Big Data Research 40 (2025) 100520

Contents lists available at ScienceDirect

Big Data Research

journal homepage: www.elsevier.com/locate/bdr

Complex data in tourism analysis: A stochastic approach to price competition

Giovanni Angelini [a], Michele Costa [a], Andrea Guizzardi [b,*]

[a] Department of Economics, University of Bologna, Piazza Scaravilli 2, 40126 Bologna, Italy
[b] Department of Statistical Sciences, University of Bologna, Via Belle Arti 41, 40126 Bologna, Italy

ARTICLE INFO

Keywords:
Complex data
Multivariate stochastic approach
Advance booking

ABSTRACT

This study examines pricing strategies and decision-making processes in the hospitality industry by analyzing "ask" prices on online travel agencies (i.e., the rates at which hoteliers are willing to sell their rooms). We face the challenge of modeling a continuous flow of big data organized as "time series of time series," where daily seasonality and advance bookings intersect. Our research combines insights from tourism, quantitative methods, and big data to improve pricing strategies, contributing to both theory and practice in revenue management. Focusing on Venice, we analyze price competition as a multivariate stochastic process using a Structural Vector Autoregressive (SVAR) approach, aligning with modern dynamic pricing algorithms.

The findings show that time-based pricing strategies, which adjust based on the day of arrival and booking, are more important than room features in setting hotel prices. We also find that price changes have a non-linear and decreasing effect as the booking date approaches. These insights suggest that hotels could create more advanced pricing strategies, and policymakers should consider these factors when addressing the challenges related to overtourism.

We study the complex competitive relationships among heterogeneous service providers with an approach applicable to any market where consumption is delayed relative to purchase time. However, we highlight that the quality and accessibility of information in the tourism sector are key aspects to be considered when using big data in this industry.

1. Introduction

In the current "Big Data" era, technological advancements provide new opportunities to explore the vast amount of available information. The tourism sector was among the first to adopt digitalization. In the early 1960s, the introduction of computer reservation systems [28], marked a significant shift from American Airlines's basic inventory control and improved communication with distributors. This innovation allowed for real-time updates on routes, seat availability, and pricing. More recently, thanks to technological advancements such as smartphones, drones, wearables, improved connectivity, and big data, digitalization in the sector has accelerated. This has shifted many scholars' perspectives on information technology in tourism research from primarily a marketing tool to a knowledge creation tool [34].

On the demand side, the large volume of data generated by tourists - such as reviews, pictures, and texts - has enabled the development of a wide range of applications, from studying tourist motivations to predict travel flows, with a common advantage: the ability to produce timely and granular insights not possible with traditional statistical sources [26].

On the supply side, the rise of Online Travel Agencies (OTA) like Booking.com and Expedia has fundamentally transformed the travel and tourism industry. These platforms have made pricing policies and business strategies, such as product customization and marketing, public and thus more transparent [29].

In this study, we point to demonstrate that the convergence of advanced data availability and sophisticated analytical methods can improves our understanding of the competitive landscape, particularly how pricing strategies evolve in response to market forces. By leveraging the wealth of big data from OTAs, our aim is to develop a framework for more informed decision-making, helping firms to set competition strategies and optimize their pricing tactics effectively. Moreover, we aim to suggest a quantitative methodology that can tackle the challenge posed by these multidimensional big data where we relate each price listed for

* Corresponding author.
E-mail addresses: g.angelini@unibo.it (G. Angelini), michele.costa@unibo.it (M. Costa), andrea.guizzardi@unibo.it (A. Guizzardi).

https://doi.org/10.1016/j.bdr.2025.100520
Received 13 September 2024; Received in revised form 11 February 2025; Accepted 12 March 2025
Available online 13 March 2025

an overnight stay on a specific arrival day to a time series of prices posted throughout the booking window.

Existing literature often treats advance booking as a linear and deterministic factor, implicitly assuming a constant pick-up rate. However, a more accurate description of price trajectories across booking periods can be achieved through a stochastic approach [22], capturing the interdependence of prices set during different booking periods. This method aligns more closely with modern dynamic pricing algorithms [30] and allows consideration of price fairness throughout the booking period.

Inspired by methodologies widely used in social sciences and financial markets, where data abundance is well-established, we adopt a Structural Vector Auto-Regression (SVAR) approach to model the intertemporal pricing structures embedded in these 'time series of time series.

Theoretically, our proposal assumes that room rates follow a stochastic dynamic process [11], driven by hoteliers' expectations about demand fluctuations for different arrival days. We put forward the idea that this causal chain should also account for the advance booking period, as expectations about occupancy rates on the arrival day change in response to incoming bookings and cancellations. Consequently, our focus is on modeling the sequence of prices posted for different arrival days and booking periods, while considering other relevant exogenous variables affecting product differentiation, such as rate fences or room features.

Moreover, our SVAR approach allows us to "visualize" Impulse Response Functions (IRFs) that quantify the dynamics of rates over both the advance booking period and the days following the arrival day, after a simulated rate adjustment due to unexpected bookings or cancellations, and/or a change in competitor behavior. This provides insights into the expected pattern of pricing adjustments after a shock, which is a critical aspect of revenue management strategies.

