DOI: 10.1177/00222437241302102

JOURNAL of MARKETING RESEARCH

Identifying Competitors in Geographical Markets using the CSIS Method

Journal:	*Journal of Marketing Research*
Manuscript ID	JMR-22-0486.R5
Manuscript Type:	Revised Submission
Topics and Methods:	Statistics < Theoretical Foundation, Cross-sectional analysis < Methods, Time series < Methods, Competive analysis < Topics

Identifying Competitors in Geographical Markets Using the CSIS Method

Xian Gu
Assistant Professor in Marketing
Kelley School of Business
Indiana University, Bloomington
HH 2100, 1309 E. 10TH St., Bloomington, IN 47405
Tel: 812-856-1073
xiangu@iu.edu

P. K. Kannan
Dean's Chair in Marketing Science
Robert H. Smith School of Business
University of Maryland
3445 Van Munching Hall, College Park, MD 20742
Tel: 301-405-2188
pkannan@umd.edu

Identifying Competitors in Geographical Markets Using the CSIS Method

Abstract

Identifying the most relevant competitors in a geographical market is crucial for businesses with a significant offline presence, such as hotels, restaurants, and retail stores. The specific location of a business and the geographical density of potential competitors are critical factors in determining the competitive structure. However, this task can be challenging when the potential number of competitors is large and the competition is asymmetric. In this study, we apply the Conditional Sure Independence Screening (CSIS) method to a system of demand functions for competitor identification. This method offers significant computational efficiency by estimating a marginal regression for each potential competitor, rather than a full model consisting of all potential competitors. To validate the effectiveness of the CSIS method and explore the boundary conditions of its performance, we conduct extensive simulation analyses under different spatial data-generating processes. Our findings demonstrate that the CSIS method outperforms multiple other variable selection methods and remains robust under spatial misspecifications. Then we apply the CSIS method to hotel competition in two U.S. geographical regions, illustrating how the competitive structure varies across geographical densities and market segments. Finally, we highlight how managers can strategically use the results and outline the potential of the method for other non-geographical applications.

Keywords: competitor identification, CSIS, variable selection methods, demand model, spatial dependence

Introduction

For businesses with a significant offline presence, such as hotels, restaurants, and retail stores, identifying their most relevant competitors in local markets is crucial for effective competitive benchmarking and marketing mix decisions. However, this can be a difficult task. For example, in the hospitality industry, there could be a large number of hotel properties competing in a dense local geographical region, with some located close to one another. In a geographical market of this nature, it is challenging to determine the size of a focal hotel's competitive set and identify the specific competitors making up the set. Without such precise information, monitoring the activities of all other hotels becomes a potentially costly or impractical endeavor for hotel managers, especially in a dynamic environment where adjustments to pricing and promotions decisions are critical. Moreover, focusing on incorrect competitors can result in inaccurate pricing strategies and lower profits. These challenges also persist in sparse geographical markets where potential competitors may be widely dispersed. Questions about which properties scattered across the region constitute a competitive set, and the size of that set, become pertinent. Consequently, it is essential for a focal hotel to identify the set of the most relevant competing properties to effectively inform strategic and tactical decisions.

There is extensive research on identifying market structures based on the demand model and cross-price elasticities (see summaries by Elrod et al. 2002 and Shugan 2014). Yet estimating an entire demand model with more than a few dozen potential competitors can be computationally challenging (Smith, Rossi, and Allenby 2019). Thus, previous studies often leverage dimensionality reduction methods such as penalized likelihood (Li, Netessine, and Koulayev 2018) and machine learning (Gabel, Guhl, and Klapper 2019) approaches to investigate market structures when the number of competitors is large. Additionally, instead of using traditional demand and price data, some recent studies use new data sources such as online customer reviews (Lee and Bradlow 2011; Ye et al. 2022), user-generated

content (Netzer et al. 2012), online search records (Ringel and Skiera 2016), and social media activities (Yang, Zhang, and Kannan 2022). However, to accurately identify competitive market structures in a geographical context, it is essential to use a method robust to various types of potentially unobserved spatial structures. Previous studies have revealed that spatial factors can have a profound impact on both the supply and demand sides of a geographical market (Bronnenberg and Mahajan 2001; Duan and Mela 2009; Thomadsen 2005; Jank and Kannan 2005). Yet, the studies we have cited on competitor and market structure identification generally do not account for spatial factors, as they are not relevant to their specific applications. Applying these approaches directly to geographical contexts could therefore result in biased or incomplete conclusions.

This paper introduces the Conditional Sure Independence Screening (CSIS), a variable selection method proposed by Barut, Fan, and Verhasselt (2016), into the context of marketing and competitor identification. We use the CSIS method to analyze a system of demand functions, considering scenarios both with and without spatial dependence among businesses. The CSIS method computes a relevance score for each candidate variable (such as a competitor's price) by maximizing the marginal likelihood function for each candidate variable independently, conditional on other control variables known to influence the outcome variable (such as the price of the focal business, which is known to influence its demand). During this process, other candidate variables are excluded when computing each relevance score. Then the candidate variables are ranked by their relevance scores in descending order, and those exceeding a certain threshold are retained in the model. The threshold is determined by computing relevance scores from randomly permuted values for each candidate variable, similar to the importance scores in Random Forest or Gradient Boosting Machine (GBM) models. This threshold represents the maximum coefficients achievable under pure noise. Therefore, candidate variables must exceed this threshold to be considered relevant and retained in the model.

While other popular variable selection and dimensionality reduction methods, like

penalized likelihood methods (e.g., LASSO regression), typically seek to maximize the overall predictive power of the model while reducing the number of predictors, the CSIS method has a different objective. It focuses on individually assessing the strength of each candidate variable's relationship with the outcome variable. Furthermore, Barut, Fan, and Verhasselt (2016) has mathematically proved that the CSIS method has a desirable "sure screening" property. That is, as the number of observations approaches infinity, the probability of the CSIS method retaining all variables that are relevant to the outcome variable in the model converges to one. This advantage becomes particularly notable when dealing with highly correlated predictors. Unlike penalized likelihood methods that require trade-offs between predictors, the CSIS method avoids this need and is thus better able to deal with correlated predictors, which are common in many marketing applications.

