Journal of the Royal Statistical Society Series A: Statistics in Society, 2023, **00**, 1–28
https://doi.org/10.1093/jrsssa/qnad085

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS
Journal of the Statistics Society
Series A
Statistics in Society

Original Article

Unveiling Venice's hotels competition networks from dynamic pricing digital market

Mirko Armillotta[1,2], Konstantinos Fokianos[3] and Andrea Guizzardi[4,5]

[1]Department of Econometrics and Data Science, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
[2]Tinbergen Institute, Amsterdam, The Netherlands
[3]Department of Mathematics & Statistics, University of Cyprus, Nicosia, Cyprus
[4]Department of Statistical Sciences, University of Bologna, Bologna, Italy
[5]Center for Advanced Studies on Tourism (CAST), University of Bologna, Bologna, Italy

Address for correspondence: Konstantinos Fokianos, Department of Mathematicsics and Statistics, University of Cyprus, P.O. Box 20537, 1678 Nicosia, Cyprus. Email: fokianos.konstantinos@ucy.ac.cy

Abstract

We study the dynamic price competition of hotels in Venice using publicly available data scraped from an online travel agency. This study poses two main challenges. First, the time series of prices recorded for each hotel encompasses a twofold time frame. For every single asking price for an overnight stay on a specific day, there is a corresponding time series of asking prices along the booking window on the online platforms. Second, the competition relations between different hoteliers is clearly unknown and needs to be discovered using a suitable methodology. We address these problems by proposing a novel Network Autoregressive model which is able to handle the peculiar threefold data structure of the data set with time-varying coefficients over the booking window. This approach allows us to uncover the competition network of the market players by employing a quick data-driven algorithm. Independent, mixed, and leader–follower relationships are detected, revealing the competitive dynamics of the destination, useful to managers and stakeholders.

Keywords: correlation, data-driven approach, dynamic pricing, leader–follower relationships, multivariate time series, network autoregression

1 Introduction

The spread of E-commerce has facilitated the creation, exchange, and processing of large amounts of information, which has led to changing business strategies for product customisation and marketing in many sectors, including the travel and tourism market (Provenzano & Baggio, 2020). Pricing strategies have also evolved from pure room inventory controls to multi-pricing approaches, allowing revenue managers to differentiate the effect on price due to demand shifts across distribution channels, day of stay and booking horizons (Bigné et al., 2021). The extensive use of online booking platforms such as Booking.com, HRS, or Expedia (i.e. online travel agencies—OTAs) by hotels and their guests has encouraged price transparency and competition (Bigné & Decrop, 2019), increasing the information available about how competing hotels react to competitor price adjustments.

This research contributes to the existing literature by providing a new tool to discover complex relationships using the statistical methods of network data analysis. In line with Skilton and Bernardes (2015), we conjecture that competitive behaviour can be represented by a network architecture. More precisely a competition network (Gimeno, 2004) whose edges are based on

Received: January 27, 2023. **Revised:** July 12, 2023. **Accepted:** July 12, 2023

competitive interdependency, market engagement assumptions and are drawn by observing public actions and responses of different competing agents (Choi et al., 2022). For this reason, we suggest that the whole set of prices a hotel publishes on an OTA can be considered an effective example of (implicit) shared knowledge regarding business price competition tactics in a real-life context (Guizzardi et al., 2021). This is important information because it is composed of primary data (collected in a raw format), posted directly by the competitors and, therefore, observed without the bias that can occurs when using secondary (big) data (e.g. statistics from Google Trends—see Lazer et al., 2014).

The proper analysis of such data poses new statistical challenges because every single asking price for an overnight stay on a certain day corresponds to a time series of asking prices along a booking window. We attack this problem by proposing to relate the inter-temporal pricing structures encoded by these 'time series of time series' with a new autoregressive network approach. This allows us to discover the leader–follower relation among decision makers without any a priori constraints on the network structure, except the very general ones pertaining to the quality segment the hotel competes in—proxied by its median rate—and the size of the geographical area of competition.

Our approach examines the day-by-day online prices, without making a priori assumptions on the factors inducing heterogeneity: pricing strategies or responses to reaction choices by competitors. However, we do evaluate ex-post the role played by some non-price factors, namely, the total capacity at the decision unit, their star rating, and the availability of special features such as restaurants and meeting rooms. These aspects influence competitive behaviours and determine their ability to compete in different market segments (e.g. business or groups markets). Moreover, by considering time-lagged interactions, we avoid the problem of endogeneity bias (Li et al., 2018), resulting from the joint determination of prices among agents, for example, a hotel's price could be simultaneously a function of its competitors' prices and vice versa.

We develop this methodology using data from 95 hotels in Venice, a world renowned cultural–historical destination in Italy, focussing on daily pricing strategies and time-lagged interactions, for up to 14 days of advance booking. This results in 14 autoregressive networks—one for each advance booking—synthesising leader–follower competition patterns over a time span of almost one year (344 days in total from 1 April 2019 to 9 March 2020).

To the best of our knowledge, this is the first attempt to design a self-defining competition network of accommodation enterprises leveraging public data characterised by high-frequency sampling, while explicitly accounting for the role of advance booking. Specifically, we aim to understand decision-making practices in the competitive environment by discovering and measuring the effect of competition on pricing, one of the key management tasks.

The rest of the paper is organised as follows: Section 2 focuses on the background of the present study. In Section 3, a preliminary data analysis is reported highlighting the properties of the data. The methodology for model estimation and competition network discovery is amply discussed in Section 4. Section 5 provides the main results emerging from the network autoregressive models estimation. The interpretation of the hotelier profiles and relations emerging from competition networks are reported in Section 6. A brief summary of the main contributions of the paper is included in Section 7. Finally, Appendix A reports the asymptotic theory for the employed estimators and contains additional tables and figures.

2 Study background

Competition occurs when firms share common resources and product markets (Choi et al., 2022; Yao et al., 2008). Gimeno (2004) suggests that rivalry is also determined when a network of interfirm relationships exists. Thus, in line with Choi et al. (2022) and Lavie (2021), we define a competition network as the overlapping relationships formed in common markets through the actions and responses of embedded rival firms.

The set of competitive relations of a competition network are commonly found in patents, annual reports, traditional and social media such as newspapers, radio, Facebook, or Twitter (Choi et al., 2022) and, specifically for a service where consumption and purchase do not coincide in time, evidence of competitive relations are also inside the online pricing strategies during a purchase time window (Guizzardi et al., 2019). This is implicit public information which is rarely codified because firms need advanced technical tools to interpret their competitors' behaviours and develop policies to obtain business opportunities or reduce threats.

Focussing on the accommodation sector, we know that competitors do not react uniformly in response to a fall in price (Kim et al., 2018) as companies with the strongest competitive position are able to price the service without taking into account the rates from other leaders in the market (Pellinen, 2003). Additionally, Bigné and Decrop (2019) highlight that the high level of dispersion in company pricing strategies also depends on the coexistence of companies who focus more on revenue expansion (providing a high-quality service for customers) and others whose goal is cost reduction. In both cases, companies are aware that customers have a high propensity to search the OTA for the best perceived value (which is not always the lowest price Chung et al., 2021). This makes the challenge of profit maximisation more complex. Moreover, the presence of both stochastic demand and a rich data set of competitors' prices and number of bookings, can be overwhelming to managers gathering information (from the Internet).

The sheer quantity of information (Olsen & Roper, 1998) can lead them to a limited and local centric search for alternative actions (Cyert & March, 1963), where the final pricing decisions become a routine practice based on partial information (Lee, 2016). In these cases, imitation is a characteristic response to uncertainty in decision-making (Gavilan et al., 2018) especially when there is a strong perception of a performance gap with respect to competitors (March & Simon, 1993). This gap is an important trigger for experiential learning (Rezvani et al., 2019); performance levels near expectations foster learning from one's own experience, while performance levels far from expectations favour exploring and learning from others (Baum & Dahlin, 2007). Accordingly, the process of analysing and selecting alternative strategies can make a larger impact than foreseen (Mohr, 1978) given the lack of strategic development skills, organisational goals may be unclear or changing (Cohen et al., 1972) and decision-making eventually depends on the context (Pellinen, 2003).

As a consequence, see also Smallman and Moore (2010), we believe that decision-making should be investigated by specifically focussing on context, accepting complex and unclear causality, and accounting for the importance of both decision-making heuristics and time. Both the time dimensions (day of consumption and advance purchase) affecting competitive dynamics on all the markets where consumption is not on day of purchase and the capacity are fixed (e.g. tickets for events and exhibitions, overnight stays in accommodation facilities, seats in means of transportation).

To fully exploit the opportunity offered by these publicly available threefold data (time and individuals), and to identify competition between decision makers only on the basis of their behaviour, we suggest a new approach. A network time-series model that does not require specifying a priori connections between the nodes (i.e. a self-defining network) precisely because identifying the structure of the competition network is the final goal of this contribution.

In recent years, a growing stream of literature on modelling network connections over multivariate time series has been developed. Zhu et al. (2017) proposed a network autoregressive model (NAR) where a continuous response variable is observed for each node of a network. The high-dimensional vector of such responses is modelled through a dynamic time series regressed on the past values of the response, measured on the node itself and the average lagged response of the neighbours connected to the node. A significant extension of the work to quantile regression has been developed by Zhu et al. (2019). Some other extensions for network time-series models, include the grouped least squares estimation of the NAR model (Zhu & Pan, 2020), and a network version for GARCH-type models (Zhou et al., 2020). Knight et al. (2020) consider a model with more elaborate neighbourhood structures, called generalised NAR, which addresses the effect of several layers of connections between the nodes of the network. In addition, an R software for fitting such models is provided, but only for continuous-valued variables. The latest applications of generalised NAR models can be found in Nason and Wei (2022).

More recently, Armillotta and Fokianos (2022a) extended this line of research to accommodate multivariate count data by specifying linear and a log-linear Poisson network autoregressive models. Further details on count time series can be found in Armillotta, Luati, et al. (2022). The same authors have proposed non-linear extensions of both count and continuous NAR models and a procedure for testing the linearity of the model (Armillotta & Fokianos, 2022b; Armillotta, Fokianos, et al., 2022). R software to perform this kind of analysis was developed by Tsagris et al. (2022) and Armillotta, Tsagris, et al. (2022). Other multivariate models for network data have recently been applied with success in several fields; see Lebacher et al. (2021) for an application on global weapon markets and Gilardi et al. (2022) for an example with car crashes data.

All previous literature develops inference by assuming that connections between the nodes of the network are known a priori, while our contribution does not rely on this kind of assumption. We introduce a new approach that brings to light the unknown network competition patterns between hotels, using a purely data-driven algorithm exploiting all the available information about hotels' price dynamics. The algorithm includes a novel specification of the network autoregressive framework, which takes the three-dimensional nature of the data set into account where the dynamic part of the model is based on advance booking effects connected to prices for each arrival time, instead of the classic lagged time effects. The parameters of the models and the network connections can vary with the advance booking to account for different competition patterns along the booking window, for example, the early or last-minute booking periods. This constitutes an additional innovation in our approach with respect to the existing literature.

We also suggest that the predictive accuracy along advance bookings could be used as a measure to identify the optimal competition network by simply observing daily organisational contexts. In this way, following Roy and Raju (2011), among others, the networks are built on the direct edges among the hotels that are then associated to three typical competition profiles: (i) independent (Bertrand-Nash) behaviour, where firms adjust prices only to maximise their revenue, (ii) Stackelberg leader–follower, where a firm acts as the leader or the follower of another one, and (iii) mixed, where a firm is alternatively leader or follower of a second firm along the advance booking window (i.e. its reaction function varies with the time-lag between purchase and consumption).

Ideally, the competition network should be inter-regional, encompassing all competitors operating at a destination including those in nearby destinations (Pellinen, 2003). However, Abrate and Viglia (2016) suggest that price competition is driven by sharing context-related attributes (primarily location), in addition to tangible and reputational ones. Accordingly, competition is also studied at a single destination level, assuming limits to the geographical distance between two competitors (e.g. almost 300 m in the case of the city of Seville, Spain; Chica-Olmo, 2020). Furthermore, Mohammed et al. (2019) conceptualises the frequency of room rate change to be determined by market structure factors (average occupancy ratio and competitor spatial concentration), hotel characteristics (chain affiliation, star rating, size, and class) and location attributes (district, distance to airport and train station). Thus, in accordance with Urtasun and Gutiérrez (2006), among others, we assume the average price (i.e. market commonality and resource similarity) as a variable to select 'true' competitors in a given geographical neighbourhood. However, we also consider the role played by some important factors that introduce heterogeneity in either pricing strategies, or responses to competitors: capacity (Park et al., 2022), star rating (Sánchez-Pérez et al., 2020) and the availability of special features such as restaurants and meeting rooms.

The approach we propose can be employed to study competition behaviour in every market where consumption is delayed with respect to the purchase time. Given that online shopping has significantly increased in recent years (with a consequent increasing availability of data), we believe we are addressing a major challenge for both academics and managers, as we tackle an important question around the complex competitive relationships among highly heterogeneous service providers.

3 Explanatory data analysis

Our empirical analysis focus on hotels in Venice, a world renowned destination in Italy. As the goal of this paper is to show how it is possible to evince hotel price competition from publicly available data, we focus on a pre-COVID-19 time span. In fact, during COVID-19 many hotels perished from the Internet or followed a virtual channel closure strategy, i.e. offering very high rates only to maintain online visibility while managing to stay closed on certain arrival dates to save on the costs of personnel, heating, and electricity. This point has been further discussed by Arabadzhyan et al. (2021) for the case of Milan. For the specific area of Venice, some evidence has been provided by the Center of Advanced Studies in Tourism (CAST) at University of Bologna; see https://site.unibo.it/indiceattivitalberghiera/en/hai-index for further information.