We apply our model to a dataset of 102 hotels in Venice, a globally renowned cultural destination, focusing on daily pricing strategies and time-lagged interactions over a 28-day booking window. However, this approach is applicable to studying competitive behavior in any market where consumption is delayed relative to purchase time. As such, we address a significant challenge for both academics and practitioners by investigating the complex competitive relationships among highly heterogeneous service providers.

The structure of this paper is as follows. In the second section, we summarize the main challenges related to the use of big data in tourism analysis. The third section presents the model, followed by an illustration of the data in the fourth section. The discussion of the results is covered in section five, and finally, the conclusions are drawn in section six.

2. Complex data challenges in tourism analysis

There are several important aspects to consider when working with big data in tourism, namely the lack of a solid theoretical foundation, challenges related to data quality, and the need for continuous attention to social, technological, and regulatory changes. Big data can uncover real-time patterns that traditional economic theories often fail to explain. As a result, it may seem easy to argue that a data-driven approach could replace theoretical models. However, this is not always the case in the tourism sector, where underlying theoretical aspects remain relevant [25]. To address this issue, it is essential to integrate insights from theories in fields such as psychology, communication, and information processing [33]. These include, among others, Social Learning Theory [3] and Emotional Contagion Theory [18], which have been used to leverage different types of big data in tourism studies.

Data quality challenges mainly arise from issues of comparability and reliability, which complicate the integration of diverse high-resolution data or their combination with official statistics. Private companies, which typically control such data, have little incentive to share their raw information. Thus, the reliability of data often depends on the company's reputation [20]. Consequently, public entities struggle to use aggregated data from private companies to develop high-resolution statistics. Demunter [12] points out that many of the legal and technical challenges faced today echo those encountered in the past when integrating public and administrative records into official statistics. However, the current landscape is more complex due to the private ownership of data, which introduces not only technical and legal issues but also political concerns.

Equally essential is the need for continuous attention to social, technological, and regulatory changes. The tourism and hospitality industry is characterized by rapid transformations in how tourists and businesses adopt new technologies or react to economic and social shifts, as well as unexpected events. The COVID-19 pandemic is a notable example, drastically altering market dynamics and consumer expectations [32] as well as providing new opportunity to gather information from unstructured data [13]. Technological advancements in search and indexing algorithms, as well as in machine learning, are reshaping how information is converted into usable metrics, affecting both data accessibility and the ability to integrate different sources. Not all big data available today will remain accessible tomorrow. Building market analysis systems based on big data requires creating information systems capable of automatically obtaining real-time data, making it comparable when necessary, and storing it for future use [7]. Additionally, confidentiality and data protection laws of many countries prohibit the storage, production, or dissemination of information that could compromise individuals' privacy [27]. These laws are primarily national and often inconsistent with each other, making them ineffective for regulating the use of data that is, by definition, globally accessible. This is particularly important in light of the growing use of artificial intelligence tools, which presents further significant challenges.

2.1. Hotels pricing competition through the lens of big data

The tourism industry has long been a leader in technological innovation, starting with the automation of booking processes through the SABRE system developed by American Airlines and IBM in the 1960s. This system revolutionized booking management by providing real-time flight information, greatly enhancing operational efficiency [2]. The emergence of OTAs, particularly Expedia, marked a pivotal shift, as industry (big) data became publicly available. This transparency allowed consumers and businesses alike to gain deeper insight into firms' competitive behavior. This evolution represents a critical moment for the tourism sector, where data transitioned from merely operational tools to public assets for strategic decision-making in an increasingly competitive landscape. OTAs provide real-time access to pricing information, allowing both hotels and consumers to monitor how competitors adjust their prices in response to market changes [6]. This revolution in data availability has been further boosted by advancements in modern Property Management Systems (PMS), which facilitate the real-time management of rates, inventory, and customer loyalty, enabling businesses to react swiftly to market (observed) conditions. As a result, pricing strategies and decision-making in the travel and tourism sector are increasingly influenced by complex data. Hotels now can publish "ask" prices on OTAs similarly to how stock prices are listed, creating a continuous flow of information that reflects decision-makers' expectations regarding their own booking curves.

Just as stock prices, hotel prices are publicly displayed and subject to constant fluctuations based on demand. This continuous flow of information allows both hotels and consumers to observe how competitors adjust their prices in response to market dynamics. The literature on dynamic pricing supports this real-time adjustment view, highlighting that pricing is primarily influenced by demand curves along the advance booking timeline [17]. Specifically, if hotels are selling rooms at a steady rate, they may increase prices. Conversely, if they experience cancellations (or slower sales), they may lower prices to attract bookings. That's why, if we look at the price competition on OTAs through a financial lens, it becomes possible to see hotel rooms as shares owned by an investor (the hotelier) who wants to sell them over the time. The OTAs act like an order book, showing only the first-level ask prices without revealing the corresponding first-level bid prices (how much customers are willing to pay for those rooms).