To validate the effectiveness of the CSIS method, we conduct extensive simulation exercises, comparing them with several alternative methods such as LASSO regression and elastic net. In the simulations, we explore a diverse range of boundary conditions that could impact the method's performance. These include the type of spatial data-generating process, the length of time periods, the number of businesses in the region, the number of true competitors, the magnitude of price coefficients, the geographical density, and the market segment. Across most scenarios, the CSIS method consistently outperforms other methods. Notably, we demonstrate the effective performance of the CSIS method across data generated under diverse spatial dependence. Overall, this highlights the CSIS method's robustness against spatial misspecifications and its versatility in accommodating different spatial dependence structures among businesses.

Then we proceed to illustrate the usefulness of the CSIS method by applying it to price competition in the U.S. hotel industry. This investigation spans multiple geographical regions and encompasses two years from 2015 to 2016. The hotel industry provides an ideal context for our method as location and pricing are identified as the top two factors influencing consumers' booking decisions (STR 2021). Using our method, we not only identify numerous

nearby competitors that align with a focal hotel's brand tier but also pick out relevant competitors located at greater distances or associated with different brand tiers. Furthermore, we show that the market structure among hotel properties varies across different geographical densities and market segments (e.g., corporate guests versus loyalty redemption guests). In addition, we highlight distinctions in the competitive set between peak and off-peak seasons, noting that, on average, the competitive sets are larger during peak seasons.

Our study builds on the extensive literature on competitor identification and competitive market structure. In particular, we draw on previous studies that use demand and price data to identify competitors, such as research on antitrust and merger analysis (Conlon and Mortimer 2021; Farrell and Shapiro 2010; Gowrisankaran, Nevo, and Town 2015; Nevo 2000). We provide a comparative review of selected previous studies on competitor identification in Table 1, highlighting our contributions. Some studies have addressed the issue of large competitive sets through dimensionality-reduction methods (Gabel, Guhl, and Klapper 2019; Li, Netessine, and Koulayev 2018; Netzer et al. 2012), and other studies have accounted for the asymmetric competition (Kannan and Sanchez 1994; Ringel and Skiera 2016). Our research makes a significant contribution to the marketing literature by introducing the CSIS method to the field and highlighting its robustness to various specifications of spatial dependence. Furthermore, our method offers a compelling computational advantage because it maximizes a marginal likelihood function for each potential competitor independently instead of a full model including all competitors. This eliminates the high dimensionality problem that typically arises when dealing with a large number of competing alternatives, making our method computationally efficient and easy to apply. In addition, our method allows for modeling asymmetric competition, which is a critical aspect of many market contexts. Overall, our method provides a more accurate and comprehensive analysis of market competition in geographical contexts.

From a substantive perspective, our research generates important empirical findings that provide insights for offline businesses. In the context of our application, the proposed method

Table 1: Previous Research on Competitor Identification

Research	Empirical Context	Competitor Identification	Robust to Spatial Dependence	Large Number of Competitors	Asymmetric Competition
Kannan and Sanchez (1994)	Grocery	Subset selection	×	×	✓
Lee and Bradlow (2011)	Digital cameras	Text mining	×	×	×
Netzer et al. (2012)	Cars and drugs	Text mining	×	✓	×
Ringel and Skiera (2016)	Televisions	Mapping	×	✓	✓
Li, Netessine, and Koulayev (2018)	Hotels	Price functions with LASSO	×	✓	×
Gabel, Guhl, and Klapper (2019)	Grocery	Mapping	×	✓	×
Smith, Rossi, and Allenby (2019)	Grocery	Separable demand model	×	×	×
Our paper	Hotels	CSIS	✓	✓	✓

enables hotel managers to identify their price competitors for each market segment accurately. Our findings can be leveraged to develop effective pricing and revenue management strategies and gain a competitive edge in the marketplace. While other studies, such as Li, Netessine, and Koulayev (2018), have also examined hotel competition, our research stands out in several critical ways. First, the CSIS method performs well in handling various spatial dependencies among businesses, a factor not addressed in previous studies. Second, instead of focusing on price correlations between competitors, we base our analysis on the relationship between price and demand, which avoids issues that will be discussed in the next section. Lastly, our study examines hotel competition across multiple market segments, providing a more comprehensive and nuanced understanding of the competitive landscape in the hotel industry.

Institutional Setting and Data

Our study focuses on the hospitality industry in several geographical regions across the United States. To determine a hotel property's competitive set, our model considers the direct influence of competitor prices on the demand for the focal hotel property within a given geographical region. Benchmarking against competitive hotel properties in the local market is crucial for owners or franchisees of a hotel property, as it can aid in increasing overall occupancy rate and return on investment (ROI) through marketing mix decisions and identifying areas of improvement such as additional amenities (Wöber 2002). Among all these

decisions, pricing and promotion decisions tend to be the most important ones as they can be quickly adjusted in the short term to address competition. Furthermore, benchmarking information is an essential input for property owners or franchisees to negotiate with hotel brands that manage and operate the property.

To identify their competitors, property owners, and hotel managers frequently concentrate on several hotel attributes such as geographical proximity, brand tier and image, price, and hotel size, with geographical proximity often being the primary factor (Baum and Lant 2003; Lee 2015; Schwartz, Webb, and Ma 2021). For example, a hotel manager may simply consider other properties in similar geographical proximity and brand tiers as competitors, which is a frequently used heuristic (Mohammed, Guillet, and Law 2014). However, such a process of competitor identification used in the hotel industry has a subjective nature, which can result in misleading benchmarking results (Schwartz, Webb, and Ma 2021). For instance, determining how many miles should be regarded as too far for a hotel to be considered a competitor is subjective (Schwartz and Webb 2022).[1] Moreover, in practice, hotels may cater to multiple market segments, and their competitive sets could differ depending on the segment. For example, corporate travelers in a downtown area may prioritize a hotel property's proximity to a particular business venue, while tourists booking vacation packages may have a broader search area.

