ORIGINAL SCIENTIFIC PAPER
UDK: 338.485.2(460)
005.31:519.246.8
DOI: 10.5937/EKOPRE2204240V
Date of Receipt: January 16, 2022

Vladimir Vasić
University of Belgrade
Faculty of Economics
Department for Statistics and Mathematics

Branislava Hristov Stančić
University of Belgrade
Faculty of Economics
Department for Economic Policy and Development

Bojan Zečević
University of Belgrade
Faculty of Economics
Department for Economic Policy and Development

EMPIRICAL MODELING OF HOTEL OCCUPANCY RATE WITH DYNAMIC PANEL DATA

Empirijsko modelovanje stope popunjenosti kapaciteta hotela dinamičkim panel podacima

Abstract

The previous research studies used mainly the occupancy rate as one of the key indicators of hotel performance. As the hotel occupancy rate varies both throughout the year and for different types of hotels, the use of panel data is more appropriate and more comprehensive compared to the cross-sectional data or time series, which have so far been most commonly used in similar research. Also, the previous research did not take into account the great heterogeneity among the analyzed hotels, nor the correlation of the occupancy rate in relation to its past values. By using the generalized method of moments within the dynamic panel data model, it is possible to take both properties into account. The analyzed data pertain to the hotel industry of Spain. Specifically, the given panel data include a sample of 49 hotels observed over a period of 12 years. The application of dynamic panel analysis shows that the values of hotel occupancy rate are influenced by the values of hotel occupancy rate with a lag one, as well as the values of total marketing expenses with a lag one. It was further determined that the values of incentive management fees, as well as the average daily rate and the consumer price index also have an impact on the observed variable. We are convinced that the presented analysis results will be of significant benefit to hotel managers.

Keywords: *hotel industry, occupancy rate, dynamic panel data, panel generalized method of moments, Sargan test, Arellano-Bond serial correlation test.*

Sažetak

U dosadašnjim istraživanjima, kao jedan od glavnih pokazatelja uspešnosti poslovanja hotela, uglavnom se uzimala stopa popunjenosti kapaciteta. Kako stopa popunjenosti kapaciteta u jednom hotelu varira tokom godine, ali isto tako varira i za različite tipove hotela, za njenu analizu prikladnija je i sveobuhvatnija upotreba forme panela podataka u odnosu na podatke preseka ili vremenske serije, koji su se do sada najčešće primenjivali u sličnim istraživanjima. Takođe, dosadašnja istraživanja nisu istovremeno uzimala veliku heterogenost među analiziranim hotelima, kao ni povezanost stope popunjenosti kapaciteta sa njenim istorijskim vrednostima. Upotrebom generalizovane metode momenata u okviru modela dinamičkih panel podataka, moguće je da se oba svojstva uzmu u obzir. Analizirani podaci pripadaju hotelskoj industriji Španije. Tačnije, dati panel podaci obuhvataju uzorak od 49 hotela posmatranih u periodu od 12 godina. Primenom dinamičke panel analize pokazano je da na vrednosti stope popunjenosti kapaciteta utiču vrednosti stope popunjenosti kapaciteta sa docnjom jedan, kao i vrednosti ukupnih troškova marketinga sa docnjom jedan. Daljom analizom utvrđeno je da na posmatranu varijablu utiču i vrednosti podsticajnih nagrada za menadžere, kao i prosečna dnevna cena sobe, kao i potrošački indeks cena. Uvereni smo da će menadžerima hotela prikazani rezultati analize biti od značajne koristi.

Ključne reči: *hotelska industrija, stopa popunjenosti kapaciteta, dinamički panel podaci, panel generalizovani metod momenata, Sarganov test, Arelano-Bond test serijske korelacije.*

Introduction

Financial performance ratios are commonly used as an indicator of business success, but also as the bottom line target, given that companies expect to generate an appropriate return on investment. Insight into these operating ratios does not provide information on the influence of intangible performance drivers, such as knowledge and customer and partner relationship, which significantly affect the hotel's performance [48, p. 600]. This shortcoming is compensated by the introduction of non-financial, i.e., qualitative indicators, which are the measure of quality, flexibility and implementation of new technologies [6, p. 149]. Hotels that follow and meet the needs and desires of customers offer far more sophisticated products thus achieving defined financial goals and a competitive advantage [41, p. 555].

Regarding operational performance, the three most commonly used performance indicators in the hotel industry are the average daily rate, revenue per available room and occupancy rate [14, p. 22], [17, p. 2], [15, p. 143], [43, p. 11]. Given that information on intangible resources such as the level of hotel guest satisfaction is often only partially available and that hotel financial information almost always remain unavailable and inconsistent for comparison purposes [25, p. 39], in scientific literature occupancy rate becomes a general measure of lodging performance [28], [34]. Jeffrey et al. [26, p. 74] consider that the occupancy rate is the only performance indicator that is widely available and which is relevant for monitoring and evaluating the hotel performance at the individual hotel level across the entire industry. Measuring the occupancy of the facility is an extremely good solution for showing the success of the hotel owing to the frequent reluctance of hotel managers to give detailed financial results. For most hotels, the occupancy rate is an effective addition to the financial results [42, p. 207]. It also enables the identification of trends and fluctuations within the industry [38, p. 176].

However, decision making based only on the occupancy rate is inadequate, because it could lead to erroneous conclusions about business success of the hotel. Occupancy information needs to be supplemented by information on prices obtained from sophisticated revenue management systems, in order to best translate the occupancy into the room net income. In particular, it is necessary to maintain a balance in the level of prices and the level of occupancy, because even a small increase in prices could lead to a significant reduction in the occupancy rate [28, p. 58].

Some studies have explored the issue of occupancy rates' changes and forecasts [27], [28], [29], [33], while some have examined key determinants of hotel occupancy rates [1], [21], [30], [31], [34], [37]. The occupancy rate forecasting is of great importance as it provides important information for both government agencies and hotel managers [10, p. 55]. On the other hand, by identifying and managing the key determinants of the occupancy rate, managers have the opportunity to achieve the desired occupancy more efficiently as well as to achieve better bottom line results.

In view of these considerations, the hotel occupancy rate was used in this research to illustrate the business performance of hotels in Spain. The study examines the key determinants of the occupancy rate in Spanish hotels. In the model, the occupancy rate is determined as a dependent variable, while four explanatory variables were identified: total marketing expenses – TME, incentive management fees – IMF, average daily rate – ADR and consumer price index – CPI. According to the best of our knowledge, the model created in this way represents an original model that has not been presented in the scientific literature so far. Its originality lies in the selection and combination of independent variables, as well as in the research methodology. Specifically, panel data were used in the paper, which represents a new and modern approach in treating this kind of issue.

Previous research in this domain relied on all types of data. Some papers used panel data [9], [12], [21], [34], while some used time series data [10], [26], [27], [28], [33] or cross-sectional data [1], [20].

Our model clearly identifies variables of interest whose management could lead to an increase in the occupancy rate providing implications for hotel management practice.