The prices that firms publish on OTAs at different stages of advance booking offer a unique opportunity to understand companies' competitive strategies based on their 'real-life' behavior. This is especially relevant since firm behavior theory itself underscores the value of such an approach. Firms are often overwhelmed by the vast amount of available information, leading managers to adopt limited, local-centric searches for alternatives [10]. As a result, pricing decisions are frequently made based on incomplete information. Consequently, imitation becomes a common strategy to navigate these challenges, particularly when firms perceive a performance gap [24]. These gaps can trigger experiential learning: when performance is close to expectations, firms tend to learn from their own experiences, whereas when performance diverges significantly, they look to others for insight [4]. Moreover, organizational goals may be unclear due to the absence or instability of strategic development skills [9], or they may differ among firms. Bigné [6] highlights that variability in pricing strategies often stems from different business orientations; some firms prioritize revenue growth through high-quality service, while others focus on cost reduction. In this context, customers tend to seek perceived value over the lowest price [8], adding complexity to profit-maximization efforts.

This theoretical background leads us to believe that decision-making can be effectively studied by focusing on raw data from OTAs, assuming that room rates follow a stochastic dynamic process (see [11]) depending on both the time dimensions (i.e., day of consumption and advance purchase).

In this complex landscape, big data from OTAs becomes an invaluable resource. By analyzing price dynamics across multiple timeframes - such as seasonality and advance booking – we advance existing literature, leveraging publicly available data to shed light on competitive behavior in the tourism market. This type of dataset poses new statistical challenges. Each asking price for an overnight stay on a specific day (referred to as daily-seasonality) corresponds to a time series of asking prices across the booking window. We argue that both dimensions - booking period and daily-seasonality - are crucial for fully understanding competition and must be considered simultaneously. Seasonality, including calendar effects, strongly influences price competition, especially in metropolitan areas and popular destinations where events, holidays, or weekends can significantly affect demand. Lower-quality hotels tend to offer deeper discounts during low seasons, while higher-priced hotels are generally less sensitive to demand fluctuations, indicating asymmetric seasonal effects based on quality differentiation [31]. However, during peak periods (e.g., fairs, major events), lower-scale hotels can attract last-minute customers willing to pay higher prices due to room shortages [16].

It is also important to note that big data on the supply side offers two key advantages over web search data. First, it consists of "raw data" directly posted by competing firms on OTAs, free from the biases that often affect secondary data sources like Google Trends, which can prevent time series comparability due to filtering and algorithmic updates [21]. Second, as noted by Guizzardi et al. [14], these data serve as an effective example of implicit shared knowledge about firms' pricing competition tactics in real-life contexts, facilitating intelligent interactions between the territory and its inhabitants, thus providing a starting point for a smart decision-making process.

3. The model

Competition arises when firms operate within shared markets, necessitating an understanding of their pricing strategies in a context where consumption and purchase do not occur simultaneously. In an industry characterized by intense competition and numerous small enterprises, pricing strategies are often implicit and rarely documented. This lack of clarity necessitates sophisticated analytical tools to interpret competitors' behaviors and inform strategic decision-making. Our proposal is to adopt a stochastic approach based on a Structural VAR model, which accounts for the interdependence between various advance bookings and incorporates a wide range of seasonal factors, such as weekends, trade fairs, and Carnival, while controlling for room characteristics.

Traditional univariate time series models, commonly used in tourism and hospitality literature, are not sufficiently general to model the two temporal dimensions available in the big data from OTAs. But when we choose the best VAR framework, we also consider that the hospitality literature [19] indicates that hotels' approaches to dynamic pricing is influenced by many unmeasurable characteristics on OTAs, such as size, chain affiliation, and level of digitalization (e.g., revenue management tools, distribution channel management, frequency of application). This evidence led us to estimate a different model for each hotel (i.e., a different distribution of residuals) even if we are aware of methodologies that would allow to model jointly also the spatial dimension, i.e., the hotel, as panel-SVAR.

While VAR approaches are not new in tourism literature, they are more commonly applied to demand forecasting rather than price modelling. One notable exception is Bergantino et al. [5], who employ a panel-VAR model to describe price dynamics and competition in the Italian airline and railway markets. However, their panel-VAR methodology becomes impractical for a large cross-sectional dimension. Typically, the case with competing hotels in a destination, require restrictions on the coefficient that are hard to be imposed, leading to results with poor economic interpretation. We focus on $Y^i_{t-j,t}$, the logarithm of the Best Available Rates (BARs) for the i th hotel, with $i = 1, 2, \ldots, N$, available on the internet in $t-j$, for the arrival date t (j indicates the advance booking and J is the maximum advance booking considered). According to Guizzardi et al. [15] and Angelini et al. [1], for each hotel we specify a SVAR model for the $J \times 1$ vector Y^i_t which has the following representation (we drop the hotel superscript i):

$$Y_t = C + AY_{t-1} + \Gamma X^s_t + \Theta X^f_t + \eta_t, \quad \eta_t \sim N(0, \Sigma) \tag{1}$$

$$\eta_t = B\varepsilon_t, \quad \varepsilon_t \sim N(0, I_k), \quad BB' = \Sigma \tag{2}$$

where X^s_t is a vector with exogenous seasonal variables and X^f_t is the vector with room-dependent variables.