Furthermore, competitive sets may be asymmetric among a dyad of properties even within a market segment. While the price charged by Hotel A may influence demand for Hotel B, the price of Hotel B may not necessarily influence demand for Hotel A. This asymmetry could be due to various reasons, such as the relative capacities of the hotel properties. For example, assume hotels A and B have 500 and 50 guest rooms, respectively. In this scenario, if Hotel A lowers its prices, it is likely to decrease the demand for Hotel B. However, the

[1] To supplement our literature review, we interviewed revenue managers from two hotel chains – one large and one smaller. These interviews revealed that managers consider factors such as geographical proximity, brand tier, market segments, amenities, hotel size, unique property features (e.g., cultural and historical heritage), guest ratings, and meta-search engine categorization when determining competitive sets. These factors are considered informally based on data availability.

opposite effect may be less significant, as Hotel B, with its limited capacity, can only attract a small portion of Hotel A's demand by lowering its prices. Given this dynamic, if Hotel B views Hotel A as a competitor, it is likely to adjust its pricing strategy in response to Hotel A's prices, leading to a strong correlation between the two hotels' pricing strategies. If we solely rely on the price relationships between hotels (e.g., Li, Netessine, and Koulayev 2018), we may incorrectly identify Hotel B as a competitor of Hotel A. Therefore, our proposed method must account for asymmetric competition based on not only prices but also demand.

Data Overview

We obtained our data from a well-known data analytics firm with extensive expertise in the U.S. hospitality industry. The firm collects comprehensive data from hotel properties and offers subscription-based benchmarking data to them. Our data includes 1,313 hotel properties across five geographical regions in the United States: Washington, D.C.; Amarillo, Texas; New York City; Miami, Florida; and Stillwater, Oklahoma. The market boundaries were carefully determined by the data provider based on surveys from property managers and potential customers. Each geographical region includes a focal city, as indicated by the region name, as well as broad surrounding areas. Our data includes most hotels in the sampled markets that are in partnership with hotel groups such as Choice Hotels, InterContinental, Hilton, Marriott, and Wyndham.

Our data covers two years from January 1, 2015, to December 31, 2016. In the data, reservation records are at the hotel property, market segment, and check-in date levels, with each record containing aggregated information from multiple customer bookings. We observe detailed information about the guest-paid revenue, hotel-collected revenue, and number of room nights for each record. Moreover, for each hotel property, we observe its property name, brand name, brand tier (e.g., economy and upscale), number of guest rooms, and precise location information such as latitude, longitude, and property address.

Competition in Dense and Sparse Geographical Regions

Given that spatial factors may play an important role in analyzing competition among hotel properties, we are particularly interested in examining hotel competition in two distinct geographical regions: a dense metropolitan area and a sparsely populated area. The dense region, exemplified by Washington, D.C., is characterized by a high concentration of hotel properties within a small area. The primary challenge in this region is identifying a smaller set of the most relevant competitors among the large number of properties located close to the focal hotel. On the other hand, in sparse regions, such as Amarillo, Texas, hotels are distributed sparsely over a large area, presenting the challenge of identifying distant hotel properties that compete with the focal hotel property in addition to the ones close by.

Figure 1 depicts the geographical distribution of hotels in Washington, D.C., and Amarillo, TX. In both regions, the hotels tend to cluster in a few spots due to agglomeration. Table 2 provides a summary of key facts about the two regions and indicates that their spatial structures are vastly different. Although the sparse region (Amarillo) covers a significantly larger area than the dense region (Washington), both regions have a similar number of observed hotel properties (101 versus 127). The average distance between hotels in the Amarillo region is significantly greater than that in the Washington region, with a distance of 117.6 miles compared to 9.1 miles. The distances are computed based on the fastest driving routes, using the latitude and longitude of hotel locations and road data from R packages, dodgr, geodist, ggmap, and osmdata.

Furthermore, the composition of hotel brand tiers differs significantly between the two regions. In the Washington region, more than 50% of the branded hotels are in the upscale or upper upscale tiers, while most hotels in the Amarillo region are in the economy, midscale, or upper midscale tiers, with no luxury hotel. As a result, the average daily rate in the Washington region is approximately twice as high as that in the Amarillo region. Lastly, the supply of guest rooms is also substantially different between the two regions, with the Washington region providing 30,645 guest rooms and the Amarillo region only offering 7,601

Figure 1: Geographical Locations of Observed Hotels

(a) Washington, D.C. **(b) Amarillo, TX**

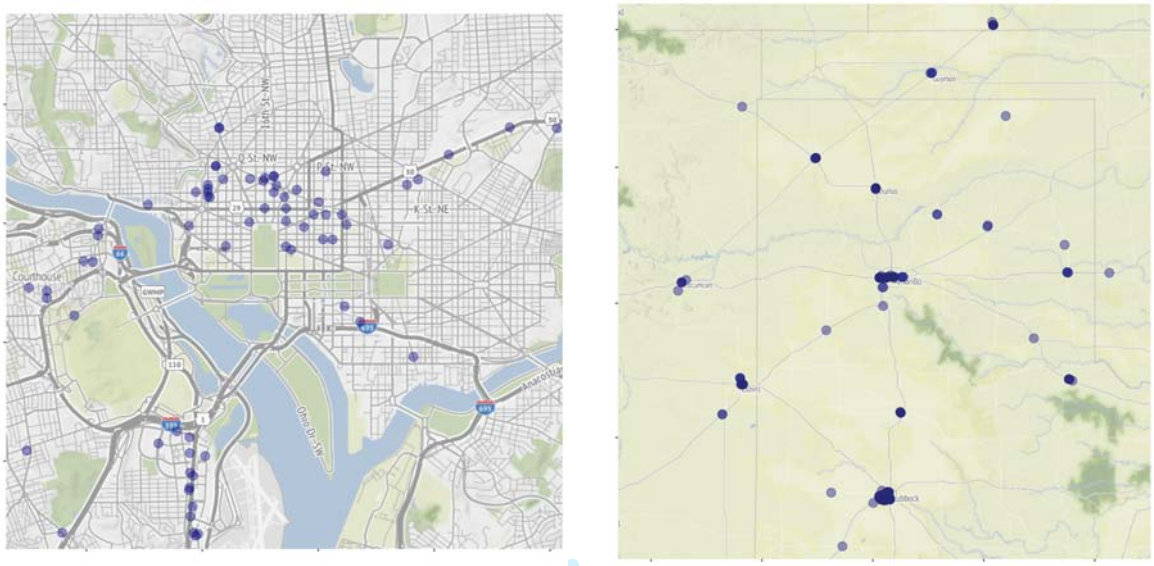


Table 2: Information about Two Geographical Regions

Information	**Dense Region** (Washington, D.C.)	**Sparse Region** (Amarillo, TX)
Region size (sq. miles)	259	58,521
Avg. driving distance between two hotels (miles)	9.1	117.6
Number of hotels	127	101
Economy	7.1%	32.7%
Midscale	3.1%	16.8%
Upper Midscale	20.5%	39.6%
Upscale	37.8%	9.9%
Upper Upscale	26.8%	1.0%
Luxury	4.7%	0%
Number of guest rooms	30,645	7,601
Avg. daily rate ($)	191.5	95.0

