the sale or purchase transactions to form favorable prices. The development of the information economy, especially the e-commerce market, facilitates faster adaptation to changes in market conditions and consumer preferences. The new opportunities allow companies to set the optimal price for a specific period of time for a specific consumer.

More and more e-commerce platforms are using dynamic pricing as a type of personalized pricing. Among the positive aspects of dynamic pricing usage can be identified the growth of business profitability, the absence of direct price discrimination, because it does not take into account information about the behavioral characteristics of a particular customer, their solvency, and so on. These special features alleviate the ethical issues that typically arise from customized pricing. Prices are the same for all consumers but depend on time factors.

The features of dynamic pricing include the speed of response to changes in the market, flexibility, control over pricing strategy, cost savings in the long term, implementation of specific software into management processes that promotes informatization and digitalization of business.

This study with the mathematical modeling implementation confirmed the use of dynamic pricing strategies on the popular online service Airbnb. Significantly, this online platform is recognized as one of the most successful businesses in e-commerce. It has many competitive advantages, which confirms the growth of its popularity compared to the popularity of Booking and other similar e-services.

This fact once again confirms the relevance of studying the company's successful experience to increase competitiveness in e-commerce and prospects of using modern data mining technologies and powerful methods to form strategies for pricing and determining factors affecting prices. This study places special emphasis on identifying the dynamic factors of pricing that have become increasingly important.

When implementing the new software to support personalized pricing, e-commerce companies can use the approach proposed in this study, namely, the PLS regression. The advantages of this model are the ability to identify hidden predictors among a large number of collinear factors (dynamic factors are often hidden and collinear) and good performance in the processing of large information arrays, even in cases where the number of observations is small, and number of factors is large. It is also important to note that the inclusion of this algorithm in the toolkit of other methods and approaches does not require significant funds, as it is present in most popular open-source libraries.

It should be noted that we can recommend the use of PLS regression in the implementation of all other types of personalized pricing because in modern conditions the constant increasing the number of factors is also due to the expansion of the range of behavioral characteristics of consumers.

Thus, the use of the PLS model answers the today's challenges and this approach should take its rightful place among the tools to support personalized pricing.