It is important to stress how our specification is flexible with respect to different formulations of vectors X^s_t and X^f_t. Both A and B matrices are constrained to be upper triangular. This allows us to assume that significant statistical relationships can be inferred only from larger to smaller advance bookings, coherently with the chronological ordering of the endogenous variables.

The estimation of the coefficients in model (1) (2) is performed under the assumption of Gaussian errors via constrained maximum likelihood, which is asymptotically equivalent to a feasible generalised least squares estimator [23].

An important feature of our model is its ability to evaluate the dynamic response of BARs to a shock in the j-th advance booking using Impulse Response Functions (IRFs). Given the structure of the data, the identification of structural shocks using a Cholesky decomposition is a natural choice.[1] To assess how shocks propagate both along the price trajectory and between subsequent arrival dates, we compute the IRFs as follows:

[1] The data indeed consist of room prices for different advance bookings, where the temporal order is clearly defined. This temporal ordering allows us to impose a recursive structure in which shocks to prices at earlier booking periods can affect prices at later periods, but not the other way around.

$$IRF_j(t+h) = A^h b_j, \quad h = 0, 1, ..., H, \quad j = 0, 1, ...J \tag{3}$$

where b_j is the *j-th* column of the matrix of structural coefficients B and H is the maximum horizon considered. The result is a triangular matrix of IRFs, with rows representing advance bookings and columns representing shocks. Each IRF can, in turn, be represented on a Cartesian plane, where the y-axis shows the (log) BARs and the x-axis represents the day after the arrival date. This line represents how a simulated price shock (equal to one standard deviation) propagates along the following arrival days, at a certain advance booking.

Overall, the coefficients Γ, Θ and B provide a broad and valuable range of information, enabling a deeper understanding of how various factors influence pricing dynamics. This comprehensive approach offers insights critical for effective decision-making in highly competitive markets.

4. The data

Big data from OTAs are a valuable source of information. By examining price dynamics across multiple time dimensions - seasonality and advance purchasing - we can leverage publicly available data to shed light on competitive behavior in the tourism market. This analysis is particularly relevant for sectors where consumption is time-sensitive, such as accommodations, event tickets, and transportation, providing valuable insights for decision-makers navigating a complex competitive landscape.

Our dataset includes daily price observations for 102 hotels highly rated (3 stars or more), located in Venice, across multiple time horizons, encompassing various advance booking periods and seasonal patterns. The large volume of data, coupled with the intricate nature of pricing strategies influenced by factors such as demand fluctuations, competitive responses, and external events, creates a complex framework that calls for advanced analytical tools to uncover meaningful trends and relationships. The ability to track competitive pricing behavior at single business level, offers a unique opportunity to explore how hotels adjust their prices in a dynamic and highly competitive market like Venice, which is a destination experiencing periodic excess demand (overtourism), providing insights that are otherwise difficult to capture through traditional data sources. Additionally, the monitored hotels are expected to compete each-other, as most of them are located in a small radius (only 800 meters) from the San Marco Sestiere (one of the six districts in the city of Venice).

We consider 3 advance bookings, namely j = 0,7,28, that corresponds to the prices offered in the last minute ($j = 0$) and at two early bookings periods of a week ($j = 7$) and a month ($j = 28$). Even though the proposed approach can be used to study the pricing behaviour in any booking window, longer horizons are excluded because we expect that competition is fiercer at the last minute [16]. With the aid of a web-scraping software we simulate a customer searching for a room at each of the above three different advance booking in a window from January 1st, 2019 to February 29th, 2020 (425 observations) in order to exclude the covid-19 period where competition was strongly affected by the ban on (international) travel. We scrape the BAR, to ensures the highest homogeneity with respect to possible (unobservable) product differentiation practices.

The exogenous seasonal variables vector X_t^s includes 6 dummy variables ($Fairs_t; Bien_t; Carn_t; Aug_t; WE_t; Hol_t$), which take value 1 in correspondence of the most visited fairs/events, the Biennale and the carnival periods, Saturdays and Sundays, August and bank holidays.

As exogenous room-dependent variables X_t^f we consider two dummies ($Breakf_t$ and $Cancel_t$ which takes value 1 if the breakfast is included and if the room can be cancelled with no penalties), and two variables $Items_t$ and $Price_t$, i.e., the number of features of the offered room (air conditioning, balcony, etc.…) and the number of prices published by the same hotel. These variables serve as proxies for the quality offered and the stock of available rooms for day t, respectively.

Table 1 presents some descriptive statistics for hotel pricing. The average BARs are higher in the early booking.