3.1 Data description

On 10 March 2020, the Italian Ministry issued a decree limiting the movement of all individuals throughout Italy, except when specifically authorised for work or healthcare. This date marks the beginning of the pandemic in Italy. Therefore, 9 March 2020 is the end date of the sample under study. With the aid of a Python-based web-scraping software—see https://pypi.org/project/selenium/—we simulate a customer searching for a room through Booking.com at each of 14 (0–13) different advance booking periods. Prices were scraped between mid-night and 05:00 am, to restrain price changes during the scraping period. All posted offers were recorded but the best available rate (BAR), i.e. the lowest price offered for a standard room was kept, when the rooms appear equivalent based on their characteristics. The most frequent choice for a room during the advance booking period is a non-refundable, double room for single use with breakfast included in the price. We choose this kind of room as our standard for the search. This implies and ensures homogeneity with respect to possible (unobservable) product differentiation practices. Data collection began on 18 March 2019.

The observed data ranges from 1 April 2019 to 9 March 2020 (344 days). We lose 13 days of data at the beginning of the time period because we consider a booking window up to 13 days (e.g. the rate for a stay on 1 April, booked 13 days in advance, can only be found scraping the data on 18 March). 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 (Guizzardi et al., 2019).

Following this criteria, we scraped 95 high- and mid-segment hotels active in online market in Venice, as they have a higher likelihood for dynamic pricing and electronic distribution practices (Dabas & Manaktola, 2007). For these hotels, some covariates describing their characteristics are publicly available: two dummies for the presence/absence of restaurant and meeting room; the total number of rooms available for each hotel; the star rating (from 1 to 5) and the geographical latitude–longitude coordinates. Other covariates (spa, swimming pool, gym, etc.) might be relevant for price competition but we will not consider them for two reasons. First, we apply a parsimony principle leveraging the well-known relation between star rating, number of rooms (as a proxy of hotel size), and customer preference ranking (Dolnicar, 2002). Second, variables for services represent a potential source of bias as they do not guarantee these facilities will actually be used (Guizzardi et al., 2016). For example, knowledge that a hotel has a swimming pool does not provide any information about its size or schedule of operation.

3.2 Descriptive statistics

The set of hotels is indexed by $i = 1, \ldots, N$, $t = 1, \ldots, T$ refers to the arrival date and $k = 0, \ldots, K = 13$ to the number of days of advance booking. Let $Y_{i,t}^{(k)} = \log(P_{i,t}^{(k)})$ denote the log of BAR, for hotel i, arrival date t and advance booking days k.

Table 1 shows some descriptive statistics for $P_{i,t}^{(k)}$, computed over all N hotels and T arrival dates, for each specific advance booking period. We note that average prices are proportional to the advance booking and their minimum is attained at $k = 0$. Venice is a very popular destination and the risk of not finding a room, when booking during last minute, costs more than 30 euros between the median asking price at $k = 13$ and the average rate at $k = 0$. The large gap between the extreme percentiles (about 417 euros on average) is partially explained by fluctuations which are typical of daily time-series rates in a destination where world-famous events are held, such as Carnival or the Biennale. Furthermore, even though the sample consists of hotels which are relatively homogeneous in terms of quality (at least 3 stars), some offer special features such as good views of Venice's main attractions. For this reason, location can lead to differentiated average room rates even when comparing facilities located a few metres apart. Mathur and Dewani (2016). For a room with a view of the Grand Canal in a 5-star hotel, the cost rarely falls below 4,000 euros per night with peaks around 7,500 during the high season.

The first column in Table 2 presents some descriptive statistics for the BAR computed over all hotels, arrival dates, and advance bookings. In the second column the same results are presented

Table 1. Descriptive statistics of best available rates for each advance booking

k	5%-quantile	Median	Mean	95%-quantile
0	44.170	113.550	166.689	457.500
1	45.320	125.515	173.538	460.565
2	46.400	128.600	175.844	457.500
3	46.400	129.500	178.704	456.120
4	48.494	129.620	175.959	450.500
5	48.500	134.540	181.094	453.300
6	48.500	135.980	181.012	451.635
7	49.400	139.500	199.609	510.854
8	49.400	139.500	189.917	501.635
9	48.652	138.500	178.665	453.500
10	48.500	139.500	177.962	450.500
11	48.500	142.730	180.014	452.721
12	48.500	143.640	183.676	453.603
13	48.500	146.500	185.306	454.443

considering only the last-minute booking, i.e. $0 \leq k \leq 3$. A summary of the other covariates is also reported. From Table 2, we can see that the median BAR is around 135.2 euros, with high variability of the different hotels available in the large market of Venice. About 28% (18%) of the hotels analysed has a restaurant (meeting) room. All the hotels included in the analysis have at least 10 rooms, with an average number of rooms around 45. Typically, each hotel has 5–6 neighbourhood hotels within a 200 m radius.

A multivariate time-series plot for all hotels at advance booking $k = 0$, see Figure 1, shows the seasonal patterns observed in Venice. Two distinct profiles become clear during the time period we consider. The green part of the plot shows the high room prices that correspond to the seasonal periods of Spring and Summer. From the median log-price time series (Figure 1, panel below), we note that major peaks are detected during Easter and August holidays. The purple part of the plot corresponds to the prices during the Winter period. This season constitutes a second profile which is characterised by lower room rates. We also observe peaks around end-of-the-year holidays and during Venice carnival. Figure A1 in Appendix A provides a better illustration of the same data by employing a cubic spline smoother. The conclusions are identical. The seasonal pattern is homogeneous across all advance bookings. (Figures with $k \neq 0$ are not shown due to space constraints, but are available upon request.)

In Figure 2, the left panel shows the result of a hierarchical cluster analysis on hotels, using as clustering variables 14 standard deviations, $\sigma_i^{(k)} = \sqrt{T^{-1} \sum_{t=1}^{T} (Y_{i,t}^{(k)} - \bar{Y}_i^{(k)})^2}$, for $k = 0, \ldots, 13$, where $\bar{Y}_i^{(k)}$ is the sample mean for the log-BARs of hotel i at advance booking k. In Figure 2 (right panel), we cluster the advance bookings $k = 0, \ldots, 13$ by using the same standard deviations, $\sigma_i^{(k)}$ as clustering variables. Finally, in Figure 3, we report the magnitude of the standard deviations. On the whole, it appears that not all hotels have the same propensity to apply dynamic pricing. From Figure 2 (left panel), the height of the dendrogram reveals the presence of two groups. Cluster 1, includes 60% of hotels with the lowest price variability (see Figure 3), while Cluster 2 groups the hotels with the highest likelihood of changing price according to seasonality.

We also note that there is a strong advance booking effect on the propensity to rely on dynamic pricing. In fact, we find a clear separation between last-minute booking ($k = 0, \ldots, 3$) and early reservations. We observe a slightly lower variability, indicating that the pricing strategies become more homogeneous during last-minute reservations—see Figure 3. This is a lower-effort/less-risky pricing strategy.

Table 2. Descriptive statistics

	BAR	BAR ($k \leq 3$)	Restaurant	Meeting	Num. rooms	Num. hotels in 200 m
1st Qu.	81.50	75.48	0.00	0.00	22.00	1.00
Median	135.20	124.50	0.00	0.00	28.00	4.00
Mean	181.70	173.57	0.28	0.18	44.98	5.28
3rd Qu.	213.50	201.90	1.00	0.00	53.00	9.00

Note. BAR = best available rate.

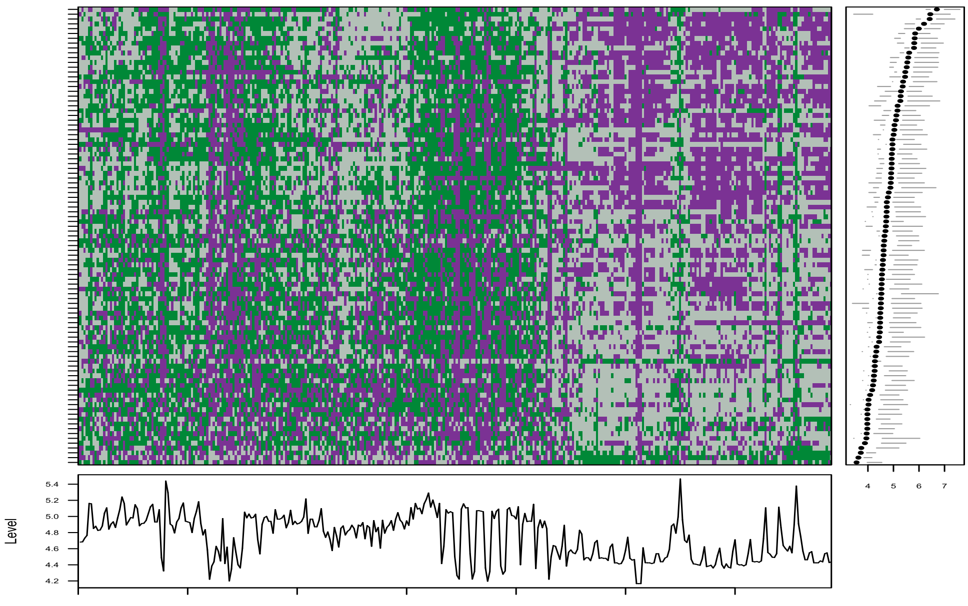

Figure 1. Multivariate time series of log-transformed best available rates. Advance booking $k = 0$. Hotels by row ordered in descending order by median log-price. Time in column (1 April 2019–9 March 2020). Colours reflect threefolds of log-price distribution: high (green), mid (grey), low (purple) obtained by dividing each hotel's log-price distribution in tertiles. Right panel: box plot distribution around the median (full dot) for each hotel time series. Below: time series of hotel median log-price levels.

Table 3 summarises some descriptive statistics. The two clusters of hotels identified are very different in terms of size, price and meeting room availability. Cluster 1 includes large and expensive accommodations offering business-oriented services. This is expected as these high-quality hotels tend to maintain a more stable pricing pattern to address price-fairness concerns. In addition, they are positioned in the off-line market where the business customers often book multiple rooms, well in advance, negotiating special rates. Such established rates become a benchmark which is difficult to change because businesses cannot price rooms online and at the same time satisfy their off-line customers.

4 Methodology

The room prices published online provide researchers an opportunity to dig deeply into the hotel's short-term price competition tactics in a real-life context. In line with Skilton and Bernardes (2015), we propose to model the heterogeneity with which competing hotels react to competitors' price adjustments by employing a novel Network Autoregressive approach (see Section 4.1) able to handle data with one cross-sectional dimension (the hotels i) and two time-series dimensions

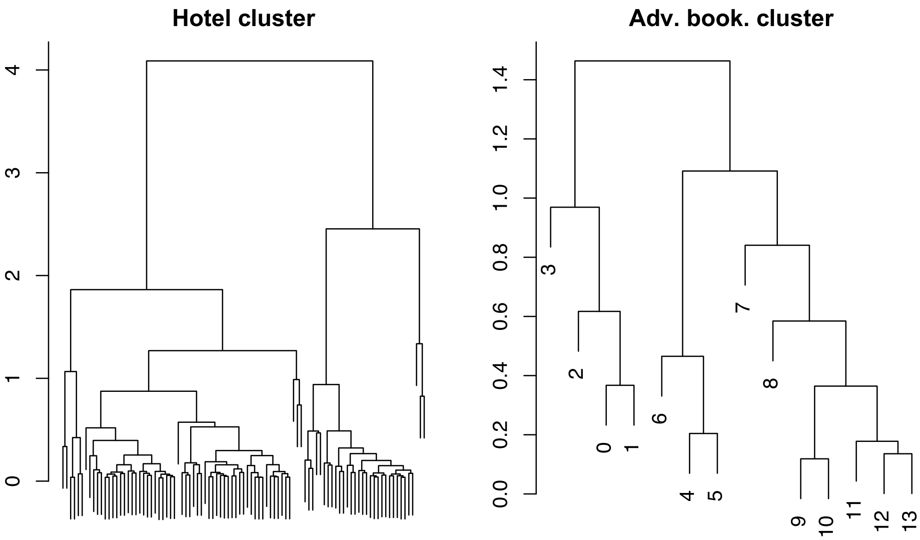

Figure 2. Hierarchical clustering with complete linkage of sample standard deviations of hotels (left panel) and advance bookings (right panel). Group dissimilarity levels on the vertical axis. For the segmentation along hotels (left panel) the clustering variables are 14 standard deviations computed over arrival dates *t*, one for each advance booking *k*. For clustering of advance booking (right panel), clustering variables are 95 standard deviations computed over arrival dates *t*, one for each hotel *i*.

(arrival dates t for the room rented and number of days of advance booking k). The structure of a network with N nodes (network size) with index $i = 1, \ldots N$ is described by an adjacency matrix $A = (a_{ij}) \in \{0, 1\}^{N \times N}$, i.e. $a_{ij} = 1$ if there exists a directed edge from i to j, $i \to j$ (hotel i follows hotel j), and $a_{ij} = 0$ otherwise. Since any hotel cannot be competing against itself, self-relationships are not allowed, i.e. $a_{ii} = 0$ for any $i = 1, \ldots, N$; this is a typical assumption (Kolaczyk & Csárdi, 2014; Wasserman et al., 1994). For each hotel (node i) the time series of BARs is measured over a time window $t = 1, \ldots, T$.

Moreover, to account for different competition behaviours along the booking window, the network matrix is assumed to depend on k, i.e. $A^{(k)} = (a_{ij}^{(k)}) \in \{0, 1\}^{N \times N}$, i.e. $a_{ij}^{(k)} = 1$ if hotel i follows the discounts/surcharges of hotel j for rooms sold k days in advance ($i \to j$), and $a_{ij}^{(k)} = 0$ otherwise. For example, if $k = 0$, the network will uncover the competition relations in the last-minute sector, while for $k = 13$ we account for a different early-booking competition network.

A focal point of our analysis is that we do not have a priori knowledge of the competition relations between hotels. Therefore, we suggest this approach to discover the price competition relations, starting from the competitors' pricing actions and reactions observed along time t and advance booking k. To this aim, we introduce the following methodology.