Literature review and hypotheses development

To manage the hotel occupancy rate and to improve the bottom line financial results, it is necessary to understand the determinants affecting hotel occupancy. Hotel occupancy rate depends both on external and internal factors. The external factors refer to the state of the economy and politics, legislation, technologies and demographics [1, p. 200]. On the other hand, there are many more internal factors and in the scientific literature they are grouped differently depending on their influence on the hotel's operations. The hotel's internal factors which could affect innovation behavior and business performance are: firm size, membership in a business group, organizational aspects, high costs of innovation, lack of qualified personnel [37, pp. 145-146]. According to Lockyer [35, p. 481], the four major internal factors which affect hotel demand, and hence the hotel's occupancy, are price, location, facilities and cleanliness. Factors that are also rated as significant for hotel occupancy are cleanliness, location, room rate and security [3, p. 13]. There are other independent variables used in modeling occupancy rates such as the size of hotel, annual average room rate, proportion of free individual travelers, proportion of domestic travelers and chain-affiliation. These variables are determinants of the hotels' characteristics and therefore could have impact on the occupancy rate performance [21, p. 25]. According to Lei and Lam [34, pp. 3-4], the most important factors that may affect hotel occupancy rate are: average room rate, total available rooms, number of tourist arrivals, gross domestic product, inflation, CPI, star rating, seasonality and casino facility. Within these variables only four were statistically significant: average room rate, total available rooms, star rating and casino facility. Kim, Cho and Brymer [30, p. 406] have also confirmed the positive impact of the number of available rooms on the occupancy rate of hotels.

Understanding cost behavior is crucial for creating accurate budgets and controlling operations to enhance hotel profitability. There has been little empirical work done in the domain of investigation of the impact of operating costs on the hotel business performance [13], [39]. These studies have shown that there is a significant impact of operating costs on bottom line indicators, but the question which arises is how these costs affect the indicator such as occupancy rate. On this basis, the following is hypothesized:

H_1: There is a significant positive impact of explanatory variable total marketing expenses on dependent variable occupancy rate.

Although trained, professional and proactive management is crucial for achieving the desired hotel performance, the question of the importance and impact of managers' salaries and their bonuses on hotel performance is almost nonexistent in the scientific literature. Specifically, the only paper that examined the relationship between hotel room revenues and gross operating profit on the one hand, and managerial fees and their bonuses on the other hand was written by Hua, DeFranco and Abbott [19, p. 4]. To the best of our knowledge, until now there was no scientific work that has examined the impact of incentive fees (bonuses) for managers on the degree of hotel occupancy. Consequently, the following hypothesis is proposed:

H_2: There is a significant positive impact of explanatory variable incentive management fees on dependent variable occupancy rate.

The ADR reflects the hotel's ability to generate revenue intensively from occupied rooms. This indicator is calculated as the mean price charged for all hotel rooms sold in a given period [14, p. 22]. The ADR and occupancy rate are usually treated as two major hotel business performance indicators [26, p. 86], [43, p. 11]. Determining the right room rates is one of the most crucial functions for any hotel [11, p. 65]. The prices charged by a hotel directly affect its daily basis performance in terms of competitive position, occupancy rate and revenues [16]. In that sense, as prices could significantly affect the behavior of hotel guests, it is important to consider whether price variations can significantly affect variations in capacity utilization. Hence, the following two hypotheses are proposed:

H_3: There is a significant negative impact of explanatory variable average daily rate on dependent variable occupancy rate.

During the past two decades in the countries of eurozone, inflation and changes in prices in the hospitality industry have received much attention from researchers and policymakers. One of the measures of the inflation rate is the consumer price index, which is even more relevant in small, open economies that have a large share of exports in the total balance of payments [46, p. 94]. As Spain is on the list of 20 leading countries in terms of imports and exports, the CPI is a good measure of inflation in this country [45]. The CPI measures changes in the prices of products and services that households procure to fulfill their needs. Spending on tourism and hotel services is closely connected to the state of the economy and economic cycle. As rapid growth or falling of prices can significantly harm a healthy economy and stable business, it is important to see how a change in this index will affect changes in tourist behavior in terms of their decision to make reservations which will be measured in this study by occupancy rate. Therefore, it is important to consider the relationship between the CPI and the performance of the hotel sector measured by occupancy rate, and thus the following hypothesis is suggested:

H_4: There is a significant positive impact of explanatory variable consumer price index on dependent variable occupancy rate.

Research methodology

The first step in econometric modeling would certainly be descriptive statistics and testing [32, p. 54]. These activities belong to the area of preliminary analysis. When testing, we usually compare the average values of the dependent variable in relation to some groups.

Panel data will require two indexes to be able to exactly identify each observation. Panel data mainly combine time series and cross-sectional data. In that case, if we observe the variable Y, its individual value is denoted by $y_{i,t}$, where the index i takes the values $i = 1,\ldots,N$, while the index t takes the values $t = 1,\ldots,M$.

In order to test the equality of the arithmetic means of the dependent variable, in relation to different groups, it is necessary to first define between and within sums of squares using the formulas: $SS_B = H\sum_{i=1}^{N}(\bar{y}_i - \bar{y})^2$ and $SS_W = H\sum_{i=1}^{N}\sum_{t=1}^{M}(y_{i,t} - \bar{y}_i)^2$ [22, p. 803]. In the previous formulas, $\bar{y}_i$ denotes the sample arithmetic mean within group i, and $\bar{y}$ denotes the overall sample arithmetic mean. We are now able to define test statistics for testing the equality of arithmetic means by groups:

$$F = \frac{SS_B/(N-1)}{SS_w/(MN-N)} \tag{1}$$

where MN denotes the total number of observations. Test statistics have an F distribution with degrees of freedom (N-1, MN-N). The assumptions for the application of testing is that the observed data have a normal distribution by groups, as well as equal variance [23, p. 450]. If the variance is not equal across groups, a robust version of the previous test statistics will be used to test the equality of the sample environments, and such a test is called the Welch test. To create the Welch test statistics, we first need to calculate the weights w_i by the formula $\frac{M}{s_i^2}$, where s_i^2 represents the sample variance in the group i. The given test statistics are presented by the formula

$$F^* = \frac{\frac{1}{N-1}\sum_{i=1}^{N} w_i(\bar{y}_i - \bar{y}^*)^2}{1 + \frac{2(N-2)}{N^2-1}\sum_{i=1}^{N}\frac{(1-h_i)^2}{M-1}} \tag{2}$$

where h_i is the normalized weight which is calculated as , and $\bar{y}^*$ is the weighted grand mean, which is calculated as $\sum_{i=1}^{N} h_i \bar{y}_i$. This robust statistic has an approximate F distribution with (N – 1, DF^*) degrees of freedom, where DF^* is calculated as $\frac{N-1}{3\sum_{i=1}^{N}\frac{(1-h_i)^2}{M-1}}$.