Revenue managers were successful in selling the top-rated rooms at higher advance booking as confirmed by the weakly increasing trend in the 10th and the 90th percentiles. Policies related to price differentiation based on breakfast offerings remain largely unchanged in the booking window and (as expected), the quote of free cancellations options strongly decrease in the last minute.

Fig. 1 shows the price distributions for the three advance bookings considered. From the distributions, some differences can be observed. The tail of the distribution for the advance booking 0 is the thickest, indicating that as the booking date approaches, the probability of observing extreme values is higher. In contrast, the distribution 28 days before the booking date shows the least variability and is more concentrated around the average values. The evidence is consistent with the hypothesis that (in the early booking phase, when you are far from the arrival date and likely have more room availability) your performance is close to your own expectations, and hotels compete by setting rates based on their own experience. In the last-minute, however, the effect of stochastic demand (i.e., unexpected bookings and cancellations) becomes more evident, and combined with the approach of the booking date, the risk of unsold rooms increases, and this makes it more likely that hoteliers observe performance diverging significantly from their expectations, leading managers to adopt limited, local-centric searches for alternative pricing policies [10], and imitation becomes a common strategy to compete. In fact, in addition to the thicker right tail (with more rooms offered at higher prices), Table 1 shows that prices become more extreme in terms of discounts as well. Specifically, the 10th percentile drops to €68 (with rooms offered below €45 in 1% of cases). These are the least equipped rooms, with an average of 21.6 items compared to 23.8 items for the 90th percentile demonstrating that, when price discrimination is limited, product quality discrimination (advanced revenue management) takes over.

5. Results and discussion

In Table 2, we present a summary of the estimated coefficients for the room-dependent variables X_t^f, focusing only on those that are statistically significant (alpha < 20%) in at least one-third of the monitored hotels.

As expected, all the effects on the BAR are positive. The impact of advance booking is particularly evident. Offering a cancellation policy tends to increase the price, especially for last-minute bookings. However, it is precisely for $j = 0$ (last-minute) that we observe the highest variability, indicating that, in order to increase the occupancy rate, the possibility of cancellation is sometimes combined with a significant price increase. Selling a room with breakfast has a strong positive effect on the price, especially for early bookings.

On the contrary, the effect of the number of features (a proxy for the offered quality) becomes clear only in the very last-minute, where more luxurious rooms are offered at inflated prices. The lack of a clear relationship between quality and price in early bookings suggests that not all hotels coordinate rate fence management with quality management. This simplification of pricing strategy reflects the limited and locally

Table 1
Best Available Rate descriptive statistics.

	$j = 0$	$j = 7$	$j = 28$
mean(BAR)	214.7	234.0	243.9
std(BAR)	243.1	247.3	241.6
10–90perc(BAR)	68–397	77–415	85–415
%cancellation	2.42%	15.13%	15.19%
%breakfast	60.09%	61.09%	62.88%
Mean(#items)	22.8	22.8	22.7

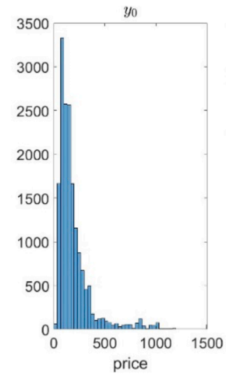

Fig. 1. Distribution of prices for the three advance booking considered.

focused search for alternative actions resulting in final pricing decisions that become routine practices and/or imitation [10].

In comparison to the coefficients estimated for the room-dependent variables, the estimated coefficients for X_t^s, reported in Table 3, show less variability across hotels and throughout the advance booking period. This suggests that seasonality plays a role in rate differentiation regardless of both the hotels' propensity to adopt dynamic pricing algorithms and their heterogeneity. As for the room-dependent variables, all effects on the BAR are positive.

It is interesting to note that the impact of all variables measuring seasonality follows a U-shaped pattern with respect to the advance booking period, reaching a minimum at $j = 7$ (one week before arrival). The only exception to this pattern is the weekend WE_t, where the shape is reversed. It should be highlighted that the seasonal effect associated with weekends has the highest coefficients compared to other estimates. Additionally, it is almost always significant, meaning that the majority of revenue managers consider it as a lever to differentiate prices. This is likely because is the most common seasonal pattern, which enables easy observation of its influence on the pick-up curve and facilitates price optimization through adjustments in room rates.

An important aspect of our SVAR proposal is its ability to compute Impulse IRFs as outlined in (3). This allows us to evaluate the impact of a price shock across the advance booking. In the first row of Fig. 2, we display $IRF_j(t+h)$, which shows the price response on the arrival day t and the following h days, resulting from a simulated one-standard deviation price increase at advance booking j.

For example, if t corresponds to June 30th the second element of $IRF_{28}(t+h)$ (shown in the last column, second row) measures the impact on the ask price posted on June 23rd for a stay on June 30th (and the subsequent days along the x-axis). This impact is caused by the price adjustment posted by the hotelier on June 2nd ($j = 28$), which was influenced by an unexpected cancellation/reservation observed on that date.