4.1 Advance booking network autoregression

In this section, we will assume that the sequence of network adjacency matrices $\{A^{(k)}, k = 0, \ldots, K-1\}$ is known, leaving the task of reconstructing such sequence of matrices to the next section, using a data-driven process able to give a structural form to competition patterns between hotels. We propose the following network autoregressive models for advance bookings $k = 0, \ldots, K-1$.

$$Y_{i,t}^{(k)} = \beta_0^{(k)} + Z_i' \gamma^{(k)} + \delta_t' \alpha^{(k)} + \beta_1^{(k)} \frac{1}{n_i^{(k)}} \sum_{j=1}^{N} a_{ij}^{(k)} Y_{j,t}^{(k+1)} + \beta_2^{(k)} Y_{i,t}^{(k+1)} + \varepsilon_{i,t}^{(k)}, \tag{1}$$

where $Y_{i,t}^{(k)} = \log(P_{i,t}^{(k)})$, $\varepsilon_{i,t}^{(k)}$ is the stochastic error, assumed to be independent, with standard deviation $\sigma^{(k)}$, $n_i^{(k)} = \sum_{j=1}^{N} a_{ij}^{(k)}$ is the total number of connections starting from hotel i, at advance

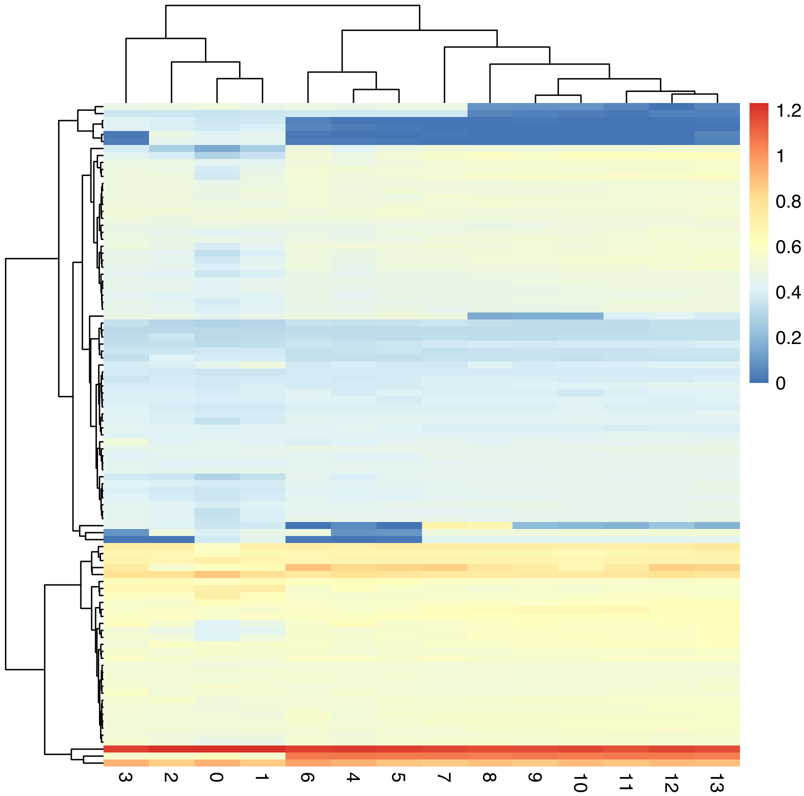

Figure 3. Heat map of the hierarchical clustering presented in Figure 2 whose dendrograms are reported over the axes (left panel → row-wise, right panel → column-wise). The magnitude of sample standard deviations over arrival time *t* for hotels (rows) and advance bookings (columns) is shown.

booking k, called out-degree.

Model (1) postulates that, for every single hotel i being a potential node of a network, the log-BAR time series of hotel i, at arrival time t, booked k days in advance, $Y_{i,t}^{(k)}$, is regressed on:

- The average log-BARs, same arrival date t, at advance booking $(k+1)$, computed over hotels $j \neq i$ which are followed by i (i.e. $i \to j$). That way we let price policies of hotels j have an impact on the prices of hotel i; such effect is then called 'network effect', associated to the parameter $\beta_1^{(k)}$.
- The log-BAR for the same hotel i, same arrival day t, but higher advance booking, $Y_{i,t}^{(k+1)}$. The advance booking autoregressive effect is measured by the parameter $\beta_2^{(k)}$, that will be henceforth simply called 'autoregressive effect'.
- A set of structural hotel-specific covariates $Z_i = (Z_{1,i}, Z_{2,i}, Z_{3,i}, Z_{4,i}, Z_{5,i})'$. More precisely, $Z_{1,i}$ ($Z_{2,i}$) is a dummy variable taking values 1 if hotel i owns a restaurant (meeting room, respectively), 0 otherwise. $Z_{3,i}$ is the normalised number of hotel i's rooms; this represent hotel i's size. $Z_{4,i}$ is the (normalised) number of hotels within 200 m from hotel i; this is assumed to measure the competition pressure that hotel i undergoes. Finally, $Z_{5,i}$ is the star rating.
- A set of seasonal dummies $\delta_t = (\delta_{1,t}, \ldots, \delta_{8,t})'$, where $\delta_{1,t}$ is the weekend dummy, taking value 1 if the arrival date t is at the end of the week (Friday/Saturday/Sunday), 0 otherwise. The others indicators correspond to a specified holiday period. We considered: Carnival, Easter, Pentecost, August, Mid-August, Halloween, Immaculate Conception, Christmas, and New Year.

Table 3. Descriptive statistics for the two hotel clusters identified in Figure 2

Cluster		BAR	BAR ($k \leq 3$)	Restaurant	Meeting	Num. rooms	Num. hotels in 200 m
1	1st Qu.	87.00	83.69	0.00	0.00	22.50	1.00
	Median	142.20	132.50	0.00	0.00	31.00	4.00
	Mean	180.30	179.13	0.29	0.22	49.75	5.00
	3rd Qu.	217.50	209.25	1.00	0.00	63.50	8.50
2	1st Qu.	66.92	58.50	0.00	0.00	20.75	1.75
	Median	124.82	108.03	0.00	0.00	26.00	4.00
	Mean	184.41	163.00	0.28	0.09	35.59	5.84
	3rd Qu.	204.74	184.50	1.00	0.00	39.50	9.25

Note. BAR = best available rate.

By rewriting model (1) in matrix form we have, for $k = 0, \ldots, K-1$

$$Y_t^{(k)} = \beta_0^{(k)} 1 + Z\gamma^{(k)} + \Delta_t \alpha^{(k)} + G^{(k)} Y_t^{(k+1)} + \varepsilon_t^{(k)}, \tag{2}$$

where $G^{(k)} = \beta_1^{(k)} W^{(k)} + \beta_2^{(k)} I$, $W^{(k)} = \operatorname{diag}\{1/n_1^{(k)}, \ldots, 1/n_N^{(k)}\} A^{(k)}$ is the sequence of row-normalised adjacency matrix, with $A^{(k)} = (a_{ij})$, so $w_i^{(k)} = (a_{ij}^{(k)}/n_i^{(k)}, j = 1, \ldots, N)' \in \mathbb{R}^N$ is the ith row vector of the matrix $W^{(k)}$, and I is an identity matrix of appropriate dimensions. Note that all the row-sums of the matrix $W^{(k)}$ are no bigger than 1, for all k. Then, $Z = (Z_1, \ldots, Z_5)$ is the $N \times 5$ matrix of covariates, with N-dimensional vectors $Z_h = (Z_{h,i})$, for $i = 1, \ldots, N$ and $h = 1, \ldots, 5$, with $\gamma^{(k)}$ being the associated 4×1 vector of parameters. Finally, $\Delta_t = (\delta_t, \ldots, \delta_t)'$ is an $N \times 8$ matrix where each row replicates the dummy vector δ_t and whose relative parameter vector is $\alpha^{(k)}$.

Model (2) assumes that the structure of the network is non-random. Once the network matrix W is known, it is then possible to estimate the $m \times 1$ vector of unknown parameters of the model $\theta^{(k)} = (\beta_0^{(k)}, \beta_1^{(k)}, \beta_2^{(k)}, \gamma^{(k)\prime}, \alpha^{(k)\prime})'$, for $k = 1, \ldots, K-1$, by using the following least squares methods.

Define the $N \times m$ matrix of regressors $Q_t^{(k+1)} = (1, W^{(k)} Y_t^{(k+1)}, Y_t^{(k+1)}, Z, \Delta_t)$, so that model (2) can be rewritten as $Y_t^{(k)} = Q_t^{(k+1)} \theta^{(k)} + \varepsilon_t^{(k)}$, for $t = 1, \ldots, T$. We then rewrite the problem as a general TN-dimensional linear model in the following way. Define $Y^{(k)} = (Y_1^{(k)\prime}, Y_2^{(k)\prime}, \ldots, Y_T^{(k)\prime})'$, $\varepsilon^{(k)} = (\varepsilon_1^{(k)\prime}, \varepsilon_2^{(k)\prime}, \ldots, \varepsilon_T^{(k)\prime})'$ and $Q^{(k+1)} = (Q_1^{(k+1)\prime}, Q_2^{(k+1)\prime}, \ldots, Q_T^{(k+1)\prime})'$ is the $TN \times m$ matrix of stacked regression matrices. The solution of the general linear system $Y^{(k)} = Q^{(k+1)} \theta^{(k)} + \varepsilon^{(k)}$ is the following least squares estimator, for $k = 1, \ldots, K-1$

$$\hat{\theta}^{(k)} = \left(Q^{(k+1)\prime} Q^{(k+1)} \right)^{-1} Q^{(k+1)\prime} Y^{(k)} = \left(\sum_{t=1}^{T} Q_t^{(k+1)\prime} Q_t^{(k+1)} \right)^{-1} \sum_{t=1}^{T} Q_t^{(k+1)\prime} Y_t^{(k)}. \tag{3}$$

Large sample properties of equation (3) are established within the typical Gauss-Markov framework (see Appendix A.2) which presupposes iid errors and iid exogenous covariates.

Note that model (2) has m parameters to be estimated. In contrast, if we had assumed the full-matrix model $G^{(k)} = (g_{ij}^{(k)})_{i,j=1,\ldots,N}$ then we would face the problem of estimating $\mathcal{O}(N^2)$ parameters. As stated in Bernanke et al. (2005), full-matrix models are not often used for economic data sets that contain more than 6–8 time series. Therefore, the parsimony of equation (2) allows us to treat high-dimensional data.

Price dynamics modelled by equation (1) introduce a novel methodology which has not been discussed in the literature, to the best of our knowledge. Related works (see Section 2) employ rather different approaches when compared to equation (1). Indeed, the dynamic part of the model is not purely 'autoregressive', i.e. based on time observation $(t-1)$ but is indeed a forward effect in the advance booking $(k+1)$, connected to each time event t. Moreover, the model allows us to deal with three-dimensional data sets, for individual (i) over a time frame (t) with respect to a certain action occurring k time units previous. Finally, all the coefficients of the models, including network, are allowed to vary with the advance booking.

4.2 Building competition network

The model defined in equation (1) is a flexible tool allowing for the examination of several network effects, according to the way the matrix W is selected. For, example, it can be chosen by computing the geographical distance between nodes. In this case, W represents the matrices of spatial effects; see Cliff and Ord (1975) and Martin and Oeppen (1975), among many others. However, several alternatives are possible, depending on available information about the nodes.

In the present work, we are particularly interested in the definition of a self-selecting competition network composed by directed edges. In this way, competitive reactions (if any) are identified by the model which takes into account published prices at different arrival dates and lagged advance bookings.

Thus, as a first step, we compute, for each hotel i, the median BAR, say $P_i = \text{median}_{k,t}(P_{i,t}^{(k)})$, by considering all the arrival dates $t = 1, \ldots, T$ and all the advance bookings $k = 0, \ldots, 13$. P_i is taken as an overall measure of hotel's competition segment, given its tangible, reputation and contextual characteristics.

As a second step, we calculate the price distance between a pair of hotels i and j, by $d_{ij} = |\log(P_i) - \log(P_j)|$. This quantity can be viewed as a measure of the difference between services offered by the i and j hotels. Then, we determine the probabilities to draw an edge from hotel i to hotel j—the probability to compete—as follows:

$$p_{ij} = p_{ij}^h p_{ij}^d, \quad \text{where } p_{ij}^h = 1 - F_h(h_{ij}, \mu, \lambda), \quad p_{ij}^d = 1 - F_d(d_{ij}, \eta, \nu), \tag{4}$$

with h_{ij} being the geodesic distance between hotels i and j computed with the haversine formula. F_h, F_d are the cumulative distribution functions (c.d.f.) of the geodesic and price distances, respectively, whose parameters $\tau = (\mu, \lambda, \eta, \nu)'$ and distributional forms depend on the characteristics of the destination under study. We appeal to complement of the c.d.f. to obtain that p_{ij}^d is decreasing function of distance, i.e. $d_{ij} \to \infty$ implies $p_{ij}^d \to 0$, and when $d_{ij} \to 0$, then $p_{ij}^d \to 1$. Similar interpretations hold for p_{ij}^h. From equation (4), two hotels are likely to compete if they are geographically close and have a similar price positioning. By contrast, hotels that are far away and/or with significantly different average price levels will not compete.

In order to determine competition probabilities we exploit the information already available about the peculiarities of Venice market. Focussing on spatial competition, we see that the most famous tourist heritage attractions in Venice are in the Sestiere San Marco district (Sestiere is one of six districts in the city of Venice). This area is only about 0.45 km^2 with borders formed by the landmark bridges Ponte dell'Accademia (West), Ponte di Rialto (North), and Ponte della Paglia (East). Accordingly, about 50% of the hotels in this study are located in a radius measuring 500 m from the centre of the San Marco district (see Figure 6). Therefore, we assume that at distance $h_{ij} > 500$ m the probability of competition between hotel i and j is zero, i.e. $p_{ij}^h = 0$. This restriction is a realistic assumption since assuming the probability of competition among hotels beyond 500 m is greater than zero, would mean each hotel competes with a very large number of other hotels. However, in practice, market players only systematically follow a limited number of competitors (Cyert & March, 1963; Li et al., 2018). At the same time, we only want to consider all the hotels within a minimum geodesic distance as potential competitors. Following the literature (e.g. Chica-Olmo, 2020 and considering the high spatial density of hotels in Venice, we consider it a realistic assumption that hotels close to each other have a high probability of being competitors.