## 7. References

- [1] B. Tanir. “E-Commerce pricing.” Prisync.com, 2018. URL: <https://prisync.com/blog/ultimate-ecommerce-pricing-strategies/>.
- [2] L. R. Weatherford, S. E. Bodily. “A Taxonomy and Research Overview of Perishable-Asset Revenue Management: Yield Management, Overbooking, and Pricing,” *Operations Research*, Vol. 40, Issue 5, September-October (1992): 831-844. DOI: 10.1287/opre.40.5.831
- [3] A. V. Den Boer. “Dynamic Pricing and Learning: Historical Origins, Current Research, and New Directions.” *Surveys in operations research and management science*, Vol. 20, Issue 1, June (2015): 1-18. DOI: 10.1016/j.sorms.2015.03.001.
- [4] A. Dolgui, J.-M. Proth. “Supply Chain Engineering: Useful Methods and Techniques.” Springer, New York, NY, 2010.
- [5] A. Faruqui. “The Ethics of Dynamic Pricing.” *The Electricity Journal*, Vol. 23, Issue 6, July (2010): 13-27. DOI: 10.1016/j.tej.2010.05.013.
- [6] P. K. Kannan, P. K. Kopalle. “Marketing in the E-Channel.” *International Journal of Electronic Commerce*, Vol. 5, Issue 35, Dec. (2001): 63-83. DOI: 10.1080/10864415.2001.11044211.
- [7] A. Sahay. “How to Reap Higher Profits with Dynamic Pricing.” *MIT Sloan Management Review* 48, (2007): 53-60.
- [8] Y. Tong, L. Wang, Z. Zhou, L. Chen, B. Du, J. Ye. “Dynamic Pricing in Spatial Crowdsourcing: A Matching-Based Approach.” *International Conference on Management of Data*, May (2018): 773-788. DOI: 10.1145/3183713.3196929
- [9] C. Obermiller, D. Arnesen, M. Cohen. “Customized Pricing: Win-Win or End Run?” *Drake Management Review* 1 (2012): 12-28.
- [10] A. N. Elmachtoub, V. Gupta, M. Hamilton. “The Value of Personalized Pricing.” *ssrn.com*, 2018. URL: <https://ssrn.com/abstract=3127719>.
- [11] Office of Fair Trading, *Personalized Pricing: Increasing Transparency to Improve Trust*, 2013. URL: [https://one.oecd.org/document/DAF/COMP\(2018\)13/en/pdf](https://one.oecd.org/document/DAF/COMP(2018)13/en/pdf).
- [12] Organization for Economic Co-operation and Development, *Personalized Pricing in the Digital Era*, 2018. URL: [http://www.oft.gov.uk/shared\\_of/market-work/personalised-pricing/oft1489.pdf](http://www.oft.gov.uk/shared_of/market-work/personalised-pricing/oft1489.pdf).
- [13] I. Iarmolenko, G. Chornous. “The Model of a Second-Hand Goods Resale Exchange under Transactional Pricing Strategy.” *Ekonomika*, Vol. 99(1), May (2020): 69-78. DOI: 10.15388/Ekon.2020.1.4.
- [14] G. Chornous, I.Iarmolenko. “Spivvidnoshennia mizh riznovydamy tsinoutvorennia v informatsiinii ekonomitsi.” [Relationship between types of pricing in the information economy], in: V. S. Ponomarenko, T. S. Klebanova (Eds.), *Tools for Modeling Systems in the Information Economy*, VSHEM-KHNEU, Kharkiv, 2019, pp. 120-135. (in Ukrainian)
- [15] M. Bichler, R. D. Lawrence, J. Kalaganam, H. S. Lee, K. Katircioglu, G. Y. Lin, A. J. King, Y. Lu. “Applications of flexible pricing in business-to-business electronic commerce.” *IBM Systems Journal*, Vol. 41, Issue 2, (2002): 287–302. DOI: 10.1147/sj.412.0287.
- [16] A. Srivastava, *Dynamic pricing models: Opportunity for action*, Cap Gemini Ernst & Young Center for Business Innovation, 2001.
- [17] A. Belyh. *Different Types of Dynamic Pricing*, 2018. URL: <https://www.cleverism.com/complete-guide-dynamic-pricing/>.
- [18] D. Bertsimas, G. Perakis, *Dynamic Pricing: A Learning Approach*, in: S. Lawphongpanich, D.W. Hearn, M.J. Smith (Eds.), *Mathematical and Computational Models for Congestion Charging. Applied Optimization*, volume 101, Springer, Boston, MA, 2006. DOI: 10.1007/0-387-29645-X\_3. doi:10.1007/0-387-29645-X\_3.
- [19] A. V. Den Boer. “Dynamic Pricing with Multiple Products and Partially Specified Demand Distribution.” *Mathematics of Operations Research*, Vol. 39, Issue 3, August (2014): 597-948. DOI:10.1287/moor.2013.0636.
- [20] A. X. Carvalho, M. L. Puterman. “Dynamic optimization and learning: how should a manager set prices when the demand functions is unknown?” *Technical Report 1117*, Instituto de Pesquisa Economica Aplicada, 2005.