All IRFs show that the BAR adjustment caused by a shock (e.g., a change in occupancy rates for a fixed arrival date) decreases monotonically as a negative exponential function of the advance booking. This implies that for all the advance bookings considered, the response to a price shock propagates with diminishing effects on prices for the same arrival date published in subsequent advance bookings, until it disappears. These effects become insignificant after 2 or 3 days depending on the advance booking considered, as shown by the confidence intervals (the shaded areas). Specifically, the shaded areas include zero after two days if last-minute bookings are considered, while the effect of the shock remains positive for 3 days after the arrival date for unexpected cancellations or reservations made more than 28 days in advance. This evidence is consistent with the average length of stay in Venice, which in the period considered is 2.5 days.

Overall, these results clearly indicate the use of pricing tactics to mitigate last-minute decreases in occupancy rates when faced with unpredictable events such as cancellations. Symmetrically, strategy becomes more important with longer advance booking, as hoteliers aim to keep price trajectories stable or prevent them from decreasing, while also maintaining low variance. This helps to deter speculative behavior by customers (i.e., canceling and rebooking) and to address customers' internal reference price. Such adaptability reflects dynamic pricing

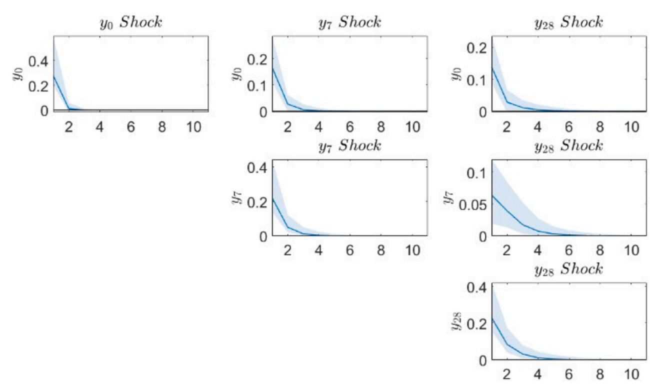

Fig. 2. Impulse Response Function to one standard deviation price shock.

Table 2
Statistically significant (alpha < 20%) coefficients in at least one-third of the monitored hotels: mean, range, and observed frequency for the room-dependent variables.

	Mean	10–90 perc	Freq		Mean	10–90 perc	Freq
Cancel_0	0.323	0.088–0.714	43%	Beakf_0	0.483	0.057–1.752	50%
Cancel_7	0.222	0.045–0.408	41%	Beakf_7	0.651	0.033–2.436	47%
Cancel_28	0.234	0.064–0.487	36%	Beakf_28	0.688	0.057–2.480	59%
Items_0	0.024	0.003–0.051	45%				

Table 3
Statistically significant (alpha < 20%) coefficients in at least one-third of the monitored hotels: mean, range, and observed frequency for the seasonal variables.

	Mean	10–90 perc	Freq		Mean	10–90 perc	Freq
Fairs_0	0.095	0.031–0.164	66%	Hol_0	0.091	0.032–0.170	74%
Fairs_7	0.085	0.027–0.169	78%	Hol_7	0.084	0.029–0.139	75%
Fairs_28	0.090	0.028–0.148	67%	Hol_28	0.086	0.030–0.153	79%
Bienn_0	0.142	0.058–0.242	93%				
Bienn_7	0.122	0.044–0.214	88%	WE_0	0.213	0.089–0.337	99%
Bienn_28	0.143	0.044–0.272	83%	WE_7	0.243	0.110–0.356	99%
Carn_0	0.210	0.079–0.408	60%	WE_28	0.210	0.069–0.333	97%
Carn_7	0.201	0.082–0.338	60%				
Carn_28	0.216	0.093–0.416	57%	Aug_0	0.084	0.038–0.142	42%

management, allowing hoteliers to maximize revenues in response to demand fluctuations.

The dynamics highlighted by the IRFs at different advance booking periods reinforce the conclusion that hoteliers forecast occupancy rates at various stages of advance booking by learning from their own experiences, or at least with a local-centric perspective. In fact, the average response to an early booking shock follows a convex trajectory, peaking at the very moment when the hotelier records the increase in occupancy rate and reaching a minimum 2–3 days before the arrival date. For last-minute bookings ($j = 0$), the effect shows a significant increase, supporting our conclusion about a tactical approach in response to unpredictable events like last-minute cancellations and walk-in guests.

6. Conclusion

In this study, we conduct a comprehensive analysis of pricing strategies in a popular tourist destination like Venice. By applying a Structural VAR (SVAR) methodology, we explore the relationship between seasonality and advance bookings, treating price discrimination patterns as a multivariate stochastic process. This approach allows us to model not only the immediate price variations resulting from new bookings or cancellations, but also the interdependencies between different advance bookings, that is, how hoteliers manage customer perceptions about relative prices in an overtourism destination like Venice.