The logistic function is a natural choice to map probabilities in a (0, 1) range starting from a real domain. Therefore, for the spatial competition we employ a truncated logistic c.d.f. with left

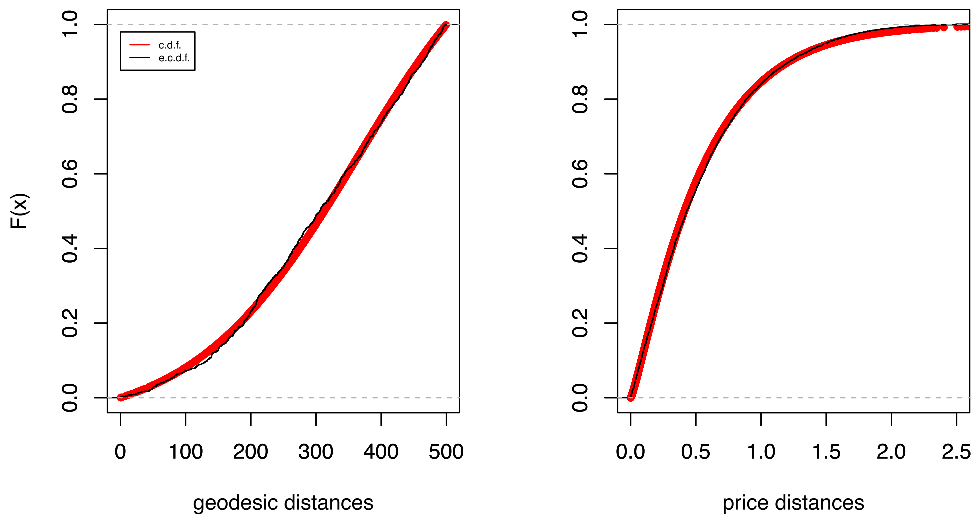

Figure 4. Left: empirical c.d.f. of geodesic distances (black) against c.d.f. of estimated truncated logistic (red). Right: empirical c.d.f. of price distances (black) against estimated gamma c.d.f. (red).

truncation to 0 (distances cannot be negative) and right truncation to 500:

$$F_h(h_{ij}, \mu, \lambda) = \frac{L_h(h_{ij}, \mu, \lambda) - L_h(0, \mu, \lambda)}{L_h(500, \mu, \lambda) - L_h(0, \mu, \lambda)}, \quad L_h(h_{ij}, \mu, \lambda) = \frac{1}{1 + \exp\left(-(h_{ij} - \mu)/\lambda\right)}.$$

Focussing on the price distance, we assume that two hotels say i and j, have a high probability of being competitors if they have similar prices while they should have a very low probability of competition if their prices are reasonably different. In this case, a logistic distribution is not adequate (Figure 4, right panel). As a reasonable choice of distribution for a non-negative variable such as the price distance, we choose $d_{ij} \sim \text{Gamma}(\eta, v)$, where the parameters denote shape and rate, respectively. We then estimate each pair of parameters by maximum likelihood,

$$(\hat{\mu}, \hat{\lambda})' = \underset{\mu,\lambda}{\text{argmax}} \sum_{i=2}^{N} \sum_{j<i} \log f_h(h_{ij}, \mu, \lambda), \quad (\hat{\eta}, \hat{v})' = \underset{\eta,v}{\text{argmax}} \sum_{i=2}^{N} \sum_{j<i} \log f_d(d_{ij}, \eta, v),$$

where f_h, f_d are the probability density functions of the truncated logistic and gamma distributions, respectively. The estimation leads to $\hat{\tau} = (\hat{\mu}, \hat{\lambda}, \hat{\eta}, \hat{v})' = (362.45, 124.41, 1.22, 2.21)'$ being all significantly different from 0 at usual significance levels. The resulting c.d.f. computed with estimated parameters provides an adequate fit of the data (see Figure 4). The Kolmogorov–Smirnov does not reject both models, at 1% level.

Define $Y^{(k)}$ the whole set of observed prices at advance booking k, Δ the whole set of dummies at all the available arrival times t, and $X^{(k)} = (Y^{(k)}, Z, \Delta)$ the whole data set of observations available to the researcher at advance booking k. With this notation, the OLS estimation of the parameters is a function of the network and the data set, i.e. $\hat{\theta}^{(k)} = \hat{\theta}^{(k)}(W^{(k)}, X^{(k)})$. Thus, the fitted values $\hat{Y}_t^{(k)} = \hat{\beta_0}^{(k)}1 + Z\hat{\gamma}^{(k)} + \Delta_t\hat{\alpha}^{(k)} + (\hat{\beta}_1^{(k)} W^{(k)} + \hat{\beta}_2^{(k)} I)Y_t^{(k+1)} = \hat{Y}_t^{(k)}(W^{(k)}, \hat{\theta}^{(k)}, X^{(k)})$ are functions of the data and the competition network structure. Since in competition market the network is unknown, we propose the following data-driven procedure to construct it. For each $k = 0, \ldots, K-1$, the off-diagonal element of a network adjacency matrix $A^{(k)}$ can be generated by a Bernoulli trial with probability p_{ij} defined as in equation (4), for each pair of hotels, i.e. for each element of the adjacency matrix, a_{ij}. Repeat the same simulation S times so that we obtain $A^{(k,s)}$ (and so $W^{(k,s)}$), for $s = 1, \ldots, S$, used to perform the estimation $\hat{\theta}^{(k,s)}$, compute fitted values $\hat{Y}_t^{(k,s)}$, and then select

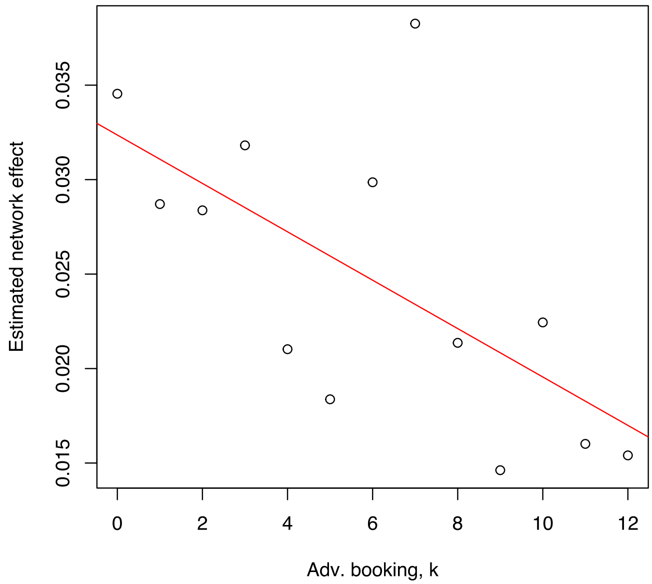

Figure 5. Scatter plot of estimated network effects $\hat{\beta}_1^{(k)}$ of model (1) against the advance bookings $k = 0, \ldots, K-1$. Red line: linear trend (p-value 0.0176).

the network matrix minimising the root mean squared error (RMSE)

$$\hat{W}^{(k)} := \underset{W^{(k,s)}}{\operatorname{argmin}} \sqrt{\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\left(Y_{i,t}^{(k)} - \hat{Y}_{i,t}^{(k,s)}\right)^2}. \tag{5}$$

This type of data-driven sampling approach selects the network matrix minimising the discrepancy between the model and the scraped data. This allows us to adaptively shape the price competition network, that dynamically forms over time, which is more likely to represent the everyday price competition context in a certain neighbourhood area.

In summary, the network estimation process consists of (i) maximum likelihood estimation of competition probabilities, by employing equation (4); for each advance booking k: (ii) simulating networks, several times, using the estimated probability; (iii) fitting of NAR model (1) to each network generated by applying equation (3); (iv) and identification of the competition network which minimises the RMSE (5). Detailed description of the steps involved in unveiling the competition network is given by Algorithm **1**.

Alternative methodology might be applied to identify competition networks. For instance, Guizzardi et al. (2019) employ a vector autoregressive model of order 1, say VAR(1), in connection to Granger causality tests for the same estimation problem. However, fitting of a VAR(1) model requires estimation of $N + N^2$ parameters for each advance booking k. The NAR based approach taken in this work requires estimation of less parameters (m for each k) and allows for covariate inclusion which can be instrumental on discovering the competition network. In addition, Algorithm **1** determines the network that minimises the discrepancy with the real world dynamic pricing.

5 Estimation results

Results of estimation from model (1), with networks generated according to the mechanism described in Section 4.2 are shown in Table 4 by generating $S = 1{,}000$ network matrices. The output is focussed on variables with corresponding significant coefficients. We do not find any deterministic seasonal effect (weekend and holiday dummy variables were not significant) because any stochastic seasonal effect is taken into account by the published price at advance booking $k + 1$. This

Algorithm 1 Algorithm for identifying competitive networks on NAR models with advance bookings dynamic prices.

Require N number of hotels; T temporal size; K total number of advance booking; S number of simulations; Y log-BAR data; Z covariates; Δ dummies; h_{ij} geodesic distances

$P_i \leftarrow \text{median}_{k,t}\left[\exp\left(Y_{i,t}^{(k)}\right)\right]$

$d_{ij} \leftarrow \left|\log(P_i) - \log(P_j)\right|$

Estimate p_{ij} according to (4)

for $k = 0$ to K **do**

  for $s = 1$ to S **do**

    Draw $a_{ij}^{(k,s)} \in \{0, 1\}$ from $Bernoulli(p_{ij})$ for all $i \neq j$

    Compute the network matrix $W^{(k,s)} = \left[w_{ij}^{(k,s)}\right]_{i,j=1,\dots,N}$ where $w_{ij}^{(k,s)} \leftarrow a_{ij}^{(k,s)} / \sum_{j=1}^{N} a_{ij}^{(k,s)}$

    $Q_t^{(k+1,s)} \leftarrow \left[1,\ W^{(k,s)} Y_t^{(k+1)},\ Y_t^{(k+1)},\ Z,\ \Delta_t\right]$

    $Q^{(k+1,s)} \leftarrow \left[Q_1^{(k+1,s)\prime},\ Q_2^{(k+1,s)\prime},\ \dots,\ Q_T^{(k+1,s)\prime}\right]^{\prime}$

    $\hat{\theta}^{(k,s)} \leftarrow \left[Q^{(k+1,s)\prime} Q^{(k+1,s)}\right]^{-1} Q^{(k+1,s)\prime} Y^{(k)}$

    $\hat{Y}_t^{(k,s)} \leftarrow \hat{\beta}_0^{(k,s)} 1 + Z\hat{\gamma}^{(k,s)} + \Delta_t \hat{\alpha}^{(k,s)} + \left[\hat{\beta}_1^{(k,s)} W^{(k,s)} + \hat{\beta}_2^{(k,s)} I\right] Y_t^{(k+1)}$

    $RMSE_{k,s} \leftarrow \sqrt{(NT)^{-1} \sum_{t=1}^{T} \sum_{i=1}^{N} \left[Y_{i,t}^{(k)} - \hat{Y}_{i,t}^{(k,s)}\right]^2}$

  end for

  $RMSE_k \leftarrow \min_{s=1,\dots,S}\left\{RMSE_{k,s}\right\}$

  $s^* \leftarrow \text{argmin}_{s=1,\dots,S}\left(RMSE_{k,s}\right)$

  $\hat{W}^{(k)} \leftarrow W^{(k,s^*)};\ \hat{Y}^{(k)} \leftarrow \hat{Y}^{(k,s^*)};\ \hat{\theta}^{(k)} \leftarrow \hat{\theta}^{(k,s^*)}$

  Compute the network matrix $\hat{A}^{(k)} = \left[\hat{a}_{ij}^{(k)}\right]_{i,j=1,\dots,N}$ where $\hat{a}_{ij}^{(k)} \leftarrow I\left(\hat{w}_{ij}^{(k)} \neq 0\right)$

  return $\left\{\hat{A}^{(k)},\ \hat{W}^{(k)},\ \hat{Y}^{(k)},\ \hat{\theta}^{(k)},\ RMSE_k\right\}$

end for

particular explanatory variable contains information for future demand fluctuations. Basically, it represents the expectation of each hotelier regarding the overall effect of seasonality at the destination due to different arrival days t, as well as its idiosyncratic effect due to single hotel features (Guizzardi et al., 2022). For a similar reason, Z_5 (star rating) is not significant for all the advance booking periods. In fact, the lagged price can be considered a proxy of the general quality level and/or of the offered features.

Therefore, model (1) was re-estimated by excluding the nonsignificant variables and the results are reported in Table 4. The analysis provides clear evidence of a significant network effect. More precisely, the coefficients $\hat{\beta}_1$ are found to be positive and significant at 5% significance level, indicating that, for competing hotels, a change in the rate published by the 'average competitor' at advance booking $k + 1$ is followed by a rate adjustment of the same sign in k. The adjustment ranges from around 1.5% in the early-booking period to more than 3.5% in the last-minute time frame. More interesting, we find that the estimated network effect follows a linear deterministic relation (see Figure 5). A simple regression where the dependent variable is a deterministic trend ($k = 0, \dots, 12$) yields a statistically significant slope. In other words, reactions to competitors' pricing policies increase as the number of advance booking days decreases; a result consistent with Guizzardi et al. (2019), among others.

The autoregressive effect $\hat{\beta}_2$ is also significant and positive. The high value of the autoregressive coefficients (0.818 on average) shows that the prices set by hoteliers are very persistent, i.e. that hoteliers follow their own pricing strategy, which is accurate in forecasting reservations/cancellations along the booking curve and/or attentive to consumers' price-fairness issues affecting profits in the long term (Malc et al., 2016). The coefficient reaches its minimum at $k = 0$ when the risk of both being left with unsold rooms and a performance gap compared to a competitor become triggers to imitating and learning from others.