guest rooms. In this way, the Washington and Amarillo regions reflect the important features of dense and sparse regions in the hotel industry, providing valuable insights into the distinct market competition dynamics in these regions.

Hotel Market Segments

Another distinctive feature of our data is that it includes hotel reservations from all market segments on a daily basis. Referring to common practices in the hospitality industry (Madhok and Doherty 2020), we categorize hotel revenues into six distinct market segments based on the source of business and how business is negotiated: (1) transient (hotel-owned channels), (2) transient (third-party channels), (3) corporate, (4) group, (5) wholesale, and (6) employee/loyalty-redemption. The six market segments are non-overlapping in the sense that each reservation is associated with only one segment based on its booking channel.

We present a detailed description of each segment in Table 3, emphasizing the critical distinctions between them. We highlight that hotels' pricing strategies differ across segments, where some segments follow a negotiation model (e.g., wholesale), while others do not (e.g., transient). Also, some segments feature one-time business contracts for specific events (e.g., group), while others use long-term agreements (e.g., corporate). Moreover, we note that hotels' operating expenses and revenue structures are also different across market segments. For instance, while customers may pay the same room rate when booking through the hotel's website or a third-party website like Expedia, hotels' revenue streams vary as hotels incur commissions payable to third-party channels.

In Table 4, we also report the revenue share and average daily rate of each market segment in both regions. The transient (hotel-owned channels) and corporate segments are the two primary revenue drivers in both regions, with revenue shares exceeding 20%. The group segment is also a significant revenue source for the Washington region, accounting for a revenue share as high as 27.6%. Wholesale and employee/loyalty-redemption segments have a relatively low revenue share in both regions. The table also shows that the average daily

Table 3: Descriptions of Hotel Market Segments

No.	Segment	Description
1	Transient (hotel-owned-channels)	Individual guests who seek short hotel stays and book at publicly listed rates through hotel-owned booking channels such as the hotel's website, mobile app, call center, and property
2	Transient (third-party channels)	Individual guests who seek short hotel stays and book at publicly listed rates through third-party booking channels including consortia (i.e., groups of independent travel agencies) and online travel agencies (OTA)
3	Corporate	Guests who book at negotiated rates given to corporates, federal and state governments, and airline crews, which is usually based on annually-renewed contracts
4	Group	Large groups of guests who book a block of rooms at reduced rates for a specific event taking place in the hotel or nearby, such as conferences, weddings, and SMERF (social, military, education, religious and fraternal) meetings. A primary group contact typically coordinates with the hotel.
5	Wholesale	Guests who book at negotiated rates through wholesale accounts (e.g., tour operators and entertainment booking agents), in which a bulk of rooms are sold
6	Employee/loyalty-redemption	Guests associated with employees of the hotel or its affiliates and guests who use loyalty rewards programs to redeem for hotel stays

rates vary across segments, with the highest in the wholesale segment and the lowest in the employee/loyalty-redemption segment. Overall, our findings suggest that a narrow focus on a single segment may provide limited insights into the competitive landscape, highlighting the need for managers to adopt a broader perspective to create successful strategies that respond to the diverse needs of customers across various segments.

Table 4: Revenue Share and Average Daily Rate (ADR)

	Washington, D.C.			Amarillo, TX		
Segment	Revenue Share (%)	ADR (Mean $)	ADR (SD)	Revenue Share (%)	ADR (Mean $)	ADR (SD)
Transient (hotel-owned channels)	36.0	189.3	91.8	62.1	96.6	35
Transient (third-party channels)	9.3	187.6	95.5	9.7	104.2	36.5
Corporate	23.8	182.1	71.8	21.4	85.2	23.1
Group	27.6	175.5	76.2	3.1	99.4	30.9
Wholesale	1.7	202.8	89.3	1.2	119.6	36.6
Employee/loyalty-redemption	1.6	72.5	61.2	2.5	42.6	31.1

Hotel Sales and Prices

Sales and prices are the key variables in our model. Specifically, we measure hotel sales using the number of room nights sold for a specific check-in date and determine hotel prices for that date by calculating the average daily rate paid for those room nights. This measurement approach is widely used in the hotel industry research (Green and Lomanno 2016; Li, Netessine, and Koulayev 2018), as well as in research on the airline (Gerardi and Shapiro 2009; Granados, Gupta, and Kauffman 2012) and cruise industries (Joo, Gauri, and Wilbur 2020).

Figure 2 illustrates the time trends of the number of room nights and the average daily rate in the two geographical regions during our two-year data period. First, there is strong co-movement between the number of room nights and the average daily rate in both regions. This co-movement can be in the same or opposite direction, indicating the common demand shocks and price-setting behavior in the hotel industry. Second, the Washington region exhibits greater fluctuations in hotel sales and prices over time than the Amarillo region. Both regions also demonstrate seasonal patterns. In the Washington region, hotel sales and prices are higher in spring and lower in winter. Meanwhile, the Amarillo region shows relatively steady hotel sales and prices throughout the year, except for a decrease during the winter season. Consequently, we emphasize the importance of controlling for the common time trends using time-fixed effects in our model to obtain accurate results. Moreover, to capture the seasonal patterns, we compare the competitive sets during both peak and off-peak seasons.