Panel covariances

Panel data are quite complex, hence for a better understanding of panel data, it is useful to analyze them from the cross-section point of view, and from the periods point of view. For this reason, we can define measures of association between cross-sections or between periods for the dependent variable. The covariances ($\sigma_{i,j}$)) for the dependent variable between cross-sections are calculated by the formula $E\{[Y_i - E(Y_i)][Y_j - E(Y_j)]\}$, where $Y_i^T = (Y_{i,1}, Y_{i,2}, \ldots, Y_{i,M})$ represents a random variable that refers to the i^{th} cross-sectional data of the panel variable Y, $i = 1,\ldots,N$. Covariance between cross-sections represents the association between data for different cross-setions, for a given moment in time [24, p. 1220].

Similarly, we can define the covariances within the cross-sections for the dependent variable Y. The given covariance is calculated by the formula $\sigma_{s,t} = E\{[Y_s - E(Y_s)][Y_t - E(Y_t)]\}$, where $Y_i^T = (Y_{1,t}, Y_{2,t}, \ldots, Y_{N,t})$, . The within cross-section covariance measures the association between data in different time periods, for a particular cross-section. For example, if we want to calculate, for a panel dependent variable, the covariance for different time periods, the following formula will be used:

$$\sigma_{s,t} = \tfrac{1}{N} \Sigma_{i=1}^{N} (Y_{i,t} - \bar{Y}_t)(Y_{i,s} - \bar{Y}_s) \quad (3)$$

where $\bar{Y}_t = \frac{1}{N} \Sigma_{i=1}^{N} Y_{i,t}$ and $\bar{Y}_s = \frac{1}{N} \Sigma_{i=1}^{N} Y_{i,s}$.

Generalized method of moments estimation

Estimation using the generalized method of moments in panel data is based on moments [18, p. 100] which are of the form $g(\beta) = \Sigma_{i=1}^{N} g_i(\beta) = \Sigma_{i=1}^{N} Z_i^T u_i(\beta)$, where Z_i represents the matrix of instruments for cross-section i, and where the error term $u_i(\beta)$ is given by the equation $Y_i - f(X_{i,t},\beta)$. Note that, when the estimation by the generalized method of moments in panel data is performed, the summation is done in some cases, in relation to periods t instead of cross-section i.

Estimation by the generalized method of moments, in essence, comes down to minimizing the following expression [47, p. 525]:

$$S(\beta) = (\Sigma_{i=1}^{N} Z_i^T u_i(\beta))^T W(\Sigma_{i=1}^{N} Z_i^T u_i(\beta)) = g(\beta)^T W g(\beta) \quad (4)$$

in relation to the parameter β and the corresponding weighting matrix W. Finally, we can conclude that, in the data panel, the estimation using the generalized method of moments includes the determination of the instruments (in the notation Z) and the determination of the weighting matrix W.

Dynamic panel data modeling

Dynamic panel data modeling is intended for panel data consisting of a large number of cross-sectional units and a small number of periods [7, p. 142], (as is the case with the panel data we analyze). Otherwise linear models of dynamic panels include p lags of dependent variables as covariates in the model. In addition to the given covariates, the model also contains cross-sectional effects [8, p. 4]. Thus, the model also contains past values of the dependent variable, as well as cross-sectional effects, which are correlated with each other, so that the estimation using standard methods is wrong, because the estimates are not consistent. This problem is overcome by using estimation with the generalized method of moments [4, p. 148].

A linear model of dynamic panel data, can be represented by an expression

$$Y_{i,t} = \Sigma_{j=1}^{p} \rho_j Y_{i,t-j} + X_{i,t}^T \beta + \delta_i + u_{i,t} \quad (5)$$

where δ_i represents cross-sectional effects. The given cross-sectional effect can be eliminated by applying the first-difference operator. In this way equation (5) becomes

$$\Delta Y_{i,t} = \Sigma_{j=1}^{p} \rho_j \Delta Y_{i,t-j} + \Delta X_{i,t}^T \beta + \delta_i + \Delta u_{i,t} \quad (6)$$

which can be estimated using the generalized method of moments. Also, efficient estimation using the generalized method of moments usually includes a different number of instruments for each period [44, p. 115]. The different number of instruments (for each period) is determined on the basis of the numbers of lagged dependent and predetermined variables that are available for a given period. For example, if we want to consider the use of lagged values of the dependent variable in equation (6) as instruments. If the error terms in equation (5) are independent and identically distributed, then the first period available for the use of the instruments is $t = 3$. We can easily see that $Y_{i,1}$ is a valid instrument because it is correlated with $\Delta Y_{i,2}$ while it is uncorrelated with $\Delta u_{i,3}$. Also, for the period $t = 4$, the potential instruments are $Y_{i,2}$ and $Y_{i,1}$. Continuing on this principle, we can create a set of instruments for case i using lags of the dependent variable:

$$H_i = \begin{bmatrix} Y_{i,1} & -1 & 0 & \dots & \dots & \dots & \dots & 0 \\ 0 & Y_{i,1} & Y_{i,2} & \dots & \dots & \dots & \dots & 0 \\ \dots & \dots & & & & & & \\ 0 & 0 & 0 & \dots & Y_{i,1} & Y_{i,2} & \dots & Y_{i,T-2} \end{bmatrix} \sigma^2.$$

Similarly, sets of instruments can be created for each predetermined variable. In this way we can determine the instruments needed for estimation using the generalized method of moments. It remains only to determine the weighting matrix. The weighting matrix used in the one-step Arellano-Bond estimator is given by the form $W = (\frac{1}{N} \Sigma_{i=1}^{N} Z_i^T A Z_i)^{-1}$, where the matrix A is defined as

$$A = \tfrac{1}{2} \begin{bmatrix} 2 & -1 & 0 & \dots & 0 & 0 \\ -1 & 2 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 2 & -1 \\ 0 & 0 & 0 & \dots & -1 & 2 \end{bmatrix} \sigma^2.$$

Finally, note that the weighting matrix used in the two-step Arellano-Bond estimator is given by the form $W = (\frac{1}{N}\sum_{i=1}^{N} Z_i^T \Delta u_i \Delta u_i^T Z_i)^{-1}$, where u_i represents residuals from one-step estimation.

Sargan test

The Sargan test [40, p. 393] is used to test the null hypothesis that the over-identifying restrictions are valid. Thus, if the instrument rank (in the notation *p*) is greater than the number of estimated coefficients (in the notation *k*), we can use the Sargan test [5, p. 173]. The Sargan statistic (referred to as *J-statistic* in most software), if the null hypothesis is correct, has an approximately chi-square distribution with a degree of freedom. The p-value may be computed using the formula:

$$p - value = 1 - P(\chi^2_{p-k} \leq J - statistic) = 1 - F(J - statistic) \\ = @chisq(J - statistic, Instrument\ rank - (\#estimated\ coefficients)) \quad (7)$$

where *F(x)* denotes cumulative distribution functions, while the expression in the last equation is a command in EViews software that is used to calculate the required p-value.

Arellano-Bond serial correlation test

The use of the generalized method of moments in dynamic panel data is correct only if there is no autocorrelation in model errors (at the level of original, undifferenced values) [2, p. 278]. Testing that there is no autocorrelation in errors in dynamic panel data models is complex, because transformed errors have a more complicated structure than untransformed (original) errors. The Arellano-Bond test for autocorrelation refers to transformed model errors, where the transformation is represented by the operator of the first difference.