The covariates Z_l, for $l = \{1, 2, 3\}$, are always significant and the overall tendency of their coefficients is positive. These variables together can be considered a proxy of the value of the

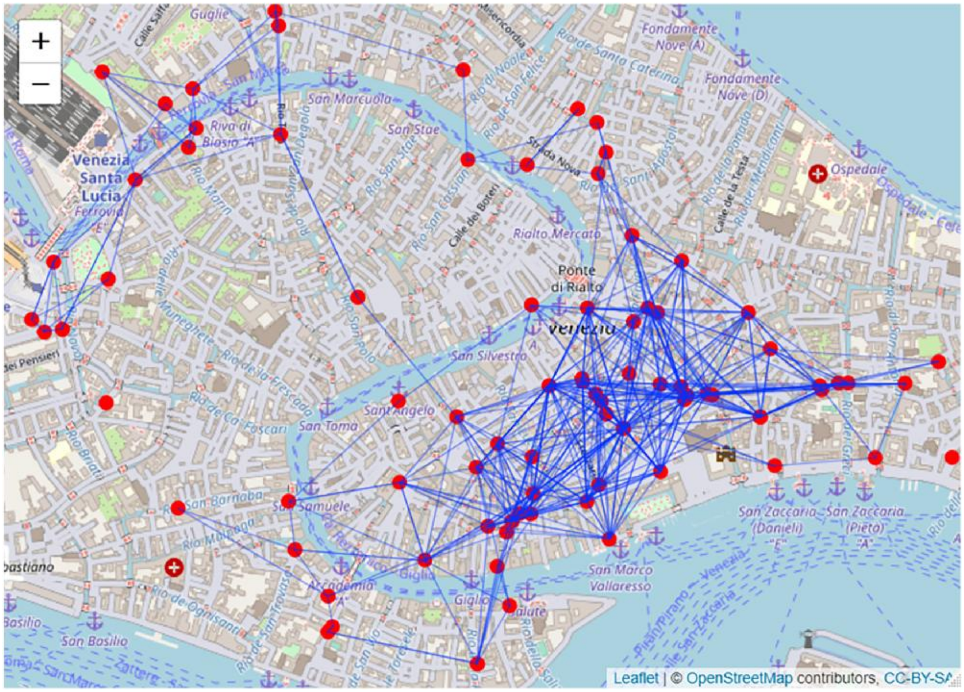

Figure 6. Estimated hotel competition network for $k = 0$ performed according to Algorithm 1. $S = 1,000$ simulations. Red dots: hotels. Blue line: connection between hotels.

property so that an upward effect on final price seems reasonable. The covariate Z_4 is significant for 11 up to 13 advance bookings and is negative. It accounts for the spatial density of hotels in a limited 200 m area, measuring the effect of competition. Indeed, the hospitality literature has shown that a higher level of competition can decrease the price of a service (Becerra et al., 2013).

Finally, we note that the minimised RMSE is low, indicating that shaped competitive relationships help to predict the pricing behaviour and—consequently—they can be considered an accurate representation of the competitors' relationships in 'real life'. See also Section 5.1.

Figures 6 and 7 show the competition network at $k = 0$ and $k = 7$, respectively. The plots illustrate that location is an important factor in differentiating the offer and therefore determine the intensity of the competition. High competition is observed in the San Marco district where all the hotels located in this area have the highest number of estimated average connections (up to 15). For a hotel, being near the Grand Canal significantly improves the possibility of differentiating the offer to reduce the intensity of competition (the number of both competitors and bidirectional edges). Moreover, hotels on opposite sides of the Canal compete less, although they are close in geographical terms, they are perceived as distant due to the difficulty of crossing the water. As expected this evidence is stronger in the early-booking period with more rooms available.

The autocorrelation functions of the residual $\varepsilon_{i,t}^{(k)}$, see Figure A2, indicates that model (1) captures the time dependence along t quite satisfactorily at $k = 0$. Similar results (available on request but not shown for space considerations) are obtained for the other advance bookings k allowing us to conclude that the fluctuations of prices for $t = 1, 2, \ldots, T$ are adequately accounted for by model (1).

5.1 Alternative model specifications

In further support of the validity of the specification (1), we compare its forecasting performance against two rival models that also include the rates posted for arrival days $t - 1$ and $t - 7$ (using the same k). In this way, we assume that competition is also driven by a time effect along the calendar

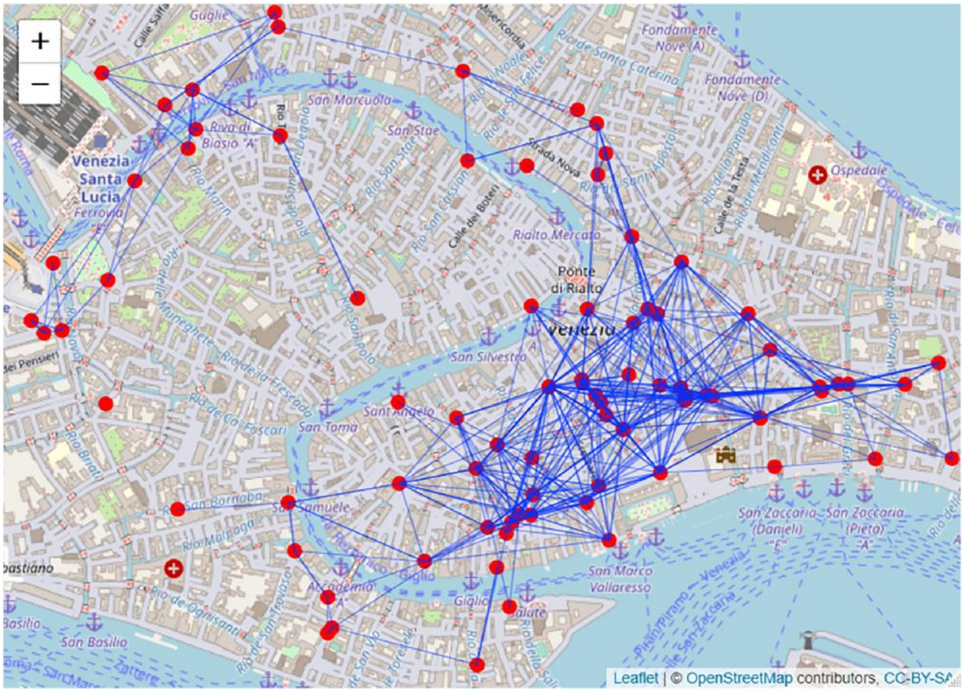

Figure 7. Estimated hotel competition network for $k = 7$ performed according to Algorithm 1. $S = 1{,}000$ simulations. Red dots: hotels. Blue line: connection between hotels.

days t, i.e. according to price adjustment between adjacent arrival days or between the same day of the week in adjacent weeks. In particular:

$$\Delta Y_{i,t}^{(k)} = \beta_0^{(k)} + Z_i'\gamma^{(k)} + \delta_t'\alpha^{(k)} + \beta_1^{(k)} \frac{1}{n_i^{(k)}} \sum_{j=1}^{N} a_{ij}^{(k)} \Delta Y_{j,t}^{(k+1)} + \beta_2^{(k)} \Delta Y_{i,t}^{(k+1)} + \varepsilon_{i,t}^{(k)}, \tag{6}$$

where for the model in first differences we have $\Delta Y_{i,t}^{(k)} = \log(P_{i,t}^{(k)}) - \log(P_{i,t-1}^{(k)})$, whereas for the model on the weekly differences $\Delta Y_{i,t}^{(k)} = \log(P_{i,t}^{(k)}) - \log(P_{i,t-7}^{(k)})$.

Since the main focus of this paper is on building a competition network, it seems reasonable to compare the models in terms of the normalised RMSE, say $NRMSE_k = RMSE_k/(Y_{\max}^{(k)} - Y_{\min}^{(k)})$, where $Y_{\max}^{(k)} = \max_{i,t}(Y_{i,t}^{(k)})$ in order to account for the different scales of the dependent variables. Considering that the NRMSE is a measure of overall forecast accuracy lying in [0, 1], we can see from Table A1 that all the three models show good predictive performances with small NRMSE never exceeding a 7.4% error. However, in levels, the model has the smallest NRMSE for all the advance booking, apart from $k = 0$.

For space constraints, we do not report estimation results (available upon request) for models (6), but will mention that all the exogenous variables are not significant (intercept included), while the network and lagged effects are significant. Interpretations of such effects are in line with the model (1) (see Section 6).

6 Network interpretation

Based on the previous findings, we have identified 13 networks of competing agents (one for each advance booking $k = 0, \ldots, K - 1$). In this section, we first illustrate the suggested criteria to classify each hotel with its typical (i.e. more frequent) competitive behaviour in the booking window. Then we highlight the principal characteristics of the groups identified looking for differences and similarities.

Table 4. Regression estimates of model (1) obtained from Algorithm 1

k	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	$\hat{\gamma}_4$	RMSE
0	1.0033*	0.0345*	0.7392*	0.1441*	0.1108*	0.0519*	−0.0481*	0.4191
	(0.0172)	(0.0016)	(0.0038)	(0.0057)	(0.0067)	(0.0027)	(0.0026)	
1	0.807*	0.0287*	0.7976*	0.0829*	0.0729*	0.0449*	−0.0299*	0.3714
	(0.0154)	(0.0014)	(0.0034)	(0.005)	(0.0059)	(0.0024)	(0.0023)	
2	0.968*	0.0284*	0.7653*	0.1133*	0.1088*	0.0524*	−0.0091*	0.385
	(0.0159)	(0.0015)	(0.0034)	(0.0051)	(0.0061)	(0.0025)	(0.0023)	
3	0.6623*	0.0318*	0.8296*	0.0765*	0.0785*	0.0401*	−0.0384*	0.3528
	(0.0149)	(0.0013)	(0.0031)	(0.0047)	(0.0056)	(0.0023)	(0.0021)	
4	0.8151*	0.021*	0.8053*	0.0577*	0.0663*	0.0328*	−0.0197*	0.3643
	(0.0152)	(0.0014)	(0.0032)	(0.0049)	(0.0058)	(0.0024)	(0.0022)	
5	0.5235*	0.0184*	0.8732*	0.0188*	0.0627*	0.0231*	−0.0187*	0.3101
	(0.0131)	(0.0013)	(0.0028)	(0.0042)	(0.005)	(0.002)	(0.0019)	
6	0.8505*	0.0299*	0.7889*	0.036*	0.1408*	−0.0121*	−0.001	0.3665
	(0.0158)	(0.0014)	(0.003)	(0.0049)	(0.0057)	(0.0024)	(0.0022)	
7	0.6407*	0.0383*	0.831*	0.0408*	0.1554*	0.0401*	−0.0209*	0.3748
	(0.0162)	(0.0016)	(0.0033)	(0.005)	(0.0061)	(0.0025)	(0.0023)	
8	0.4546*	0.0214*	0.8918*	0.0495*	−0.0406*	0.0614*	−0.0179*	0.3078
	(0.0134)	(0.0012)	(0.0027)	(0.0041)	(0.0048)	(0.002)	(0.0019)	
9	0.7034*	0.0146*	0.8397*	0.0556*	0.0349*	0.0269*	−0.0047	0.3373
	(0.0146)	(0.0013)	(0.003)	(0.0046)	(0.0053)	(0.0022)	(0.002)	
10	0.7737*	0.0224*	0.8175*	0.0749*	−0.0323*	0.0401*	−0.0254*	0.3491
	(0.0151)	(0.0016)	(0.0032)	(0.0047)	(0.0055)	(0.0023)	(0.002)	
11	0.7773*	0.016*	0.8223*	0.0539*	0.0483*	0.023*	−0.0099*	0.3417
	(0.0146)	(0.0013)	(0.003)	(0.0046)	(0.0054)	(0.0022)	(0.002)	
12	0.7267*	0.0154*	0.8327*	0.0534*	0.0698*	0.0244*	−0.0081*	0.3461
	(0.0149)	(0.0013)	(0.0031)	(0.0046)	(0.0058)	(0.0022)	(0.0021)	

Note. Standard errors in brackets. $S = 1{,}000$ simulation performed. The minimum root mean squared error (RMSE) (5) obtained at the constructed network matrix is also reported. Significant coefficients at 5% (with Bonferroni correction) denoted by *.

6.1 Classification criteria

We start defining the following matrices which contain counts:

1. Row-to-column matrix (RtC) counting the number of advance bookings where hotels by row (i) follow hotels by column (j), and not vice versa, i.e. $i \to j$ but $j \nrightarrow i$.
2. Col-to-row matrix (CtR) counting the number of advance bookings where hotels by row (i) are followed by hotels by column (j), and not vice versa, i.e. $i \leftarrow j$ but $j \nleftarrow i$.
3. Mixed matrix counting the number of advance bookings where two different hotels, say i and j, have reciprocal follower–leader relation, i.e. $i \to j$ and $j \to i$.

For each couple of hotels (i, j), we then calculate the relative frequency of follower connections ($f_{i\to j}$), leader connections ($f_{i\leftarrow j}$) and mixed ones ($f_{i\leftrightarrow j}$) considering the total number of advance bookings (recall that $K = 13$). That way, $f_{i\perp j} = 1 - f_{i\to j} - f_{i\leftarrow j} - f_{i\leftrightarrow j}$ is the frequency of advance bookings where hotels i and j have independent behaviour.

Average frequencies are obtained as $\bar{f}_{\to} = \bar{f}_{\leftarrow} = \sum_{i,j=1}^{N} f_{i\to j}/n_{\to}$ and $\bar{f}_{\leftrightarrow} = \sum_{i,j=1}^{N} f_{i\leftrightarrow j}/n_{\leftrightarrow}$, where $n_{\to} = \sum_{i,j=1}^{N} I(f_{i\to j} \neq 0)$, $n_{\leftrightarrow} = \sum_{i,j=1}^{N} I(f_{i\leftrightarrow j} \neq 0)$ and $I(\cdot)$ is the indicator function. Note that $\bar{f}_{\to} = \bar{f}_{\leftarrow}$ because the matrix CtR corresponds to the transposed RtC matrix. The respective standard deviations $\sigma_{\to} = \sigma_{\leftarrow} = \sqrt{\mathrm{Var}(f_{i\to j})}$ and $\sigma_{\leftrightarrow} = \sqrt{\mathrm{Var}(f_{i\leftrightarrow j})}$ are also determined. The competitive behaviour of hotel i with respect to hotel j is classified as follows:

- follower if $f_{i\to j} > \max\{0.5, \bar{f}_{\to} + \sigma_{\to}\}$;
- leader if $f_{i\leftarrow j} > \max\{0.5, \bar{f}_{\leftarrow} + \sigma_{\leftarrow}\}$;
- mixed if $f_{i\leftrightarrow j} > \max\{0.5, \bar{f}_{\leftrightarrow} + \sigma_{\leftrightarrow}\}$.