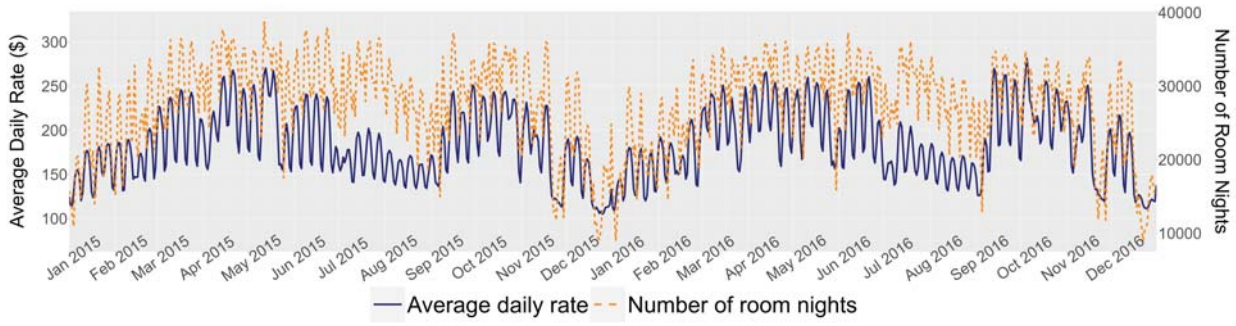
Model

The Demand Model

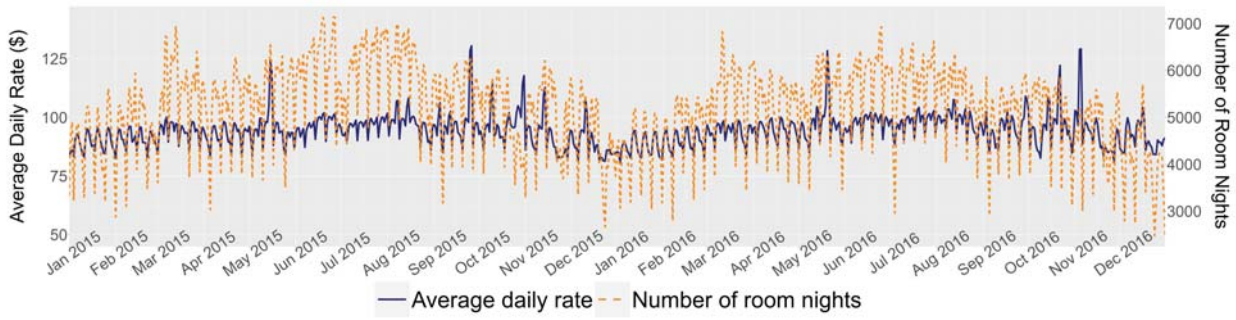
In line with well-established literature in economics and marketing (Bucklin and Srinivasan 1989; Russell 1992; Smith, Rossi, and Allenby 2019), we rely on the relationship

Figure 2: Dynamics of Average Daily Rate and Number of Room Nights

(a) Washington, D.C.



(b) Amarillo, TX



between price and demand to identify competitors. We consider a standard log-linear system of demand functions for N hotels within a geographical region. The log-linear functional form is linear in price coefficients and flexible for substitution patterns and thereby is widely used in marketing and economics research (DellaVigna and Gentzkow 2019).

$$\log q_{sit} = \alpha_{si} + \beta_{s,ii} \log p_{sit} + \sum_{j \neq i} \beta_{s,ij} \log p_{sjt} + z_t' \gamma_{si} + \epsilon_{sit} \tag{1}$$

$$s = 1, \ldots, S \quad \text{market segments},$$

$$i = 1, \ldots, N \quad \text{hotels},$$

$$t = 1, \ldots, T \quad \text{days},$$

where the demand q_{sit} is measured as the number of room nights sold by hotel i in segment s for check-in date t, and the price p_{sit} is the average daily rate charged by hotel i in segment

s for check-in date t. z_t is a set of time dummy variables for each calendar month, holiday, and weekend. It captures seasonality and other time-related factors that influence hotel demand. The error terms ϵ_{sit} capture idiosyncratic demand shocks for hotel i's segment s on check-in date t. The vector of error terms, ϵ_{st}, follows a multivariate normal distribution: $\epsilon_{st} = (\epsilon_{1st}, \ldots, \epsilon_{Nst}) \sim \mathcal{N}(0, \Sigma_s), \forall t$. We assume the demand shocks are correlated across hotels in the same segment and geographical region.

The term α_{si} is the hotel-and-segment-specific fixed effect, allowing us to capture any hotel and segment characteristics such as hotel brand, room capacity, geographical location, unique features, and facilities which could affect demand for hotel i in segment s. For example, large hotels in general have more group reservations than small hotels as they have enough room capacity to accommodate large guest groups.

Furthermore, our demand model considers the hotel-and-segment-specific coefficients, $\beta_{s,ij}$ and γ_{si}, to account for varying weight on prices and time-related factors across hotels and segments. For instance, hotels that target corporate guests may experience higher demand on weekdays compared to weekends. Conversely, hotels that cater to leisure travelers may observe greater demand on weekends.

Price Endogeneity

In modeling the demand functions, it is almost impossible to measure all the relevant factors that contribute to the demand. For instance, we do not observe local events such as conferences and concerts that could increase the demand for hotel stays during a specific period. However, we expect the hotel prices to reflect these unobserved demand shocks as hotel managers actively adjust hotel room rates based on their anticipation of upcoming events. The correlation between hotel prices and unobserved demand shocks can result in a price endogeneity problem, leading to biased and inconsistent price coefficients.

To tackle the issue of price endogeneity, we employ the classic Hausman instruments (Hausman 1996; Nevo 2001), which are hotel prices in other geographical markets. They

provide sufficient variation both over time and across hotels. Our exclusion restriction relies on a reasonable assumption that conditional on hotel and segment fixed effects (capturing time-invariant hotel and segment characteristics) and time fixed effects (capturing common time trends), prices across geographical markets are correlated due to cost shocks rather than demand shocks. For example, demand shocks such as events and conferences in one geographical region are unlikely to influence hotel demand in another.