If the untransformed model errors are independent and identically distributed, then the values of the first-difference model errors will be autocorrelated. So, the rejection of the null hypothesis of the non-existence of autocorrelation of order one, at the first-difference model errors, is quite correct. However, the rejection of the null hypothesis of the lack of second-order autocorrelation indicates that the application of the generalized method of moments in dynamic panel data is not correct. For that reason, we will calculate the autocorrelation of the first and second order. Therefore, to test the null hypothesis of the non-existence of autocorrelation in model errors, two test statistics will be created. One for first-order autocorrelation testing and the other for second-order autocorrelation testing.

If the untransformed error values of the model are independent and identically distributed, we expect the first-order autocorrelation value to be negative and statistically significant, while the second-order autocorrelation value to be insignificant. Thus we calculate the test statistics as:

$$m_j = \frac{AVE\ (p_j)}{\sqrt{VAR\ (p_j)}} \quad (8)$$

where $AVE(\rho_j)$ represents the average *j*th order autocovariance, which is obtained by the formula $\frac{1}{M-3-j}\sum_{t=4+j}^{M} \rho_{t,j}$, while autocovariance $\rho_{t,j}$ is obtained by applying the expected value operator $E(\Delta u_{i,t}, \Delta u_{i,t-j})$.

Descriptive statistics and sample

Five variables are described and used in this research. All data are secondary data provided by a global hotel market research company STR. The sample contains annual data for 49 hotels based in Spain spanning the period of twelve years, from 2006 through 2017. In the field of hospitality, this period of twelve years is especially

Table 1: Description and possible impact of explanatory variables in a dynamic model

Label	Name	Unit of measure	Role	Possible impact
OCC	Occupancy rate	capacity occupancy rate	dependent	
TME	Total marketing expenses	in U.S. $	explanatory	positive
IMF	Incentive management fees	in U.S. $	explanatory	positive
ADR	Average daily rate	in U.S. $	explanatory	negative
CPI	Consumer price index	2010 = 100	explanatory	positive

Source: The result of the analysis conducted by the authors.

challenging for research because it covers periods of both stable markets and periods of crisis. Spain was chosen as the subject of analysis as it is one of the most important tourist destinations in Europe.

Table 1 shows all the variables that will be included in the modeling. In addition to the variable name, a label of variables is given, which will be used in the labeling in the tables that follow. There will be four explanatory variables in the model, which are assumed to have a positive (except for one variable) effect on the dependent variable. It is only assumed that the explanatory variable "average daily rate" will have a negative impact, as dictated by economic logic. All explanatory variables (except one) were measured in U.S. dollars, only the variable "consumer price index" as its name indicates, is given as an index (with 2010 as the base year). The dependent variable is measured as a percentage, hence its range by definition is from 0% to 100%.

Table 2 shows the descriptive statistics of the analyzed variables. Looking at the realized values of descriptive statistics, we can see the basic characteristics of the analyzed variables, in the given period from 2006 to 2017. When considering the characteristics of the analyzed variables, it is important to observe whether the data distributions are symmetric. This is one of the basic assumptions to be examined, before we include the variable in the econometric model. The distribution of data is symmetric if mean statistics and median statistics have similar values, and if the skewness statistics is close to zero (approximately, in the interval from -1 to +1).

Analyzing the variable TME (total marketing expenses), it is observed that the value of statistics means is much higher than the value of median statistics, which indicates a potential problem of positive data asymmetry. This feature is confirmed by skewness statistics whose value is 3.7 (which is significantly higher than the value of +1). If the variable has the property of positive asymmetry, by using the transformation, the natural logarithm of the variable acquires the property of symmetry. The good side of the natural logarithm transformation is that the transformed variables have the same structure as the original variables. Because the data structure does not change, the transformed values are usually referred to as the original values. For the same reasons, the same

Table 2: Dependent and explanatory variables descriptive statistics

	OCC	TME	IMF	ADR	CPI
Mean	67.72453	387649.9	105601.8	111.9332	102.2904
Median	68.80000	143536.0	99.00000	93.78000	103.1961
Maximum	99.10000	4420058.	2618657.	374.9800	108.3753
Minimum	35.40000	0.000000	0.000000	32.06000	92.09113
Std. Dev.	11.94259	617619.0	249622.4	62.04151	5.247204
Skewness	-0.157994	3.740425	5.442653	1.981410	-0.590437
Kurtosis	2.604123	20.05336	45.05804	6.895838	1.988170
Observations	583	583	583	583	583

Source: Authors' calculations using EViews 12 software.

Table 3: Average values of dependent and explanatory variables by time periods

DATEID	OCC	TME	IMF	ADR	CPI	Obs.
2006	69.9	305353.5	66437.0	110.6	92.1	49
2007	69.8	336742.8	81528.4	126.9	94.7	49
2008	66.7	352555.0	88162.7	137.2	98.5	49
2009	60.7	322774.2	62038.5	109.7	98.2	49
2010	64.3	396338.6	0.0	105.0	100.0	49
2011	66.2	418048.9	636.8	112.9	103.2	48
2012	64.3	443234.3	129898.5	102.6	105.7	49
2013	65.6	409939.3	124162.6	105.0	107.2	48
2014	67.9	412862.0	140738.2	107.2	107.0	49
2015	70.7	371523.9	142620.2	96.1	106.5	48
2016	73.8	423275.7	206408.9	106.3	106.3	47
2017	73.2	461350.4	227696.4	122.9	108.4	49

Note: Categorized by values of *DATEID*; Sample: 2006-2017; Included observations: 583.
Source: Authors' calculations using EViews 12 software.

transformation was applied to the variables IMF (incentive management fees) and ADR (average daily rate).

As already mentioned, we are considering panel data. Panel data are two-dimensional data, hence the analysis of all data simultaneously (both in relation to the cross-section, and in relation to the periods) cannot see in detail the structure and descriptive characteristics of the analyzed variables. For this reason, when it comes to preliminary analysis of panel data, it is desirable to calculate descriptive statistics of variables, both in relation to cross-section, and in relation to periods.

Table 3 presents the average values of the analyzed variables in relation to years. Observing the dependent variable OCC, we notice that the capacity utilization rate varies from 60.7% in 2009 to 73.8% in 2016. Whether the given variation of some 13% can be attributed to randomness, and whether it can be stated that during the analyzed time the capacity utilization rate remains unchanged, or there are statistically significant differences, will be checked by applying the statistics given by formulas (1) and (2). Using formula (1), we perform a standard F-test. The realized value of the test statistics is 5.52, while the calculated degree of freedom is (11, 571). The corresponding p-value is 0.000, hence the null hypothesis of the equality of arithmetic means over time is rejected. When it comes to the Welch F-test, whose test statistics are given by formula (2), the realized value of the test statistics is 6.07, while the calculated degree of freedom is (11, 224.78). The corresponding p-value is 0.000, therefore, using this test, the null hypothesis of the equality of arithmetic means over time is rejected.