- [21] M. S. Lobo, S. Boyd. "Pricing and learning with uncertain demand." Technical report, Stanford University, 2003.
- [22] V.A. Parkhimenko, "Zadacha dinamicheskogo opredeleniya optimal'noj ceny s ispol'zovaniem neparametricheskogo podhoda." [The problem of dynamically determining the optimal price using a nonparametric approach]. Mezhdunardnaya nauchnaya konferenciya BGUIR, (2014), Minsk. (in Russian)
- [23] S. Carta, A. Medda, A. Pili, R. Recupero, R. Saia. "Forecasting E-Commerce Products Prices by Combining an Autoregressive Integrated Moving Average (ARIMA) Model and Google Trends Data." *Future Internet*, Vol. 11, Issue 1, December (2018). DOI: 10.3390/fi11010005.
- [24] W. W. Liu, Y. Liu, N. H. Chan. "Modeling EBay Price Using Stochastic Differential Equations." *Journal of Forecasting*, Vol. 38, Issue 1, January (2019): 63-72. DOI: 10.1002/for.2551.
- [25] K.-K. Tseng, R. F.-Y. Lin, H. Zhou, K.J. Kurniajaya, Q. Li. "Price Prediction of E-Commerce Products Through Internet Sentiment Analysis." *Electronic Commerce Research*, Vol. 18, Issue 1, March (2018): 65-88. DOI: 10.1007/s10660-017-9272-9.
- [26] H. Tyrallis, G. Papacharalampous. "Variable Selection in Time Series Forecasting Using Random Forests." *Algorithms*, Vol. 10, Issue 4, October (2017): 114. DOI: 10.3390/a10040114.
- [27] L. M. A. Chan, Z. J. M. Shen, D. Simchi-Levi, J. Swann, Coordination of pricing and inventory decisions: A survey and classification, in: *Handbook on Supply Chain Analysis: Modeling in the E-Business Era*, Series in Operations Research and Management Science, Kluwer Academic Publishers, 2005.
- [28] W. Elmaghraby, P. Keskinocak. "Dynamic pricing. Research overview, current practices and future directions." *Management Science*, Vol. 49, Issue 10, October (2003): 1287–1309. DOI: 10.1287/mnsc.49.10.1287.17315
- [29] W. Elmaghraby, Auctions and pricing in e-marketplaces, in: *Handbook of Quantitative Supply Chain Analysis: Modelling in the E-Business Era*, International Series in Operations Research and Management Science, Kluwer Academic Publishers, 2005.
- [30] F. Bernstein, A. Federgruen. "Pricing and replenishment strategies in a distribution system with competing retailers." *European Journal of Operational Research*, Vol. 51, Issue 3, June (2003): 409–426. DOI: 10.1287/opre.51.3.409.14957
- [31] X. Cao, H. Shen, R. Milito, P. Wirth. "Internet pricing with a game theoretic approach: concepts and examples." *ACM Transactions on Networking*, Vol. 10, Issue 2, April (2002): 208–216. DOI: 10.1109/90.993302
- [32] J. Yang, C. Zhao, C. Xing, On this page Abstract Introduction Literature Review Conclusions Data Availability Conflicts of Interest Acknowledgments References Copyright Special Issue Applications of Machine Learning Methods in Complex Economics and Financial Networks View, 2019.
- [33] M. Gupta, K. Ravikumar, M. Kumar. "Adaptive strategies for price markdown in a multiunit descending price auction: A comparative study, in: *Proceedings of the IEEE Conference on Systems, Man, and Cybernetics*, 2002, pp. 373–378.
- [34] E. V. Zagainova, "Model' dinamicheskogo cenoobrazovaniya na rynke pasazhirskih aviaperevozok." [A Model of Dynamic Pricing in the Air Passenger Market]. *ZHurnal ekonomicheskoy teorii*, (2017): 177-182 (in Russian).
- [35] V. Esposito, J. Henseler, H. Wang, W. W. Chin, *Handbook of Partial Least Squares: Concepts, Methods and Applications*, 1st. ed., Springer-Verlag Berlin Heidelberg, 2010.
- [36] P. Ye, J. Qian, J. Chen, C. Wu, Y. Zhou, S. D. Mars, F. Yang, L. Zhang. Customized Regression Model for Airbnb Dynamic Pricing, in: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Association for Computing Machinery, New York, NY, USA, 2018, pp. 932–940. DOI: 10.1145/3219819.3219830.
- [37] J. L. Wang, D. Nicolau. "Price determinants of sharing economy based accommodation rental: a study of listings from 33 cities on Airbnb.com." *International Journal of Hospitality Management*, Vol. 62, April (2017): 120-131. DOI: 10.1016/j.ijhm.2016.12.007.
- [38] Inside Airbnb, InsideAirbnb.Com. URL: <http://insideairbnb.com/get-the-data.html>.