In particular, we use hotel prices in the New York City and Miami, Florida regions as instruments for hotel prices in the Washington region, and the matched hotel prices are in the same segment. We do so based on the premise that these areas are all metropolitan regions with an abundance of hotels in densely populated areas. We expect that hotels in these regions face similar utility costs and labor markets, which result in correlated hotel prices due to the common cost shocks. Following this logic, we also use hotel prices in the Stillwater, Oklahoma region as instruments for hotel prices in the Amarillo region, as these areas share sparse hotel coverage.

Moreover, we match hotels by their brand names. For instance, we match a Hilton Garden Inn hotel in the Washington region with Hilton Garden Inn hotels in the New York City and Miami regions. In cases where brand name matches are not available, we match hotels of the same hotel group and brand tier (e.g., matching Hilton Garden Inn with Doubletree by Hilton). We posit that hotels within the same hotel group and brand tier share comparable costs for staff training and benefits, labor scheduling, and revenue management, which together comprise a substantial portion of hotel operation expenses. Additionally, hotels of the same hotel group would likely experience similar cost shocks in the event of financial distress for the hotel group.

We address the endogeneity concerns associated with both the own price variable ($\log p_{sit}$) and competitors' price variables ($\log p_{sjt}$). Taking $\log p_{sit}$ as an example, we first regress $\log p_{sit}$ on the hotel and segment fixed effects $\zeta_{si,0}$, the instrumental variables $\log p^m_{sit}$, and the time dummy variables z_t. Here, p^m_{sit} stands for the matched hotel prices. If a focal hotel i

has multiple matched hotels in segment s, p^m_{sit} will be a vector of prices of all matched hotels. From the regression, we obtain a predicted value of the focal hotel's price, $\widehat{\log p_{sit}}$, and we replace $\log p_{sit}$ with $\widehat{\log p_{sit}}$ in Equation 1.

$$\log p_{sit} = \zeta_{si,0} + \zeta_{si,1} \log p^m_{sit} + z'_t \zeta_{si,2} + e_{sit} \tag{2}$$

$$\log p_{sjt} = \zeta_{sj,0} + \zeta_{sj,1} \log p^m_{sjt} + z'_t \zeta_{sj,2} + e_{sjt}, \quad \forall j \neq i \tag{3}$$

We follow a similar procedure for competitors' price variables and replace $\log p_{sjt}$ with $\widehat{\log p_{sjt}}$ in Equation 1. In total, we estimate 794 first-stage regressions, one for each hotel in each segment and region. The partial-F tests indicate no presence of weak instruments for 97.5% of these regressions (Staiger and Stock 1997).

Competitor Identification

Our research aims to identify the most relevant competitors for each hotel based on their price coefficients. However, estimating the $N \times N$ parameter matrix for each market segment is computationally challenging when N is large, given that the number of parameters to be estimated in the price coefficient matrix, B_s, increases quadratically with N.

$$B_s = \begin{pmatrix} \beta_{s,11} & \beta_{s,12} & \cdots & \beta_{s,1N} \\ \beta_{s,21} & \beta_{s,22} & \cdots & \beta_{s,2N} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{s,N1} & \beta_{s,N2} & \cdots & \beta_{s,NN} \end{pmatrix}$$

It is reasonable to assume that the price coefficient matrices, B_s, are sparse and they contain many zero entries (Fan and Lv 2008). From the customers' perspective, it is unlikely that customers within a segment compare all hotels in a geographical region before deciding

where to stay, and as such, it is realistic to assume that most customers evaluate only a subset of the hotels. As a result, a hotel's demand is more likely to be influenced by a smaller subset of competitors. From the hotels' perspective, hotel managers find it costly and challenging to track all other hotels in a given region and segment and to adjust rates in reaction to their activity. Given that hotel pricing and availability are extremely volatile and changing regularly (Li, Netessine, and Koulayev 2018), it is essential that our study identifies a subset of the most relevant competitors in each segment that managers can focus on efficiently.

Conditional Sure Independence Screening (CSIS)

Instead of estimating the entire demand model as presented in Equation 1 and the full price coefficient matrices B_s, we conduct competitor identification using the conditional sure independence screening (CSIS) method developed by Barut, Fan, and Verhasselt (2016). It is a statistical method that identifies relevant variables in high-dimensional datasets. In modern applications such as genomics, neuroscience, and finance, datasets can contain thousands or even millions of variables, making it difficult to identify relevant variables for prediction or inference. CSIS addresses this challenge by assessing the strength of the relationship between potentially relevant variables and the outcome conditional on any other variables that are known to be relevant.

The intuition behind the CSIS method in our context is to measure how strongly a focal hotel i's demand is related to the prices of other hotels in the region. This is achieved by computing a relevance score for each hotel j's price, which indicates its contribution to hotel i's demand conditional on hotel i's own price, hotel-specific fixed effects, and time fixed effects. We then rank all competitive hotels in descending order of their relevance scores and select only those whose scores surpass a specific threshold. The relevance score of hotel j's price for hotel i's demand in segment s is given by the conditional maximum marginal likelihood estimator $\beta^*_{s,ij}$, which is defined below.

$$\{\alpha^*_{si}, \beta^*_{s,ii}, \beta^*_{s,ij}, \gamma^*_{si}\} = \underset{\alpha_{si};\beta_{s,ii};\beta_{s,ij};\gamma_{si}}{\arg\max} \left\{ \sum_{t=1}^{T} l\left(\alpha_{si} + \beta_{s,ii}\widehat{\log p}_{sit} + \beta_{s,ij}\widehat{\log p}_{sjt} + z_t'\gamma_{si}; \log q_{sit}\right) \right\}, \quad (4)$$

$$l(\cdot) = -\frac{1}{2}\ln\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\left[\log q_{sit} - \left(\alpha_{si} + \beta_{s,ii}\widehat{\log p}_{sit} + \beta_{s,ij}\widehat{\log p}_{sjt} + z_t'\gamma_{si}\right)\right]^2$$

where $l(\cdot)$ is the log value of the probability density function of a normal distribution, and σ^2 is the variance of the normal distribution. According to Barut, Fan, and Verhasselt (2016), the marginal coefficient in CSIS differs from the full regression parameter so that when the full regression parameter exceeds a certain threshold, the marginal coefficient exceeds another threshold, and thereby the marginal coefficients can provide useful probes for variable selection.