Table 4 presents the average values of the analyzed variables in relation to the hotels in the sample. Observing the dependent variable OCC, we notice that the occupancy rate varies from 51.9% for a hotel with code 108682288 to 88.5% for a hotel with code 106346849. The variation of 26.6% between hotels in the sample is probably statistically significant, and this can be verified by testing the statistical hypothesis that there is no variation in the occupancy rate between hotels. Using formula (1), we perform a standard F-test. The realized value of the test statistics is 21.6, while the calculated degree of freedom is (48, 534). The corresponding p-value is 0.000, hence the hypothesis

Table 4: Average values of dependent and explanatory variables by cross-sections

HOTEL_ID	OCC	TME	IMF	ADR	CPI	Obs.
100166086	80.2	278827.6	148744.6	76.1	102.3	12
100267940	65.8	180831.3	67895.3	91.2	102.3	12
100400406	55.9	79259.6	11927.8	65.7	102.3	12
100495193	64.3	38481.9	14523.1	76.3	102.3	12
100508515	83.1	352497.3	59422.8	104.7	102.3	12
101016492	84.1	399393.8	6.8	120.1	102.3	12
101316567	71.6	704317.7	160309.1	153.5	102.3	12
101435073	73.6	85602.3	70871.8	112.5	102.3	12
101846036	71.0	43453.8	21936.9	89.7	102.3	12
101869218	72.3	1015026.0	242510.5	162.4	102.3	12
102126665	63.1	509393.2	128719.3	94.8	102.3	12
102796244	57.0	51034.3	11884.4	76.4	102.3	12
103253890	74.1	77724.3	54508.1	111.8	102.3	12
103469179	79.6	228916.3	114094.3	100.7	102.3	12
103558280	77.3	1501121.0	211890.8	287.8	102.3	12
103759870	55.6	218901.9	33574.8	97.6	102.3	12
103910813	70.9	181407.9	43624.8	80.4	102.2	11
103978148	58.4	880068.1	96338.7	139.2	102.3	12
104408061	68.0	250938.5	77535.6	151.2	101.5	10
104412699	61.4	566058.3	281164.6	146.9	102.3	12
104578397	77.6	98499.8	65434.7	101.5	102.3	12
104964822	55.1	101458.3	17284.2	74.1	102.3	12
105063067	81.2	245207.3	158969.1	58.1	102.3	12
105070623	76.3	1516236.0	296737.3	166.5	102.3	12
105138211	61.2	35812.7	4325.0	38.9	102.3	12
105256687	64.0	822621.2	359250.7	305.2	102.3	12
105579030	59.5	374080.4	193468.2	116.7	102.3	12
105681357	67.7	666833.9	205094.9	75.1	102.3	12
105905349	62.4	90452.4	12753.0	73.6	102.3	12
106128128	62.9	233976.3	36882.9	89.6	102.3	12
106346849	88.5	129219.7	45397.6	94.5	102.3	12
106399414	85.0	251934.0	309100.9	66.1	102.3	12
106733672	58.5	42960.5	16878.6	79.7	101.9	11
107271489	55.2	31992.3	31.1	66.0	102.0	11
107284228	66.1	914102.4	215007.1	253.1	102.3	12
107434770	52.3	315210.9	34119.3	58.3	102.3	12
108056278	53.4	105929.8	12941.3	58.9	102.3	12
108425004	63.4	82011.5	74255.3	114.2	102.3	12
108629194	60.9	56242.8	37028.9	84.1	102.3	12
108682288	51.9	95131.2	22.6	88.8	102.3	12
109024518	53.2	78408.6	24935.5	65.5	102.3	12
109062560	58.7	182122.4	63000.0	99.5	102.3	12
109130453	76.5	57738.4	59126.5	161.7	102.3	12
109578035	71.7	3764618.0	757028.8	286.3	102.3	12
109632926	76.0	632431.1	105864.6	110.7	102.3	12
109790851	73.8	58625.8	37084.3	93.0	102.3	12
109994400	72.1	98487.5	46819.7	85.6	102.3	12
110008462	69.1	88668.0	33390.6	81.9	102.3	12
110056395	76.0	82240.5	74737.7	95.8	102.3	12

Note: Categorized by values of HOTEL_ID; Sample: 2006-2017; Included observations: 583.
Source: Authors' calculations using EViews 12 software.

that there is no variation in the occupancy rate between hotels is rejected. When it comes to the Welch F-test, whose test statistics are given by formula (2), the realized value of the test statistics is 30.5, while the calculated degree of freedom is (48, 185.066). The corresponding p-value is 0.000, thus even using this test, the null hypothesis that there is no variation in the occupancy rate between hotels is rejected.

By previous analysis, we observed the complexity of the structure of the dependent variable OCC. Its average values differ statistically significantly, both in the cross-sectional dimension and in the period dimension, therefore it is justified to analyze it as a panel structure.

Within the panel structure, the relationship of the dependent variable will be observed. Table 5 shows both the panel covariance and the panel correlation coefficients of the dependent variable in relation to the periods. Covariance coefficients are shown on the main diagonal, as well as in the upper right triangle of the coefficient matrix, which is given in Table 5. In the lower left triangle, in brackets, the correlation coefficients are shown. Covariance coefficients are calculated using formula (3), while correlation coefficients are obtained by applying the same formula, but over previously standardized data.

By analyzing the panel correlation coefficients, it is observed that there is a high correlation between the values of the dependent variable in different time periods, so that the past values of the dependent variable could be included in the model. This is made possible by the application of dynamic panel data model.

Empirical results

When the dynamic panel data model (given by formula (6)) is applied over the variables we have explained, and when the estimation is performed by the generalized method of moments, the estimates of unknown coefficients are given in Tables 6 and 7. Thus, when formula (6) is applied, we get the following form:

Table 6 gives a point estimate of the model coefficients, while Table 7 shows both 90% and 95% confidence interval values for estimates of unknown coefficients in the model. Table 6 in the last column shows the realized p-values of testing the statistical significance of unknown coefficients. Observing the given p-values, and comparing them with the value 0.05, we come to the conclusion that the following variables in the model are statistically significant: *OCC (-1)*, *LN_TME (-1)*, *LN_IMF*, *LN_ADR* and *CPI*.

Based on Table 6, we see that the variable OCC(-1) has a statistically significant and positive effect on the dependent variable OCC, which means that the occupancy rate of the hotel from the previous period has a positive effect on the occupancy rate of the hotel in the current period. Then, that the explanatory variable TME(-1) has a statistically significant and positive effect on the dependent variable. This means that the total marketing costs from the previous period positively affect the occupancy rate of the hotel in the current period. Then, that the explanatory variable IMF is statistically significantly and positively affects the dependent variable. Therefore, incentive management fees for managers have a positive effect on

Table 5: Panel covariance (correlation) coefficient values of the dependent variable