To gain some intuition, the first rule implies that hotel i would be a follower of hotel j if the frequency of advance bookings where $i \to j$ is higher than the average overall frequency of followers plus its standard deviation; in any case such event should occur for more than half of the advance bookings.

Then we define the following binary indices: $m_{i\to j} = I(f_{i\to j} > \max\{0.5, \bar{f}_{\to} + \sigma_{\to}\})$, $m_{i\leftarrow j} = I(f_{i\leftarrow j} > \max\{0.5, \bar{f}_{\leftarrow} + \sigma_{\leftarrow}\})$, and $m_{i\leftrightarrow j} = I(f_{i\leftrightarrow j} > \max\{0.5, \bar{f}_{\leftrightarrow} + \sigma_{\leftrightarrow}\})$, taking value 1 if the condition in the brackets is true, 0 otherwise. Note that each hotel can have a different behaviour with respect to the others, so for example $m_{i\to j} = 1$, while $m_{i\leftarrow l} = 1$, for $j \neq l$.

Finally, we categorise the characteristic competitive behaviour of hotel i looking at the following sums $m_{i\to} = \sum_{j=1}^{N} m_{i\to j}$, $m_{i\leftarrow} = \sum_{j=1}^{N} m_{i\leftarrow j}$ and $m_{i\leftrightarrow} = \sum_{j=1}^{N} m_{i\leftrightarrow j}$. In particular, we apply the following criteria:

- Independent, if $m_{i\to} = m_{i\leftarrow} = m_{i\leftrightarrow} = 0$. An independent hotel has no relationship with competitors, either as a leader or as a follower.
- Mixed, if $m_{i\to} = m_{i\leftarrow} = 0$ and $m_{i\leftrightarrow} \neq 0$ or if $m_{i\to} = m_{i\leftarrow}$. A mixed hotel only has reciprocal relationships being, alternatively, leader or follower of a second hotel in the advance booking window. However, we also consider 'mixed' as the behaviour of a hotel with an equal number of follower and leader relationships (with different hotels).
- Leader, if $m_{i\to} < m_{i\leftarrow}$. A hotel is (typically) a leader if it turns out to be a leader by at least one more hotel than the number it follows. For example, if a hotelier is a follower 1 time, a leader 3 times and is mixed 4 times, with respect to other hotels, then, by considering the mixed result as equivalent to leader and follower at the same time, the hotel will be ranked a leader for a total of 7 times and a follower 5 times. This hotel would then be classified as a leader.
- Follower, if $m_{i\to} > m_{i\leftarrow}$. A hotel is (typically) a follower if it follows at least one hotel more than the hotels considering it a leader.

6.2 Analysis of the profiles of the competitive behaviours

In this section, we study decision makers' propensity to lead or follow competitor pricing strategies as a function of their structural characteristics. We first apply the classification criteria developed

Table 5. Average statistics for classified competitive relations on all advance booking ($k = 0, \ldots, 13$)

Category	Frequency	BAR	Stars	Restaurant	Meeting	Num. rooms	Num. hotels in 200 m
Follower	14	158.77	3.6	0.43	0.21	48.5	7.7
Independent	40	200.00	3.7	0.35	0.28	47.9	3.1
Leader	13	150.19	3.2	0.15	0.00	31.5	6.8
Mixed	28	181.60	3.5	0.18	0.11	45.4	6.5

Note. Maximum levels in green, minimum levels in red. BAR = best available rate.

Table 6. Average statistics for classified competitive relations on last-minute booking ($k \leq 3$)

Category	Frequency	BAR	Stars	Restaurant	Meeting	Num. rooms	Num. hotels in 200 m
Follower	33	177.80	3.5	0.33	0.15	41.6	6.0
Independent	14	214.17	3.9	0.36	0.36	68.1	1.9
Leader	32	175.25	3.6	0.22	0.19	41.7	6.7
Mixed	16	174.16	3.3	0.25	0.06	38.4	4.1

Note. Maximum levels in green, minimum levels in red. BAR = best available rate.

in Section 6.1, then we calculate statistics for each of the four different types of competitors that coexist in Venice. This allows us to connect management practices and investment opportunity to competition patterns.

From Section 3, we know that there is a clear difference between last-minute ($k \leq 3$) and early-booking ($k > 3$) pricing behaviour. Accordingly, we find a different panorama if we focus on the competitive behaviour over the whole booking window considered—see Table 5—or in the first four advance booking days ($k = 0, \ldots, 3$; see Table 6).

When we look at the competition along a two-week advance booking window (see Table 5), we find that the independent hotels are the most expensive and the highest rated. They target richer customers and therefore are expected to maintain high and homogeneous prices along the advance booking period to communicate the exclusive offers. Looking at size and frequency of meeting rooms and restaurants, these hotels appear able to target large events like general meetings, congresses and large exhibitions. This is a particular market segment where it is common to sell a large number of rooms, well in advance, in the off-line market at special negotiated rates. These prices become (unobservable) thresholds for the online rates that cannot be crossed. The likelihood of competing on price is also reduced by their exclusive location, with a very small number of potential competitors in a 200 m radius.

Leaders, are hotels with low star rating, without meeting rooms and few restaurants. They target the low-price customers and their prices are more flexible. This is another market segment with customers booking (online) well in advance with respect to the arrival date. This characteristic makes them 'price makers' leading the online (early-booking) price competition.

Followers are similar to independent hotels except that they cannot leverage an exclusive location because they are located in the area of the highest competitor density in a 200 m radius. They offer meeting rooms and restaurants (also), positioning themselves in the business segment. However, their low average price tells us that they cannot even differentiate themselves from their competitors with respect to the services offered. Thus, price becomes a main competitive factor forcing them to follow one or more leaders in the neighbourhood.

Finally, the proposed network-based methodology, discovers the presence of many hotels that are simultaneously leaders and followers. Mixed hotels associate a low average star rating with a high average price, a clear indication that they are able to differentiate themselves from the competitors leveraging exclusive tangible, reputation or, more likely, context attributes. Their 'average' characteristics allow them to target the whole spectrum of business and leisure customers, from individual or small groups of business travellers (i.e. lower added value tourists compared to conference and incentive customers) to the richest part of the leisure segment (i.e. wedding). This wide 'business mix' increases the spectrum of pricing strategies, allowing them to be both leaders or followers over the whole 13 days of advance booking considered here.

When we focus on the competition of the last-minute bookings ($k = 0, \ldots, 3$), we get quite a different picture. Price competition becomes stronger and many hotels that were independent or mixed, when observed during a two-week booking window, become leaders or followers. Roughly speaking the independent hotels with the worst characteristics begin to compete on price while the larger and more expensive mixed-behaviour hotels (with the best features) 'specialise' their actions becoming price takers or price makers.

The hotels leading the last-minute online price competition are mainly located in areas with a high concentration of competitors (which can, therefore, be considered areas of greater interest for demand). With respect to the early-booking framework, they show a higher average star rating and a non-zero quote of hotel with meeting rooms. This indicates that the last-minute leaders are also composed of well-equipped business-hotels seeking customers in the leisure segment.

Thus, we can conclude they are prone to offering the last (low-quality) unsold rooms at lower rates, leading the last-minute price competition with aggressive last-minute discount policies. This is not due to them having the largest number of competitors in the neighbourhood, as they are in a situation where anticipating rate reductions could be a worth differentiation strategy.

The number of followers increase significantly when $k \leq 3$, especially by reducing the number of mixed hotels. Within a few days of the arrival date, the perceived performance gap with competitors appears more difficult to bridge and pairs of hotels that, observed on a larger booking window, influence each other, start to mimic one leader tactics. Last-minute followers have a lower number of stars and a higher average price (with respect to followers at the higher advance bookings). Thus, the prototype of last-minute follower (in Venice) is a hotel which is able to leverage a more 'exclusive' location in terms of rates (see also the sharp decrease in the average number of hotels in a radius of 200 m) but without enough services.

Hotels that remain independent in the last-minute period have the highest number of stars and size (the best tangible attributes to target both high-income business and leisure clients) and very few competitors nearby (1.9 on average). Therefore, they can leverage their 'unique' location and features to apply proper pricing strategies. As we know that the positive effect of horizontal differentiation is higher for higher priced hotels (see Sánchez-Pérez et al., 2020), we can infer that these hotels compete by implementing horizontal differentiation strategies, e.g. maintaining a good online reputation or offering special features, with a very low tendency to change rates at the last minute. In this way, they also manage the reference price that customers desire, which is key to travellers perception of price acceptability.

Finally, the cluster of hotels that are simultaneously leaders and followers in competition at the last minute constitutes hotels with the lowest price, size and number of star (mainly 3-star hotels). However, the most important distinctive feature is that they are located in areas with a low concentration of competitors. This allows them to play as leader (or follower) even if they have very little chance of differentiating themselves from competitors in terms of tangible attributes. In other words, we can infer that they are small-sized low-rate ('periphery') hotels that aim to saturate capacity by leading or imitating (a few) competitors' simple and aggressive last-minute discount policies.

7 Conclusions

We study dynamic price competition in the hotel market in Venice through the lens of publicly available data scraped from an OTA, monitoring three dimensions: individuals (hotels), arrival days and booking days (advance booking).

This type of data set poses two main challenges. First, the time series of prices recorded for each hotel possess a twofold time frame. In fact, every single price for an overnight stay on a certain day, corresponds to a time series of asking prices along the booking window. Second, the competition relations between hoteliers are unknown and need to be represented dynamically.

In line with the economics and management literature, we conjecture that competitive behaviour can be represented by a network architecture whose edges are drawn by observing the (public) actions and responses of different competing agents. Accordingly, we propose a novel network autoregressive model, with time-varying coefficients over the booking window, suited to handle the peculiar threefold data structure of the data set. Identifying the structure of the competition network is the final goal of our contribution.

The approach we propose can be employed to study competition behaviour on every market where consumption is delayed with respect to the purchasing time. Considering that E-commerce and selling online has significantly increased in recent years (with a consequent increasing availability of data), we think that we are taking up a major challenge for both academics and managers. Accordingly, this paper provides both methodological and empirical results as it studies decision-making through a strong focus on context, accepting complex and unclear causality, shaping the competition between decision makers on the basis of their 'real-life' behaviour.

7.1 Methodological results

We introduce a new network autoregressive model useful to discover unknown competition patterns between hotels by using a purely data-driven algorithm. In particular, the model employs the predictive accuracy along advance booking as a metric to determine the optimal network architecture, modelling competition only by observing the everyday organisational contexts. Parameters are easily estimated following a least squares method as we rewrite the problem as a general TN-dimensional linear model. Additionally, we provide the solution to the general linear system establishing consistency and asymptotic normality of the model estimators.

To the best of our knowledge, this is the first time that the dynamic part of a network architecture is based on pricing practices along the advance booking for each arrival time, instead of on the lagged time effects typically encountered in the previous literature. As a consequence, the parameters of the models and the network connections can vary with the advance booking to account for possible difference in the competition patterns along the booking window.

7.2 Empirical and managerial results

Venice is a very popular destination with a strong seasonality, driven by world-famous events such as Carnival or the Biennale, and a large variety of hotels prone to dynamic pricing. In our sample of hotels with 3–5 stars, we find that large and expensive accommodations, offering exclusive accommodation and business-oriented services, tend to keep a more stable pricing pattern across arrival days. More in general, looking at the advance booking effect on the likelihood of relying on dynamic pricing, we show that pricing strategies become more homogeneous during last-minute reservation period. As a consequence, we find a stronger network effect at the last minute, when the risk of both being left with unsold rooms and experiencing a performance gap with competitors becomes a trigger to imitating and learning from others.

In line with previous literature (Roy & Raju, 2011), we categorise three typical price competition patterns: independent, where firms adjust prices without a clear relationship to competitor pricing policies and leader (or) follower, where a firm anticipates or follows the rate changes of another one. We also find a fourth (mixed) behaviour, where a firm is leader or follower, depending on the advance booking period (i.e. the time-lag between purchase and consumption).

We shape these competition behaviours through different self-building competition networks whose edges are built looking at competitors' actions and reactions along the advance booking. We find a positive and significant network effect, i.e. we show that the rates published by a hotel, for a stay in t, at advance booking k also depends on the rates published on $k + 1$ by the competitors (same arrival day). Moreover, this network effect has an intensity that follows a linear deterministic relation, indicating that the reaction to competitors' pricing policies increases in intensity as the advance booking period decrease.

However, the most important determinant of the online rate in k is the price published by the hotelier himself in $k + 1$, same arrival day t. A result suggesting that hoteliers fix prices looking at both the (expected) demand seasonality at different arrival days t and the (expected) 'attractiveness' of the tangible, reputation or context attributes on the (expected) demand mix in t. This autoregressive component shows an estimated coefficient that reaches its minimum at $k = 0$ allowing us to conclude that the effect of hoteliers' expectations on dynamic pricing choices reaches its minimum when the risk of both being left with unsold rooms and/or having a performance gap with a competitor triggers them to imitate and learn from others.

We also find that the overall spatial density pressure—measured by Z_4, the number of hotels in a 200 m radius—exerts a negative effect on the rate level. Spatial density measures both agglomeration and competition effects. The hospitality literature has shown that a higher level of competition can decrease the price of a service (Becerra et al., 2013) but, on the other hand, it has also provided evidence of a positive relation between agglomeration and rates level; see Sánchez-Pérez et al. (2020). We find evidence that advance booking modifies the balance between the two effects. In fact, the negative effect of hotels' spatial density on the rate level is higher in the last-minute period than in the early booking, leaving room for the hypothesis that agglomeration benefits succeed in partially balancing competition when it is weaker (i.e. in the early booking).