The marginal log-likelihood function above is estimated independently for each hotel j, market segment s, and region. Therefore, the number of functions to be estimated increases linearly with the number of potential competitors while the number of parameters in each function does not increase. This procedure offers exceptional computational efficiency, making it ideal for managing a large number of competitors.

Determining Threshold

Instead of using arbitrary values, the CSIS method determines the threshold through the creation of a null model. For each competitor j in segment s, we first obtain a vector of its fitted values of prices, $\widehat{\log p}_{sj}$, from Equation 3. Next, we randomly permute the elements in this vector and generate a new price vector $\widetilde{\log p}_{js}$. Then we obtain a marginal estimator $\check{\beta}_{s,ij}$ by applying the model in Equation (4) to the permuted data $\left\{\log q_{sit}, \widehat{\log p}_{sit}, \left\{\widetilde{\log p}_{sjt}\right\}_{j\neq i}, z_t\right\}_{t=1}^{T}$.

Since the vectors $\widetilde{\log p}_{js}$ and $\log q_{si}$ are decoupled, the marginal estimator $\check{\beta}_{s,ij}$ measures a noise level under the null model. To stabilize the threshold value, we conduct the permutation

for $K = 10$ times, resulting in the following values:

$$\left\{\left|\breve{\beta}_{s,ij}^{(k)}\right|,\ \forall j \neq i\right\}_{k=1}^{K}, \tag{5}$$

We define the threshold for $|\beta_{s,ij}^{*}|$ as $\delta_{si} = \kappa\lambda_{si,\tau}$, where $\lambda_{si,\tau}$ is the $\tau-$quantile of the above values. We choose $\tau = .99$ as it is more stable than the maximum value (Barut, Fan, and Verhasselt 2016).

We also offer managers an extra layer of flexibility in managing the size of their competitive sets by allowing for a scaling parameter, $\kappa \in [1, +\infty)$, on the threshold value. In our empirical applications, we use a five-fold cross-validation to determine $\kappa = 8$ and $\kappa = 5$ for the Washington and Amarillo regions, respectively. Hotel managers can begin with the kappa value derived from cross-validation. For those equipped with the resources and capabilities to track a larger number of competitors, selecting a κ smaller than that estimated by cross-validation might be advantageous. If hotel managers seek a comprehensive view of their competitive landscape, they might set κ to 1. Conversely, for managers focusing on a few competitors, it is recommended to adjust the κ value upwards until the desired size of the competitive set is achieved. This strategic approach ensures that our method aligns seamlessly with varying managerial needs and objectives.

Thus, the set of identified relevant competitors for hotel i in segment s, J_{si}^{c}, is given by:

$$J_{si}^{c} = \left\{j \neq i : \left\{\left|\beta_{s,ij}^{*}\right| \geqslant \delta_{si}\right\}\right\} \tag{6}$$

We also discuss incorporating a spatial specification into the demand model under the CSIS method in Web Appendix B.

Advantages of CSIS

We adopt the CSIS method for three primary reasons and summarize our comparison between the CSIS method and the penalized likelihood methods in Table W1. First, the

model objectives of the CSIS method and the penalized likelihood methods are different. The penalized likelihood methods are primarily designed to maintain the overall predictive power of the model while reducing the number of predictors. Specifically, the objective function of a penalized likelihood method is to minimize the model's root mean square error (RMSE) with a penalty function, $R(\beta_{ij})$, which depends on coefficient magnitudes (Tibshirani 1996):

$$\min_{\alpha_{si},\beta_{s,ii},\beta_{s,ij},\gamma_{si}} \left\{ \sum_{t=1}^{T} \left(\log q_{sit} - \alpha_{si} - \beta_{s,ii}\widehat{\log p}_{sit} - \sum_{j\neq i} \beta_{s,ij}\widehat{\log p}_{sjt} - z_t'\gamma_{si} \right)^2 + \eta \cdot R(\beta_{ij}) \right\}, \forall s, i \quad (7)$$

For example, the penalty functions for LASSO regression and ridge regression are given by $R(\beta_{ij}) = \sum_{j\neq i} |\beta_{ij}|$ and $R(\beta_{ij}) = \sum_{j\neq i} \beta_{ij}^2$, respectively. However, in our research context, retaining two competing hotels with small price coefficients might improve the predictive power of the model compared to only keeping one competing hotel with a large price coefficient. Therefore, the penalized likelihood methods may overlook important competitors and may not be suitable for our research purpose. In contrast, the CSIS method does not have this problem since it retains variables with the highest marginal estimator without worrying about the predictive power of the full model.

Second, the CSIS method has an advantage over the penalized likelihood methods in dealing with highly correlated predictors, such as the price variables in our context. When the predictors are highly correlated, it is difficult to obtain reliable estimates of β_{ij} in the penalized likelihood methods as they are estimated jointly. For instance, the penalty function for LASSO regression, $R(\beta_{ij}) = \sum_{j\neq i} |\beta_{ij}|$, makes it tend to pick one variable to keep (rather arbitrarily) and drop other correlated variables, potentially excluding important competitors. The above issue can be amplified when the dataset contains spatial or temporal dependency in the predictors, and this could be another factor contributing to the poor performance of the penalized likelihood methods in our context. However, the CSIS method estimates the conditional marginal likelihood estimator for each predictor separately as in Equation 4, and only focuses on the relationship between each predictor and the outcome variable. So, even if predictors are highly correlated, the CSIS method does not require them to be traded off. Previous literature on sure independence screening (e.g., Simulation II in Fan and Lv 2008) and our simulation analyses support these statements.

Finally, Barut, Fan, and Verhasselt (2016) has rigorously proved that CSIS has a desirable

"sure screening" property under reasonable regularity conditions. The sure screening property is that, as the number of observations tends towards infinity (in our case, as $T \to \infty$), the probability of the CSIS method retaining all variables that are relevant to the outcome variable in the model converges to one. The number of observations utilized by Barut, Fan, and Verhasselt (2016) ranges from 100 to 500, and our number of observations $T = 731$ is adequate. According to their proof, the regularity conditions (Conditions 1 and 2 in Barut, Fan, and Verhasselt 2016) are satisfied in our analysis because we use a linear model and normalize the price variables for variable selection. Previous literature has also proved that the penalized likelihood methods can recover the correct model under certain assumptions. However, many of these results rely on the condition that the variables with non-zero coefficients are not highly correlated with those whose coefficients shrink to zero (Hastie et al. 2009). In our context, where price variables exhibit high correlations, this condition is unlikely to hold.