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
2006	138.3	115.3	106.6	71.5	88.1	95.0	86.4	73.3	59.2	54.0	49.8	62.7
2007	(0.93)	111.2	100.3	73.9	88.2	94.9	90.4	75.0	59.9	52.3	39.8	53.6
2008	(0.85)	(0.90)	112.5	84.3	95.8	102.9	91.0	81.5	64.4	55.4	54.3	60.9
2009	(0.60)	(0.70)	(0.79)	101.5	115.0	118.8	111.0	98.6	81.4	72.8	62.0	58.0
2010	(0.61)	(0.68)	(0.74)	(0.93)	149.9	156.0	145.6	133.5	113.8	100.8	77.5	68.9
2011	(0.59)	(0.66)	(0.71)	(0.87)	(0.94)	184.8	167.0	161.3	142.0	121.3	98.5	87.8
2012	(0.56)	(0.65)	(0.65)	(0.84)	(0.90)	(0.93)	172.7	163.6	142.1	122.8	91.1	82.2
2013	(0.46)	(0.52)	(0.57)	(0.72)	(0.80)	(0.87)	(0.92)	184.1	162.1	129.8	101.9	83.5
2014	(0.41)	(0.46)	(0.50)	(0.66)	(0.76)	(0.85)	(0.88)	(0.98)	150.0	116.3	89.1	72.1
2015	(0.44)	(0.48)	(0.50)	(0.70)	(0.79)	(0.86)	(0.90)	(0.92)	(0.91)	107.8	76.3	61.3
2016	(0.45)	(0.40)	(0.55)	(0.66)	(0.68)	(0.78)	(0.74)	(0.81)	(0.78)	(0.79)	86.9	71.7
2017	(0.62)	(0.59)	(0.67)	(0.67)	(0.65)	(0.75)	(0.73)	(0.72)	(0.68)	(0.69)	(0.89)	74.0

Note: Sample: 2006-2017; Included observations: 583; Analysis of clustered (between periods) relationships; Number of periods employed: 12; Balanced sample (listwise missing value deletion).
Source: Authors' calculations using EViews 12 software.

the occupancy rate of the hotel. Further, the explanatory variable ADR has a statistically significant and negative effect on the dependent variable. This means that the reduction of the average daily price of a hotel room has a positive effect on the occupancy rate of the hotel. Finally, the macro-explanatory variable CPI has a statistically significant and positive effect on the dependent variable, that is, the consumer price index has a positive effect on the hotel occupancy rate.

Based on the given results, we can infer that all hypotheses, which were set in this research, are accepted.

To verify the validity of the dynamic panel data model, if it is over-identified, the number of estimated coefficients in the model will first be established. By analyzing Table 7, we observe that the number of estimated coefficients is 16. In Table 8, we have information that the instrument rank is 49. Thus, the instrument rank is greater than the number of estimated coefficients, therefore the dynamic

Table 6: Dependent variable dynamic modeling using the panel generalized method of moments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
OCC(-1)	0.627126	0.052644	11.91264	0.0000
OCC(-2)	-0.001272	0.023734	-0.053600	0.9575
LN_TME	-0.788545	1.035373	-0.761605	0.4500
LN_TME(-1)	1.486149	0.695413	2.137074	0.0377
LN_IMF	0.312245	0.102334	3.051218	0.0037
LN_ADR	-18.10350	4.513205	-4.011229	0.0002
CPI	5.171299	1.717308	3.011281	0.0041
@LEV(@ISPERIOD("2009"))	-6.003858	1.206782	-4.975097	0.0000
@LEV(@ISPERIOD("2010"))	-2.270099	3.316700	-0.684445	0.4970
@LEV(@ISPERIOD("2011"))	-17.11041	5.304079	-3.225896	0.0023
@LEV(@ISPERIOD("2012"))	-19.92523	4.535985	-4.392703	0.0001
@LEV(@ISPERIOD("2013"))	-5.991343	2.447494	-2.447950	0.0181
@LEV(@ISPERIOD("2014"))	3.169607	0.366827	8.640615	0.0000
@LEV(@ISPERIOD("2015"))	1.567405	1.022415	1.533043	0.1318
@LEV(@ISPERIOD("2016"))	3.654011	1.230511	2.969507	0.0046
@LEV(@ISPERIOD("2017"))	-10.03055	3.704190	-2.707893	0.0094

Note: Sample: Transformation: First differences; Sample (adjusted): 2009-2017; Periods included: 9; Cross-sections included: 49; Total panel (unbalanced) observations: 436; White period (period correlation) instrument weighting matrix; White period (cross-section cluster) standard errors and covariance (d.f. corrected); Standard error and t-statistic probabilities adjusted for clustering; Instrument specification: *@DYN(OCC,-2, -7) LN_TME LN_TME(-1) LN_IMF LN_ADR CPI @LEV(@SYSPER);* Constant added to the instrument list.
Effects specification: Cross-section fixed (first differences), Period fixed (dummy variables).
Source: Authors' calculations using EViews 12 software.

Table 7: Coefficient confidence intervals values

		90% CI		95% CI	
Variable	Coefficient	Low	High	Low	High
OCC(-1)	0.627126	0.540344	0.713909	0.523648	0.730604
OCC(-2)	-0.001272	-0.040397	0.037853	-0.047924	0.045379
LN_TME	-0.788545	-2.495347	0.918256	-2.823703	1.246613
LN_TME(-1)	1.486149	0.339768	2.632530	0.119226	2.853072
LN_IMF	0.312245	0.143547	0.480942	0.111093	0.513396
LN_ADR	-18.10350	-25.54347	-10.66353	-26.97478	-9.232214
CPI	5.171299	2.340334	8.002264	1.795709	8.546888
@LEV(@ISPERIOD("2009"))	-6.003858	-7.993226	-4.014490	-8.375943	-3.631773
@LEV(@ISPERIOD("2010"))	-2.270099	-7.737645	3.197448	-8.789499	4.249301
@LEV(@ISPERIOD("2011"))	-17.11041	-25.85413	-8.366688	-27.53625	-6.684560
@LEV(@ISPERIOD("2012"))	-19.92523	-27.40276	-12.44771	-28.84130	-11.00917
@LEV(@ISPERIOD("2013"))	-5.991343	-10.02601	-1.956675	-10.80221	-1.180480
@LEV(@ISPERIOD("2014"))	3.169607	2.564897	3.774317	2.448563	3.890652
@LEV(@ISPERIOD("2015"))	1.567405	-0.118035	3.252846	-0.442282	3.577093
@LEV(@ISPERIOD("2016"))	3.654011	1.625526	5.682495	1.235284	6.072738
@LEV(@ISPERIOD("2017"))	-10.03055	-16.13687	-3.924231	-17.31161	-2.749490

Note: Sample: 2006-2017; Included observations: 436.
Source: Authors' calculations using EViews 12 software.

panel data model is over-identified. The validity of the over-identified dynamic panel data model is tested using the Sargan test.

Table 8: Sargan test

Root MSE	6.174196	Mean dependent var	0.766055
S.D. dependent var	5.870487	S.E. of regression	6.290700
Sum squared resid	16620.62	J-statistic	36.60987
Instrument rank	49	Prob(J-statistic)	0.304829

Source: Authors' calculations using EViews 12 software.

To apply the Sargan test, we need test statistics, which are in fact Sargan statistics, and are usually denoted by *J-statistic*. The given statistics are calculated and shown in Table 8, and their value is 36.6. We are now able to calculate the p-value using formula (7):

$$p-value = 1 - P(\chi^2_{p-k} \leq 36.6) = 1 - F(36.6) =$$
$$@chisq(36.6,33) = 0.30$$

The calculated p-value is greater than the value 0.05, hence we do not reject the null hypothesis of the validity of the excessive identification of the dynamic panel data model.