Moreover, the SVAR setting allows us to assess how competitive dynamics evolve in relation to seasonality, with calendar effects treated as fixed effects, whose value is estimated for each hotel. Our findings reveal that time-based pricing is central to understanding price competition strategies. Most hotels use simplified pricing methods, where seasonality - specifically the day of arrival - takes precedence over room-dependent variables. However, these strategies do not always, nor uniformly, depend on the advance booking period. In particular, August, the non-business season, positively affects the BARs, but not in a dynamic way. Conversely, events like bank holidays, trade fairs, Carnival, and the Biennale are more sensitive to advance booking. The impact of these variables follows a U-shaped pattern with respect to the advance booking period, reaching a minimum at $j = 7$ (one week before arrival). The exception is weekends, where the pattern is reversed due to the limited presence of business customers. Lower BARs in early bookings are typically designed to attract leisure travellers, who tend to book in advance and are more sensitive to online discounts, while higher last-minute rates may deter walk-in guests, leaving rooms unoccupied.

Few hotels consider room-related factors such as availability, quality, and rate fences to modulate price competition, and primarily in the last minute. Surprisingly, prices are often higher during early bookings, an outcome that leaves room for speculative behaviours like cancellations and rebooking, which is mitigated by offering a low frequency of refundable BAR rates. This unexpected trend may be attributed to the dynamics of overtourism, where hoteliers are confident in achieving expected occupancy rates. As a result, price competition intensifies in the last-minute period, with lower-quality rooms being the last to sell, aligning with dynamic pricing principles. This local-centric competition tends to focus on imitation rather than true differentiation.

Furthermore, our findings indicate that the response to simulated price shocks weakens as the booking horizon extends, lasting from 3 days in early bookings to 2 days in last-minute scenarios.

Several managerial implications arise from these insights. Hotels with a higher capacity for horizontal differentiation (i.e., segmentation of customers) can leverage high demand pressure to risk leaving premium rooms unsold until the last minute. This approach allows them to stand out in terms of quality during the periods of highest competition. In the last-minute context, the perceived value of differentiated assets increases due to limited availability. For instance, a room with a canal view balcony can command significantly higher prices closer to the arrival date, as its availability becomes highly desirable for guests seeking a unique experience.

Moreover, the relatively simple revenue management techniques currently employed, which are largely based on daily seasonality, present significant opportunities for hotels willing to develop more complex pricing strategies. By doing so, they can more accurately forecast the pick-up curve, even in overtourism destinations, without necessarily relying on advanced revenue management algorithms or highly skilled professionals.

Finally, analyzing price competition through the lens of advance bookings can help policymakers and businesses better address and manage the challenges posed by a massive and uncoordinated influx of tourists. Understanding early booking pricing patterns can inform about future demand peaks, both positive and negative, as price and demand dynamics are closely correlated. The daily occupancy rate indicated by hoteliers through pricing for future dates offer crucial information, not only for short-term pricing tactics but also for managing resources, such as transportation, security, and urban space, which are shared by both tourists and residents. This allows for more efficient use of public resources and better budget management.

This study acknowledges some limitations. It focuses solely on one city and analyzes only the lowest price for a single-person overnight stay. The reliance on OTA data and the sampling method chosen means we do not account for important time-related factors, such as hotel reputation, occupancy rates, offline sales strategies, or visibility on other platforms like Expedia. However, the methodology is flexible enough to incorporate these variables in future research and can be applied to other industries with fixed capacities and advance bookings, such as event tickets or transportation. Additionally, our methodology is limited to accommodations that consistently publish online prices and requires the development of a custom data collection process, as no comprehensive historical database of hotel rates currently exists. This highlights a broader issue with the quality and accessibility of information in the tourism sector, underscoring one of the key challenges of using big data in this industry.

CRediT authorship contribution statement

Giovanni Angelini: Writing – review & editing, Writing – original draft, Validation, Supervision, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Michele Costa:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Methodology, Investigation, Formal analysis, Conceptualization. **Andrea Guizzardi:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] G. Angelini, M. Costa, A. Guizzardi, Price competition and overtourism: a stochastic approach leveraging complex data, Italian Stat. Soc. Ser. Adv. Stat. 2 (2024) 1–6.
[2] S. Baker, K. Baker, Mine over matter, J. Bus. Strat. 19 (4) (1998) 22–26.
[3] A. Bandura, Social learning theory of aggression, J. Commun. 28 (3) (1978) 12–29.
[4] J.A. Baum, K.B. Dahlin, Aspiration performance and railroads' patterns of learning from train wrecks and crashes, Organ. Sci. 18 (3) (2007) 368–385.
[5] A.S. Bergantino, C. Capozza, M. Capurso, Pricing strategies: who leads and who follows in the air and rail passenger markets in Italy, Appl. Econ. 50 (2018) 4937–4953.