When we look at the competitive behaviour, we show that the advance bookings, in addition to having a strong impact on the attitude towards dynamic competition, plays a crucial role in determining the reactions to competitors' decisions, i.e. the businesses typical competitive behaviour. Price competition becomes stronger at the last minute, and many hotels categorised as independent or mixed, if observed on a two-week booking window, start to be leaders or followers.

Furthermore, we find significant differences among the different competition profiles, in terms of: average BAR, capacity, services, location and density of neighbouring hotels. Roughly speaking, the independent hotels with the worst characteristics begin to compete on price while the larger and more expensive mixed-behaviour hotels (with the best amenities) 'specialise' their actions becoming price takers or price makers.

More in detail, in the early-booking period, the independent hotels are the most expensive and differentiated (in terms of services). They tend to maintain higher and more stable prices both for image and price-fairness reasons (e.g. high price communicates exclusivity to the high value leisure segment). Moreover, they can offer services (e.g. meeting rooms) to target both the leisure market segment (less elastic to price) and the business segment which is more attentive to price-quality and price-fairness issues. In fact, business-men are usually frequent buyers with a clear reference price, which is key to their perception of price acceptability. Independents strengthen in this profile at the last minute when the less isolated and lower rated hotels exit this cluster. The propensity to compete on price is reduced even further (the average price in this cluster increases sharply despite the decreasing average price at the destinations, see Table 1 suggesting that they mainly compete using horizontal differentiation strategies.

The cluster of hotels that are both leaders and followers (mixed) in the 13-day booking window considered, is composed of by hotels with average characteristics, allowing them to target the whole spectrum of business and leisure customers. They differentiate themselves from the competitors leveraging the whole spectrum of attributes: price, reputation and/or context, avoiding systematically resorting to 'price wars', as demonstrated by the fact they combine a low average star rating with a high average price. In the last-minute period, the number of hotels in this cluster reduces significantly. Hotels with more rooms and especially those located in the most important tourist areas (with the highest average prices) exit this cluster, increasing the number of pure leaders or pure followers. Consequently, the typical profile of a hotel acting both as a leader or a follower is 3-star hotels that are not differentiated in terms of tangible attributes, located in areas with a low concentration of competitors, willing to practice very aggressive last-minute discount policies leading or imitating (few) competitors.

The hotels leading the early-booking price competition are hotels with a low star rating that target the segment of more price-elastic and early leisure (online) customers. For them, dynamic pricing appears as the only possible differentiation strategy. Leaders grow in number at the last minute where well-equipped business-hotels seeking customers in the leisure segment join the cluster. As a consequence, the typical last-minute leader has a higher number of stars and amenities with respect to the early-booking leader. It also has a very high number of nearby competitors and is in a situation where is worth getting ahead of the competitors by combining dynamic pricing and horizontal differentiation strategies. As an example, it can promote special offers on low-quality and/or low-services rooms in order to attract customers that accept paying more than their standard price for a hotel with exclusive features and/or a higher reputation with respect to the hotels located nearby.

Followers in the early-booking price competition are more similar to independent hotels except that they cannot leverage an exclusive location. In fact, they are located typically in the San Marco District, with few ways to differentiate themselves from leaders. The number of followers hotels strongly increases in the last-minute period when the perceived performance gap with competitors appears more difficult to bridge. Smaller and low-rated hotels join the cluster, especially if located in areas with a lower concentration of tourist attractions. They can leverage a more 'exclusive' location in terms of a higher rate (compared to early-booking followers) even if they do not have enough services or strategic skills to develop their own online pricing tactics.

This work gives managers and stakeholders a powerful tool to make informed decisions. As a revenue manager, unveiling competition network helps leveraging structural and quality differences with competitors (i.e. to increase revenues). For example, discovering that many followers cannot offer a room with an exclusive view attaches added value to those hotel rooms with this unique feature. Similarly, a manager who discovers that his/her hotel is not followed by anyone does not worry about reactions of the competition to any pricing policies. Even knowing that there are many similar hotels following a certain leader allows managers to change (or choose) pricing policy. In terms of policy, uncovering quality and size of competition between companies enables assessing results of product or communication policies developed over time. Moreover, the possibility of analysing the competition at a high spatial level and along the advance booking window grants control (and prediction) of the potential hazards of a massive and uncoordinated influx of tourists. The competition in the early booking informs about (expected) daily tourist demand peaks (positive or negative). This is a useful information to manage short-term operational tactics so that avoiding asset shortfalls/excess that tourists share with residents (i.e. transportation, security, water , or urban space), and a more efficient management of budget and public resources.

7.3 Discussion

This contribution gives several directions for future research. For example, the NAR model (1) assumes that only hotels directly followed by i (i.e. $i \to j$) possibly have an impact on its BAR. This assumption can be relaxed by allowing also a second order relationship among hotels, i.e. the competing hotels of competitors have an impact on each hotel, that is $i \to j$ and $j \to l$ but $i \nrightarrow l$. In this way, we can also study the effect $i \to j \to l$. Innovation terms might be correlated which implies that a different estimation technique, like generalised least squares, should be developed. Finally, extending the network model to include time-varying competition probabilities (4) will give further insight on the way that competition networks evolve.

A limitation of this study is that the methodology can only be applied to accommodation structures that regularly publish prices online and it requires a data collection process. As to date, there is no historical repository of hotel rates. We have not considered factors such as the hotel position on the Booking.com search page and the rating given by customers, which could also be considered to better explain pricing choices. We show that data from the OTAs are an effective example of (implicit) shared knowledge regarding business price competition tactics challenging statisticians. However they do not inform about factors which might be important for determining the online price of a room. For example, whether or not the hotel has undergone recent renovation, if it relies on external professionals for pricing management or whether it reserves a significant portion of rooms for off-line sales. By appealing to expert knowledge regarding the unique features of Venice tourism market, we define a competition network by employing reasonable assumptions related to distance and pricing policies of hotels. If a different competition network was under consideration, i.e. tourism market of New York, then competition probabilities need to be defined in a different way to take into account the new networks' particular characteristics. However, methodology and algorithmic details developed in this work are still applicable and offer useful insights, especially to markets where consumption is delayed when there exists fixed capacity (e.g. tickets for events and exhibitions, seats in means of transportation).

Acknowledgments

The authors cordially thank the Editor, the AE, and two reviewers for constructive comments which improved as earlier version of the manuscript. This work was done while M.A. was with the Department of Mathematics and Statistics at the University of Cyprus.

Conflict of interest: None declared.

Funding

M.A. acknowledges financial support from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 101108797.

Data availability

Data are available only for academic purposes and can be obtained from the authors after asking for permission. Code to replicate the analyses in the paper is available at https://github.com/mirkoarmillotta/Venice˙hotel.

Appendix A

A.1 Additional tables and plots

Table A1. Normalied root mean squared errors (NRMSE) for models estimated by Algorithm 1 in the levels (Y), in the first ($\Delta 1$) and seven days ($\Delta 7$) differences

k	0	1	2	3	4	5	6	7	8	9	10	11	12
Y	0.071	0.061	0.063	0.057	0.060	0.049	0.058	0.060	0.051	0.059	0.060	0.055	0.055
$\Delta 1$	0.069	0.064	0.065	0.060	0.061	0.057	0.065	0.066	0.053	0.063	0.062	0.065	0.061
$\Delta 7$	0.074	0.066	0.068	0.066	0.067	0.060	0.069	0.067	0.058	0.063	0.063	0.063	0.055

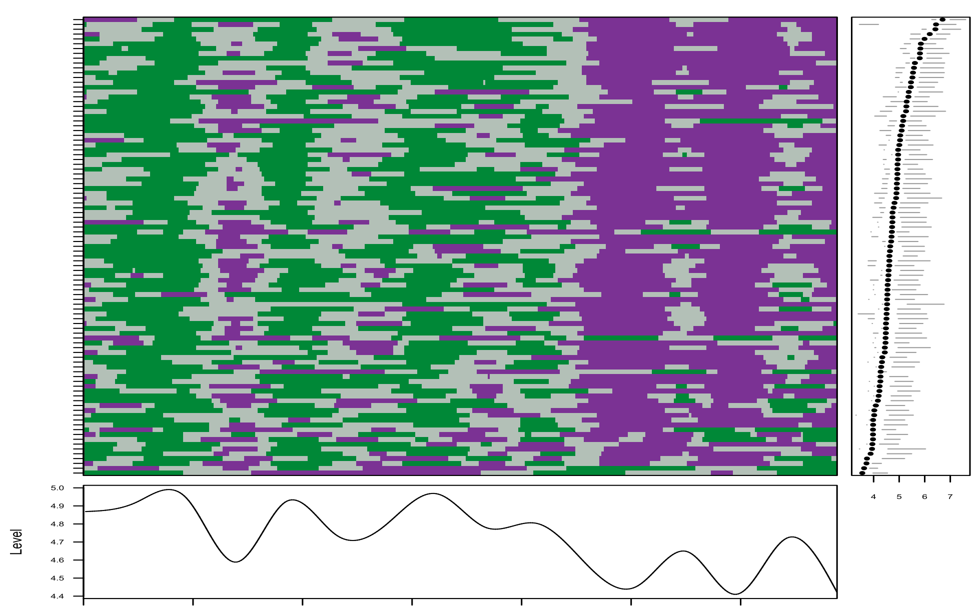

Figure A1. Natural spline smoother of the multivariate time series of log-transformed best available rates for all hotels at advance booking $k = 0$ showed in Figure 1. Hotels by row ordered in descending order by median log-price. Time in column (1 April 2019–9 March 2020). Colours reflect threefolds of log-price distribution: high (green), mid (grey), low (purple) obtained by dividing each hotel's log-price distribution in tertiles. Right panel: box plot distribution around the median (full dot) for each hotel time series. Below: time series of hotel median log-price levels.

A.2 Consistency and asymptotic normality of advance booking network autoregressive model estimator

Under the following suitable conditions, consistency and asymptotic normality of the least squares estimator (3) are derived:

1. For $k = 0, \ldots, K - 1$, the errors $\varepsilon_{i,t}^{(k)}$ are iid with mean 0 and standard deviation σ_k.
2. The covariates Z_i are independent and iid random vectors with mean vector 0 and covariance matrix Q_z.
3. The errors related to different advance bookings are independent, in symbols $\varepsilon^{(k)} \perp \varepsilon^{(s)}$, for $k \neq s$.
4. Errors and covariates are mutually independent, $\varepsilon^{(k)} \perp Z$, for $k = 0, \ldots, K - 1$.
5. $Q^{(k+1)}$ is full rank, for $k = 0, \ldots, K - 1$.

Assumption 1 is a typically required condition for least squares estimation and it would be realistic to assume that the errors regarding different hotels and different arrival dates are not dependent. Assumption 2 states a similar condition for the covariates, this is also a standard requirement. It is also realistic to assume that exogenous covariates for each hotels were generated independently from the same process. Assumption 3 is an extension of 1, by considering the independence of the errors also over the advance bookings. Assumptions 4 and 5 are standard in least squares methods and correspond to exogeneity of the regressors Z.

Define $\theta_0^{(k)}$ the true value of the parameters of model (1) and $\xrightarrow{p}$, $\xrightarrow{d}$ denote the convergence in probability and in distribution, respectively. Finally, set $\mathrm{E}(Q_{i,t}^{(k+1)} Q_{i,t}^{(k+1)\prime}) = Q_k$, where the $1 \times m$

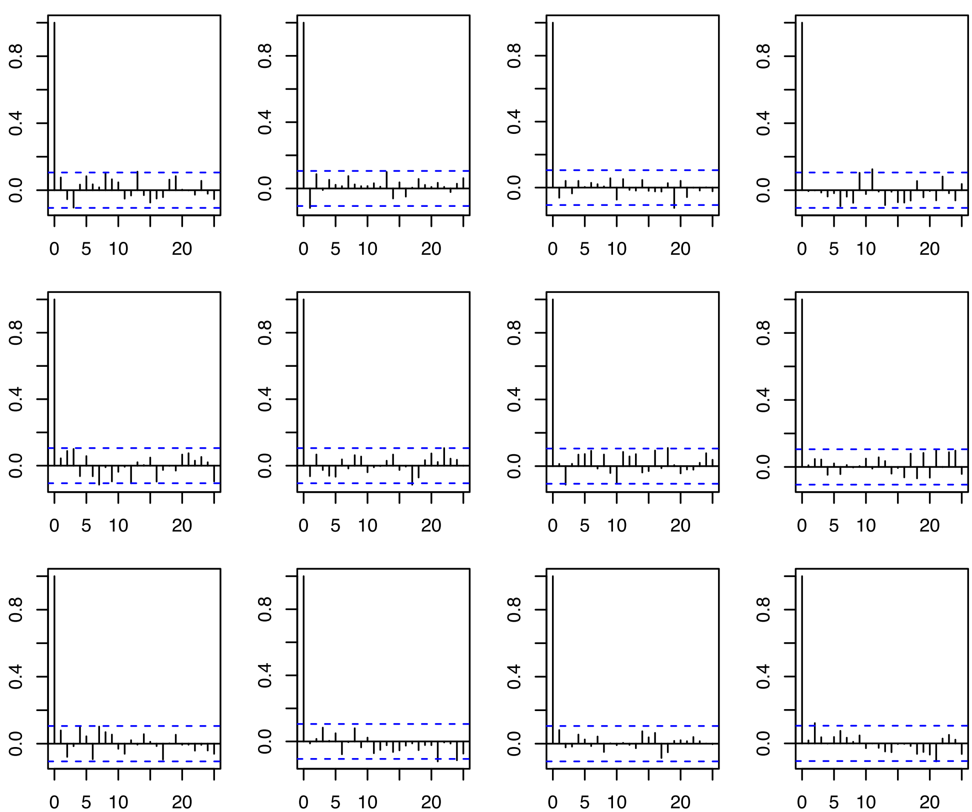

Figure A2. Model standardised residuals' autocorrelation function for 12 randomly chosen hotels at advance booking $k = 0$.

row vector $Q_{i,t}^{(k+1)\prime}$ is the (i, t) row of the matrix $Q^{(k+1)}$. We can now derive the large sample properties of the estimator (3) when both the number of the hotels and the temporal sample size grow together. We define $\min\{N, T\} \to \infty$ as shorthand for simultaneous double asymptotic regime $T \to \infty$ and $N \to \infty$, where N and T are not constrained to be related.