Table 9: Arellano-Bond serial correlation test

Sample: 2006 2017
Included observations: 436

Test order	m-Statistic	rho	SE(rho)	Prob.
AR(1)	-2.393830	-6381.675067	2665.884844	0.0167
AR(2)	-0.504390	-761.046063	1508.843979	0.6140

Note: This test is only available for equations estimated by GMM using first-difference cross-sectional effects.
Source: Authors' calculations using EViews 12 software.

Finally, it remains to check whether the model error terms from equation (5) are autocorrelated. To examine this, the Arellano-Bond serial correlation test is applied to the model given by equation (6). The test results are shown in Table 9. In fact, two tests were applied, one for determining if there is first-order autocorrelation in the model error terms (in the notation *AR(1)*), and one for determining if there is second-order autocorrelation (in the notation *AR(2)*). If the model error terms from equation (5) are uncorrelated, then we expect that the model error terms from equation (6) have a negative and statistically significant first-order autocorrelation without having a second-order autocorrelation. Table 9 shows that there is a statistically significant first-order autocorrelation (because the p-value is less than 0.05) and is negative (its value is -2.39), and that there is no second-order autocorrelation, because its p-value of 0.61 is greater than 0.05.

Limitations

The limitation of this research is that we cannot generalize the results for a given region of the European western Mediterranean, but only for the state of Spain, for the period from 2006 to 2017. In the next research, the results obtained could be validated over a larger sample that would in addition include hotels from France, Monaco, Italy, Malta and Gibraltar. Also, the analysis of the data did not take into account information on the types of hotels in terms of the number of hotel stars, their hotel chain affiliation, location, age or level of service. Future research should certainly take this information into account when identifying key variables of capacity utilization, since the different types of hotels have differing operating characteristics such as TME, IMF and level of ADRs.

For a deeper analysis in terms of giving the answer to the question as to why variables such as lagged occupancy rate, lagged incentive management fees, total marketing expenses, ADR and CPI have impact on occupancy rate, it is necessary to make an additional, qualitative research in the field of functioning, organization and earnings policy in hotels. It would also be necessary to analyze the state of the Spanish economy during the observed period as well as further analyze the structure of price changes included in the consumer price index.

Conclusion

This research investigated the impact of four factors on the occupancy rates of the hotels in Spain, hence four hypotheses were formed and tested. A generalized method of moments was used to estimate unknown coefficients in a dynamic panel data model. The use of panel data models proved to be justified, because the dependent variable hotel occupancy rate has statistically significant different average values, both in relation to cross sections and in relation to time sections. Also, the results of the analysis indicate that the current values of hotel occupancy rate are statistically significantly affected by the values of hotel occupancy rate with a lag one, the values of total marketing expenses with a lag one, as well as the values of incentive management fees. In addition to these business and economic variables, the average daily rate as well as the consumer price index have a statistically significant impact. The given results were obtained on the basis of the analysis of 49 hotels from Spain for a period of 12 years. Also, the results of this analysis should be interpreted with caution, because the results have not been validated. For that reason, a more comprehensive research is planned, which will include other countries in the western part of the European Mediterranean. However, we still expect the results of this analysis to be very useful to the hotel management, should the focus of their activities be the hotel occupancy rate.

References

1. Abdullah, A. A., & Haan, M. H. (2012). Internal success factor of hotel occupancy rate. *International Journal of Business and Social Science*, *3*(22).
2. Arellano, M. & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies 58*(2), 277–297.
3. Atkinson, A. (1998). Answering the eternal question: What does the customer want?. *The Cornell Hotel and Restaurant Administration Quarterly, 29*(2), 12-14.
4. Baltagi, B. H. (2008). *Econometric analysis of panel data (4th ed.).* Chichester: John Wiley & Sons.
5. Baltagi, B. H. (2009). *A Companion to Econometric Analysis of Panel Data*. Chichester: Wiley.
6. Bergin-Seers, S., & Jago, L. (2007). Performance measurement in small motels in Australia: (Funded by the Sustainable Tourism Co-operative Research Centre). *Tourism and Hospitality Research*, *7*(2), 144-155.
7. Bond, S.R. (2002). Dynamic panel data models: A guide to micro data methods and practice. *Portuguese Economic Journal, 1*(2), 141-162.
8. Brañas-Garza, P., Bucheli, M., & García-Muñoz, T. M. (2011). Dynamic panel data: A useful technique in experiments. Department of economic theory and economic history of the University of Granada*,* No. 10/22, 1-13.
9. Casado-Díaz, A. B., & Sellers-Rubio, R. (2020). The effect of short-term rentals on regional hotel efficiency. *Current Issues in Tourism*, 1-6.
10. Chow, W. S., Shyu, J. C., & Wang, K. C. (1998). Developing a forecast system for hotel occupancy rate using integrated ARIMA models. *Journal of International Hospitality, Leisure & Tourism Management*, *1*(3), 55-80.
11. Cross, R.G., Higbie, J.A., & Cross, D.Q. (2009). Revenue management's renaissance: A rebirth of the art and science of profitable revenue generation. *Cornell Hospitality Quarterly, 50*(1), 56-81.
12. Dogru, T., Hanks, L., Mody, M., Suess, C., & Sirakaya-Turk, E. (2020). The effects of Airbnb on hotel performance: Evidence from cities beyond the United States. *Tourism Management*, *79*, 104090.
13. Enz, C. A., & Potter, G. (1998). The impacts of variety on the costs and profits of a hotel chain's properties. *Journal of Hospitality & Tourism Research*, *22*(2), 142-157.
14. Enz, C. A., Canina, L., & Walsh, K. (2001). Hotel-industry averages: An inaccurate tool for measuring performance. *Cornell Hotel and Restaurant Administration Quarterly*, *42*(6), 22-32.
15. Enz, C. A., Peiró-Signes, Á., & Segarra-Oña, M. D. V. (2014). How fast do new hotels ramp up performance?. *Cornell Hospitality Quarterly*, *55*(2), 141-151.
16. Hanks, R. D., Cross, R. G., & Noland, R. P. (2002). Discounting in the hotel industry: A new approach. *Cornell Hotel and Restaurant Administration Quarterly*, *43*(4), 94-103.
17. Heo, C. Y. (2017). New performance indicators for restaurant revenue management: ProPASH and ProPASM. *International Journal of Hospitality Management*, *61*, 1-3.
18. Hsiao, C. (2014). *Analysis of Panel Data, Third Edition*. New York, NY: Cambridge University Press.
19. Hua, N., DeFranco, A., & Abbott, J. (2020). Management fees and hotel performance in the US. *Tourism Management*, *79*, 104093, 1-11.
20. Hua, N., Wei, W., DeFranco, A. L., & Wang, D. (2018). Do loyalty programs really matter for hotel operational and financial performance?. *International Journal of Contemporary Hospitality Management*, *30*(5), 2195-2213.
21. Hung, W. T., Shang, J. K., & Wang, F. C. (2015). Exploring the determinants of hotel occupancy rate: A dynamic panel data approach. *Abstract of Economic, Finance and Management Outlook*, *3*, 24-26.
22. IBM (2017). *IBM SPSS Statistics Algorithms*. Armonk, NY: IBM Corporation.
23. IHS Markit (2020). *EViews 12 User's Guide I*. Seal Beach, CA: IHS Global Inc.
24. IHS Markit (2020). *EViews 12 User's Guide II*. Seal Beach, CA: IHS Global Inc.
25. Jang, S., & Yu, L. (2002). Analysis of return on hotel investment: a comparison of commercial hotel companies and casino hotel companies. *Journal of Hospitality & Tourism Research*, *26*(1), 38-53.
26. Jeffrey, D., Barden, R. R., Buckley, P. J., & Hubbard, N. J. (2002). What makes for a successful hotel? Insights on hotel management following 15 years of hotel occupancy analysis in England. *Service Industries Journal*, *22*(2), 73-88.
27. Jeffrey, D., & Hubbard, N. J. (1988). Temporal dimensions and regional patterns of hotel occupancy performance in England: a time series analysis of midweek and weekend occupancy rates in 266 hotels, in 1984 and 1985. *International Journal of Hospitality Management*, *7*(1), 63-80.
28. Jeffrey, D., & Hubbard, N. J. (1994a). A model of hotel occupancy performance for monitoring and marketing in the hotel industry. *International Journal of Hospitality Management*, *13*(1), 57-71.