[6] E. Bigne, A. Decrop, Paradoxes of postmodern tourists and innovation in tourism marketing. The Future of Tourism, Springer, 2019, pp. 131–154.
[7] D. Buhalis, R. Leung, Smart hospitality - Interconnectivity and interoperability towards an ecosystem, Int. J. Hosp. Manage 71 (2018) 41–50, https://doi.org/10.1016/j.ijhm.2017.11.011.
[8] H.C. Chung, N. Chung, J.-Y. Kim, How do online hotel consumers perceive room rates? J. Vacation Market. 28 (3) (2021) 350–365.
[9] M.D. Cohen, J.G. March, J.P. Olsen, A garbage can model of organizational choice, Adm. Sci. Q. (1972) 1–25.
[10] R. Cyert, J. March, Behavioral Theory of the Firm, Wiley, 1963.
[11] R. Croes, K.J. Semrad, Discounting works in the hotel industry: a structural approach to understanding why, Tourism Economics 18 (2012) 769–779.
[12] C. Demunter, Tourism Statistics: early Adopters of Big Data?. Measuring Sustainable Tourism, 2017, p. 25.
[13] F. Giambona, C. Galluccio, A. Petrucci, Tourism sustainability in news media data, in: Book of Short Papers 2nd Italian Conference on Economic Statistics (ICES 2024) Statistical Analysis of Complex Economic Data: Recent Developments and Applications, 2024, pp. 105–109.
[14] A. Guizzardi, F.M.E. Pons, G. Angelini, E. Ranieri, Big data from dynamic pricing: a smart approach to tourism demand forecasting, Int. J. Forecast. 37 (3) (2021) 1049–1060.
[15] A. Guizzardi, G. Angelini, F. Pons, Does advance booking matter in hedonic pricing? A new multivariate approach, International Journal of Tourism Research 22 (2020) 277–288.
[16] A. Guizzardi, F. Pons, E. Ranieri, Advance booking and hotel price variability online: any opportunity for business customers? Int. J. Hosp. Manage 64 (2017) 85–93.
[17] A. Guizzardi, L.V. Ballestra, E. D'Innocenzo, Hotel dynamic pricing, stochastic demand and covid-19, Ann. Tour. Res. 97 (2022) 103495.
[18] E. Hatfield, J.T. Cacioppo, R.L. Rapson, Emotional contagion, Curr. Dir. Psychol. Sci. 2 (3) (1993) 96–100.
[19] S. Ivanov, G. Del Chiappa, A. Heyes, The research-practice gap in hotel revenue management: insights from Italy, Int. J. Hosp. Manage 95 (2021) 102924.
[20] R. Kitchin, Big Data and Official Statistics: opportunities, Challenges and Risks, SSRN Electronic Journal (2015), https://doi.org/10.2139/ssrn.2595075.
[21] D. Lazer, R. Kennedy, G. King, A. Vespignani, The parable of Google Flu: traps in big data analysis, Science (1979) 343 (6176) (2014) 1203–1205.
[22] M. Lee, Modeling and forecasting hotel room demand based on advance booking information, Tour. Manage 66 (2018) 62–71.
[23] H. Lutkepohl, New Introduction to Multiple Time Series Analysis, Springer, Berlin, 2005.
[24] J.G. March, H.A. Simon, Organizations, John Wiley & Sons, 1993.
[25] J.A. Manzanec, Hidden theorizing in big data analytics: with a reference to tourism design research, Ann. Tour. Res. (2020) 83.
[26] M. Ortu, M. Romano, A. Carta, Semi-supervised topic representation through sentiment analysis and semantic networks, Big Data Research 37 (2024).
[27] H. Padrón-Ávila, R. Hernández-Martín, Tourist tracking techniques as a tool to understand and manage tourism flows, in: H. Séraphin, T. Gladkikh, T. Vo Thanh (Eds.), Overtourism, Springer International Publishing, 2020, pp. 89–105, https://doi.org/10.1007/978-3-030-42458-9_6.
[28] W.R. Plugge, M.N. Perry, American Airlines'" Sabre" electronic reservations system, in: Papers presented at the May 9-11, 1961, western joint IRE-AIEE-ACM computer conference, 1961, pp. 593–602.
[29] D. Provenzano, R. Baggio, E-tourism economics: editorial for the special issue, Tourism Economics 26 (6) (2020) 847–852.
[30] D. Shin, S. Vaccari, A. Zeevi, Dynamic pricing with online reviews, Manage Sci. 69 (2) (2023) 824–845.
[31] X. Wang, J. Sun, H. Wen, Tourism seasonality, online user rating and hotel price: a quantitative approach based on the hedonic price model, Int. J. Hosp. Manage 79 (2019) 140–147.
[32] E.H. Wu, J. Hu, R. Chen, Monitoring and forecasting COVID-19 impacts on hotel occupancy rates with daily visitor arrivals and search queries, Current Issues in Tourism 25 (3) (2022) 490–507.
[33] D.C. Wu, S. Zhong, J. Wu, H. Song, Tourism and hospitality forecasting with big data: a systematic review of the literature, J. Hosp. Tour. Res. (2024), 10963480231223151.
[34] Z. Xiang, From digitization to the age of acceleration: on information technology and tourism, Tour. Manage Perspect. 25 (2018) 147–150.