Proposition A.1 Assume assumptions 1–5 hold. Then, for $k = 1, \ldots, K-1$, as $\min\{N, T\} \to \infty$, $\hat{\theta}^{(k)} \xrightarrow{p} \theta_0^{(k)}$ and $\sqrt{NT}(\hat{\theta}^{(k)} - \theta_0^{(k)}) \xrightarrow{d} N(0, \sigma_k^2 Q_k^{-1})$.

Proof. Assuming the linearity of the model, the consistency of the estimator follows by A2–A4 and equations (4)–(19) in Greene (2018). A2 is our assumption 5. Assumptions 1, 3, and 4 imply A3 and A4. Equations (4)–(19) is obtained by Greene (2018, A5a), which is satisfied in our case by adding assumption 2 to 1, 3 and 4. Under the same set of assumptions, the Grenander conditions described in Greene (2018, Tab. 4.2) hold for the regressors $Q^{(k+1)}$. Then, the asymptotic normality of the estimator follows by Greene (2018, Thm. 4.3). □

References

Abrate G., & Viglia G. (2016). Strategic and tactical price decisions in hotel revenue management. *Tourism Management*, *55*, 123–132. https://doi.org/10.1016/j.tourman.2016.02.006

Arabadzhyan A., Figini P., & Zirulia L. (2021). Hotels, prices and risk premium in exceptional times: The case of Milan hotels during the first COVID-19 outbreak. *Annals of Tourism Research Empirical Insights*, *2*(2), 100023. https://doi.org/10.1016/j.annale.2021.100023

Armillotta M., & Fokianos K. (2022a). 'Poisson network autoregression', arXiv, arXiv:2104.06296, preprint: not peer reviewed.

Armillotta M., & Fokianos K. (2022b). 'Testing linearity for network autoregressive models', arXiv, arXiv:2202.03852, preprint: not peer reviewed.

Armillotta M., Fokianos K., & Krikidis I. (2022). Generalized linear models network autoregression. In *International Conference on Network Science* (pp. 112–125). Springer.

Armillotta M., Luati A., & Lupparelli M. (2022). Observation-driven models for discrete-valued time series. *Electronic Journal of Statistics*, *16*(1), 1393–1433. https://doi.org/10.1214/22-EJS1989

Armillotta M., Tsagris M., & Fokianos K. (2022). 'The R-package PNAR for modelling count network time series', arXiv, arXiv:2211.02582, preprint: not peer reviewed.

Baum J. A., & Dahlin K. B. (2007). Aspiration performance and railroads' patterns of learning from train wrecks and crashes. *Organization Science*, *18*(3), 368–385. https://doi.org/10.1287/orsc.1060.0239

Becerra M., Santaló J., & Silva R. (2013). Being better vs. being different: Differentiation, competition, and pricing strategies in the Spanish hotel industry. *Tourism Management*, *34*, 71–79. https://doi.org/10.1016/j.tourman.2012.03.014

Bernanke B. S., Boivin J., & Eliasz P. (2005). Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach. *The Quarterly Journal of Economics*, *120*(1), 387–422. https://doi.org/10.1162/0033553053327452

Bigné E., & Decrop A. (2019). Paradoxes of postmodern tourists and innovation in tourism marketing. In: E. Fayos-Sola, & C. Cooper (Eds.), *The future of tourism* (pp. 131–154). Springer.

Bigné E., Nicolau J. L., & William E. (2021). Advance booking across channels: The effects on dynamic pricing. *Tourism Management*, *86*, 104341. https://doi.org/10.1016/j.tourman.2021.104341

Chica-Olmo J. (2020). Effect of monumental heritage sites on hotel room pricing. *International Journal of Hospitality Management*, *90*, 102640. https://doi.org/10.1016/j.ijhm.2020.102640

Choi S., Kim W., & Kim N. (2022). International alliance formations: The role of brokerage in technology competition networks. *Journal of Business Research*, *144*, 440–449. https://doi.org/10.1016/j.jbusres.2022.02.007

Chung H. C., Chung N., & Kim J.-Y. (2021). How do online hotel consumers perceive room rates? *Journal of Vacation Marketing*, *28*(3), 350–365. https://doi.org/10.1177/13567667211066326

Cliff A., & Ord J. K. (1975). Space-time modelling with an application to regional forecasting. *Transactions of the Institute of British Geographers*, *64*, 119–128. https://doi.org/10.2307/621469

Cohen M. D., March J. G., & Olsen J. P. (1972). A garbage can model of organizational choice. *Administrative Science Quarterly*, *17*(1), 1–25. https://doi.org/10.2307/2392088

Cyert R. M., & March J. G. (1963). *Behavioral Theory of the Firm*. Wiley.

Dabas S., & Manaktola K. (2007). Managing reservations through online distribution channels: An insight into mid-segment hotels in India. *International Journal of Contemporary Hospitality Management*, *19*(5), 388–396. https://doi.org/10.1108/09596110710757552

Dolnicar S. (2002). Business travellers' hotel expectations and disappointments: A different perspective to hotel attribute importance investigation. *Asia Pacific Journal of Tourism Research*, 7(1), 29–35. https://doi.org/10.1080/10941660208722107

Gavilan D., Avello M., & Martinez-Navarro G. (2018). The influence of online ratings and reviews on hotel booking consideration. *Tourism Management*, 66, 53–61. https://doi.org/10.1016/j.tourman.2017.10.018

Gilardi A., Jorge M., Riccardo B., & Lovelace R. (2022). Multivariate hierarchical analysis of car crashes data considering a spatial network lattice. *Journal of the Royal Statistical Society Series A*, *185*(3), 1150–1177. https://doi.org/10.1111/rssa.12823

Gimeno J. (2004). Competition within and between networks: The contingent effect of competitive embeddedness on alliance formation. *Academy of Management Journal*, *47*(6), 820–842. https://doi.org/10.2307/20159625

Greene W. H. (2018). *Econometric analysis*. Pearson.

Guizzardi A., Ballestra L. V., & D'Innocenzo E. (2022). Hotel dynamic pricing, stochastic demand and covid-19. *Annals of Tourism Research*, *97*, 103495. https://doi.org/10.1016/j.annals.2022.103495

Guizzardi A., Monti A., & Ranieri E. (2016). Rating hotel quality for corporate business travel departments. *International Journal of Contemporary Hospitality Management*, *28*(12), 2842–2863. https://doi.org/10.1108/IJCHM-07-2015-0357

Guizzardi A., Pons F. M. E., Angelini G., & Ranieri E. (2021). Big data from dynamic pricing: A smart approach to tourism demand forecasting. *International Journal of Forecasting*, *37*(3), 1049–1060. https://doi.org/10.1016/j.ijforecast.2020.11.006

Guizzardi A., Pons F. M. E., & Ranieri E. (2019). Competition patterns, spatial and advance booking effects in the accommodation market online. *Tourism Management*, *71*, 476–489. https://doi.org/10.1016/j.tourman.2018.10.028

Kim M., Lee S. K., & Roehl W. S. (2018). Competitive price interactions and strategic responses in the lodging market. *Tourism Management*, *68*, 210–219. https://doi.org/10.1016/j.tourman.2018.03.013

Knight M., Leeming K., Nason G., & Nunes M. (2020). Generalized network autoregressive processes and the GNAR package. *Journal of Statistical Software*, *96*(5), 1–36. https://doi.org/10.18637/jss.v096.i05

Kolaczyk E. D., & Csárdi G. (2014). *Statistical analysis of network data with R* (*Vol. 65*). Springer.

Lavie D. (2021). Theoretical foundation for the study of competition networks and their performance implications. *Strategic Management Review*, to appear.

Lazer D., Kennedy R., King G., & Vespignani A. (2014). The parable of Google flu: Traps in big data analysis. *Science*, *343*(6176), 1203–1205. https://doi.org/10.1126/science.1248506

Lebacher M., Thurner P. W., & Kauermann G. (2021). A dynamic separable network model with actor heterogeneity: An application to global weapons transfers. *Journal of the Royal Statistical Society Series A*, *184*(1), 201–226. https://doi.org/10.1111/rssa.12620

Lee S. H. (2016). How hotel managers decide to discount room rates: A conjoint analysis. *International Journal of Hospitality Management*, *52*, 68–77. https://doi.org/10.1016/j.ijhm.2015.09.014

Li J., Netessine S., & Koulayev S. (2018). Price to compete··· with many: How to identify price competition in high-dimensional space. *Management Science*, *64*(9), 4118–4136. https://doi.org/10.1287/mnsc.2017.2820

Malc D., Mumel D., & Pisnik A. (2016). Exploring price fairness perceptions and their influence on consumer behavior. *Journal of Business Research*, *69*(9), 3693–3697. https://doi.org/10.1016/j.jbusres.2016.03.031

March J. G., & Simon H. A. (1993). *Organizations*. John Wiley & Sons.

Martin R. L., & Oeppen J. (1975). The identification of regional forecasting models using space: Time correlation functions. *Transactions of the Institute of British Geographers*, *66*, 95–118. https://doi.org/10.2307/621623

Mathur S., & Dewani P. P. (2016). Influence of cultural heritage on hotel prices, occupancy and profit: Theory and evidence. *Tourism Economics*, *22*(5), 1014–1032. https://doi.org/10.5367/te.2015.0487

Mohammed I., Guillet B. D., & Law R. (2019). Last-minute hotel-booking and frequency of dynamic price adjustments of hotel rooms in a cosmopolitan tourism city. *Journal of Hospitality and Tourism Management*, *41*, 12–18. https://doi.org/10.1016/j.jhtm.2019.08.005

Mohr L. B. (1978). Ambiguity and choice in organizations. *American Political Science Review*, *72*(3), 1033–1035. https://doi.org/10.2307/1955137

Nason G. P., & Wei J. L. (2022). Quantifying the economic response to COVID-19 mitigations and death rates via forecasting purchasing managers' indices using generalised network autoregressive models with exogenous variables. *Journal of the Royal Statistical Society Series A*, *85*(4), 1778–1792. https://doi.org/10.1111/rssa.12875

Olsen M. D., & Roper A. (1998). Research in strategic management in the hospitality industry. *International Journal of Hospitality Management*, *17*(2), 111–124. https://doi.org/10.1016/S0278-4319(98)00012-7

Park C., Kim Y.-R., & Frye W. D. (2022). Keeping the competition close: The impact of competitor distance in the lodging industry. *International Journal of Tourism Research*, *24*(3), 402–412. https://doi.org/10.1002/jtr.2510

Pellinen J. (2003). Making price decisions in tourism enterprises. *International Journal of Hospitality Management*, *22*(2), 217–235. https://doi.org/10.1016/S0278-4319(03)00019-7

Provenzano D., & Baggio R. (2020). E-tourism economics: Editorial for the special issue. *Tourism Economics*, *26*(6), 847–852. https://doi.org/10.1177/1354816619890222

Rezvani E., Assaf A. G., Uysal M., & Lee M. (2019). Learning from own and others: The moderating role of performance aspiration. *International Journal of Hospitality Management*, *81*, 113–119. https://doi.org/10.1016/j.ijhm.2019.03.001

Roy A., & Raju J. S. (2011). The influence of demand factors on dynamic competitive pricing strategy: An empirical study. *Marketing Letters*, *22*(3), 259–281. https://doi.org/10.1007/s11002-010-9124-z

Sánchez-Pérez M., Illescas-Manzano M. D., & Martínez-Puertas S. (2020). You're the only one, or simply the best. Hotels differentiation, competition, agglomeration, and pricing. *International Journal of Hospitality Management*, *85*, 102362. https://doi.org/10.1016/j.ijhm.2019.102362

Skilton P. F., & Bernardes E. (2015). Competition network structure and product market entry. *Strategic Management Journal*, *36*(11), 1688–1696. https://doi.org/10.1002/smj.2318

Smallman C., & Moore K. (2010). Process studies of tourists' decision-making. *Annals of Tourism Research*, *37*(2), 397–422. https://doi.org/10.1016/j.annals.2009.10.014

Tsagris M., Armillotta M., & Fokianos K. (2022). *PNAR: Poisson network autoregressive models*. R package version 1.4. https://CRAN.R-project.org/package=PNAR.

Urtasun A., & Gutiérrez I. (2006). Hotel location in tourism cities: Madrid 1936–1998. *Annals of Tourism Research*, *33*(2), 382–402. https://doi.org/10.1016/j.annals.2005.12.008

Wasserman S., & Faust K. (1994). *Social network analysis: Methods and applications* (*Vol. 8*). Cambridge University Press.

Yao E., Ferrier W. J., Yu T., & Labianca G. J. (2008). Rivals by association: Competition networks and firm performance. In *Atlanta Competitive Advantage Conference*, Atlanta, Georgia, US.

Zhou J., Li D., Pan R., & Wang H. (2020). Network GARCH model. *Statistica Sinica*, *30*(4), 1–18. https://doi.org/10.5705/ss.202018.0234

Zhu X., & Pan R. (2020). Grouped network vector autoregression. *Statistica Sinica*, *30*(3), 1437–1462. https://doi.org/10.5705/ss.202017.0533

Zhu X., Pan R., Li G., Liu Y., & Wang H. (2017). Network vector autoregression. *The Annals of Statistics*, *45*(3), 1096–1123. https://doi.org/10.1214/16-AOS1476

Zhu X., Wang W., Wang H., & Härdle W. K. (2019). Network quantile autoregression. *Journal of Econometrics*, *212*(1), 345–358. https://doi.org/10.1016/j.jeconom.2019.04.034