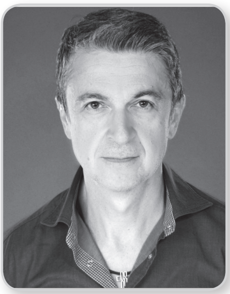
29. Jeffrey, D., & Hubbard, N. J. (1994b). Spatial-temporal patterns in the English hotel and tourist industries: a time series analysis of competitive occupancy performance at the individual hotel level. *Geoforum*, *25*(1), 73-86.
30. Kim, W. G., Cho, M., & Brymer, R. A. (2013). Determinants affecting comprehensive property-level hotel performance: The moderating role of hotel type. *International Journal of Hospitality Management*, *34*, 404-412.
31. Lam, C. C., & Lei, W. S. (2010). Hotel occupancy rate in a gambling destination: a longitudinal study of the Macao hotel Industry 2004-2007. *Journal of Hospitality & Tourism Management, 17*(1), 34-43.
32. Larose, D. T., & Larose, C. D. (2015). *Data Mining and Predictive Analytics* (Second Edition). Hoboken, New Jersey: John Wiley & Sons, Inc.
33. Law, R. (2004). Initially testing an improved extrapolative hotel room occupancy rate forecasting technique. *Journal of Travel & Tourism Marketing*, *16*(2-3), 71-77.
34. Lei, W. S. C., & Lam, C. C. C. (2015). Determinants of hotel occupancy rate in a Chinese gaming destination. *Journal of Hospitality and Tourism Management*, *22*, 1-9.
35. Lockyer, T. (2005). Understanding the dynamics of the hotel accommodation purchase decision. *International Journal of Contemporary Hospitality Management*, *17*(6), 481-492.
36. López-Fernández, M. C., Serrano-Bedia, A. M., & Gómez-López, R. (2011). Factors encouraging innovation in Spanish hospitality firms. *Cornell Hospitality Quarterly*, *52*(2), 144-152.
37. Mattsson, J., & Orfila-Sintes, F. (2014). Hotel innovation and its effect on business performance. *International Journal of Tourism Research*, *16*(4), 388-398.
38. O'Neill, J. W. (2011). Hotel occupancy: is the three-year stabilization assumption justified?. *Cornell Hospitality Quarterly, 52(*2), 176-180.
39. O'Neill, J. W., Hanson, B., & Mattila, A. S. (2008). The relationship of sales and marketing expenses to hotel performance in the United States. *Cornell Hospitality Quarterly*, *49*(4), 355-363.
40. Sargan, J. (1958). The estimation of economic relationships using instrumental variables. *Econometrica 26*(3), 393-415.
41. Sin, L. Y., Alan, C. B., Heung, V. C., & Yim, F. H. (2005). An analysis of the relationship between market orientation and business performance in the hotel industry. *International Journal of Hospitality Management*, *24*(4), 555-577.
42. Singh, A., Dev, C. S., & Mandelbaum, R. (2014). A flow-through analysis of the US lodging industry during the great recession. *International Journal of Contemporary Hospitality Management*, *26*(2), 205-224.
43. Singh, A., & Schmidgall, R. S. (2002). Analysis of financial ratios commonly used by US lodging financial executives. *Journal of Retail & Leisure Property*, *2*(3), 201-213.
44. Stata Press Publication (2019). *STATA longitudinal-data/ panel-data – reference manual release 16*. College Station, TX: StataCorp LLC.
45. The Observatory of Economic Complexity 4.0. Retrieved from https://oec.world/en/profile/country/esp.
46. Vašícek, B. (2011). Inflation dynamics and the new Keynesian Phillips curve in four Central European countries. *Emerging Markets Finance & Trade*, *47*(5), 71–100.
47. Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data, second edition*. Cambridge, Massachusetts: The MIT Press.
48. Zigan, K., & Zeglat, D. (2010). Intangible resources in performance measurement systems of the hotel industry. *Facilities*, *28*(13/14), 597-610.

Vladimir Vasić

was born in Belgrade in 1969. He graduated from the Faculty of Economics in 1994, completed a Master course in Statistical Analysis in 1999 and acquiring a PhD title in Statistics at the Faculty of Economics in 2003. He is employed as Assistant Professor at the University of Belgrade – Faculty of Economics (Department of Statistics and Mathematics) at courses: Multivariate Analysis and Nonparametric Statistical Methods. He has dealing with the implementation of data mining and data science in the business economy. He is involved in many commercial projects in the region. He is the author of over 20 papers in journals on the SCI and SSCI lists.

Branislava Hristov Stančič

is Teaching Assistant at the Faculty of Economics, University of Belgrade. She lectures in Business Economy and Management (Management of tourism Enterprises and Marketing in Tourism) and in the field economic policy and development (Tourism Economics). Her academic interest is focused on management in tourism and hospitality with special emphasis on Revenue management in hotel industry. She has published numerous scientific papers in domestic journals.



Bojan Zečević

graduated and obtained his master's degree and doctorate at the Faculty of Economics, University of Belgrade. He was elected full professor in 2017. Main areas of interest and research: tourism economics, management and marketing in tourism and hotel companies, planning and development of tourist destinations. Courses lectured on the undergraduate level: Tourism Economics, Marketing in Tourism and Hospitality and Revenue Management in Tourism and Hospitality. He also teaches Master and PhD courses. He is a member of the Scientific Society of Economists and the Serbian Marketing Association (SEMA). One of the leading regional advisors in the tourism and hotel industry with expertise in tourism research and consulting gained on numerous projects with clients such as ministries, local governments, international organisations, development agencies and hotel investors.