Validation by Simulation Analysis

Given that the true market structures are unknown and there is no ground truth against which to evaluate our CSIS method, we leverage simulations to gauge the performance of our methods and to compare them with alternative variable selection methods in recovering true market structures.

Data Simulation

Table 5 summarizes the simulation conditions used in our analysis. These conditions are crucial to ensure that our simulated datasets cover a broad range of specifications and are representative of different real-world situations. By simulating a large number of datasets with known market structures, we can systematically evaluate their performance across diverse market structures and parameter values. Importantly, we use both randomly generated and actual price variables and these two types of price variables allow us to investigate how our method performs in comparison with other methods under low and high correlations among price variables. While there is little correlation among randomly generated prices, the median correlation among actual prices is .47.

Table 5: Data Simulation Conditions

Simulation Conditions	Random Prices	Actual Prices
1. Data generating process	Non-spatial, spatial drift, spatial error, and spatial lag	
2. Length of time periods	1 month, 6 months, 1 year, and 2 years	
3. Number of hotels in the region	20, 100, and 200 (random locations)	5, 20, and 100 (actual locations)
4. Number of true competitors	between 1 and 5, between 15 and 20	
5. Average magnitude of price coefficients	0.5 and 1.5	
6. Geographical density	Washington (dense) and Amarillo (sparse)	
7. Market segments	–	6 segments
Number of simulated datasets	**384 datasets**	**1,536 datasets**

Random prices

We simulated 384 datasets using randomly generated prices according to the following steps:

1. We select one of the four data-generating processes outlined below, which include not only the non-spatial scenario but also various types of spatial dependence (Bradlow et al. 2005).

$$\text{Non-spatial:} \quad \log q_{sit} = \alpha_{si} + \beta_{s,ii} \log p_{sit} + \sum_{j \neq i} \beta_{s,ij} \log p_{sjt} + z_t' \gamma_{si} + \epsilon_{sit} \tag{8a}$$

$$\text{Spatial drift:} \quad \log q_{sit} = \alpha_{si} + \beta_{s,ii} \log p_{sit} + \sum_{j \neq i} \left(\tilde{\beta}_{s,ij} + \rho_{si} w_{s,ij}\right) \log p_{sjt} + z_t' \gamma_{si} + \epsilon_{sit} \tag{8b}$$

$$\text{Spatial error:} \quad \log q_{sit} = \alpha_{si} + \beta_{s,ii} \log p_{sit} + \sum_{j \neq i} \beta_{s,ij} \log p_{sjt} + z_t' \gamma_{si} + u_{sit} \tag{8c}$$

$$u_{sit} = \rho_{si} \sum_{j \neq i} w_{s,ij} u_{sjt} + \epsilon_{sit}$$

$$\text{Spatial lag:} \quad \log q_{sit} = \rho_{si} \sum_{j \neq i} w_{s,ij} \log q_{sjt} + \alpha_{si} + \beta_{s,ii} \log p_{sit} + \sum_{j \neq i} \beta_{s,ij} \log p_{sjt} + z_t' \gamma_{si} + \epsilon_{sit} \tag{8d}$$

where ρ_{si} is a scaling parameter, and $w_{s,ij} = \exp\left(-d_{ij}^2 / 2h_{si}^2\right)$ is a Gaussian kernel spatial weighting function, with d_{ij} being the driving distance between hotel i and j, and h_{si} being the kernel bandwidth.

2. We select a time window (i.e., data size), which could be one month, six months, a year, or

two years.

3. We determine the geographical boundaries of all observed hotels in the Washington or Amarillo region. Within these boundaries, we randomly generate 20, 100, or 200 pairs of longitudes and latitudes, which we use as the locations of simulated hotels in a given region. This approach of using randomly generated locations, rather than actual ones, allows us to explore broader boundary conditions with more hotels than we observe in the data.

4. For each hotel, we randomly select a number of true competitors, which can be between 1 and 5, or between 15 and 20.

5. We randomly generate the price coefficients, setting their average magnitude at either .5 or 1.5. We also randomly generate other model parameters in Equation 8. For instance, the spatial weighting scaling parameter, ρ, is drawn from a normal distribution with a mean of .1 and a standard deviation of .05.

6. We randomly generate price variables based on the average prices in our data and simulate the sales variables according to Equation 8. Only the information about true competitors selected in Step 4, such as their prices, demand, and weights, is used in Equation 8. We apply the above steps to both Washington and Amarillo regions.

Actual prices

We simulated 1,536 datasets using actual hotel prices. The process was identical to the one used for random prices, with a few exceptions noted below.

3. We randomly select 5, 20, or 100 hotels from all observed hotels within each segment and region. However, if the total number of hotels in a specific segment and region is lower than the required number for the simulation, we omit that simulation condition for that particular segment and region. For instance, in the Washington region, there are only 27 hotels in the wholesale segment, so we omit the simulation condition of 100 hotels in this case. Also, unlike the random-price simulations, we do not use 200 hotels as a simulation condition here because no segment has such a large number of hotels.

6-7. We use actual prices from the corresponding market and segment to simulate the sales variables according to Equation 8. Only the information about the true competitors selected in Step 4 is used in Equation 8. We apply the above steps to all six segments in both the Washington and Amarillo regions.

Methods for Competitor Identification

We apply our proposed CSIS method to the simulated datasets and compare the resultant market structures with the true market structures. We set the scaling parameter of the CSIS threshold to $\kappa = 1$ (i.e., no scaling) in the simulation analysis to ensure fair comparisons with other methods.

Moreover, we compare the results of our CSIS method with those of alternative variable selection methods to further validate our method. First, we use two popular penalized likelihood methods, LASSO regression, and elastic net, to identify the most relevant competitors. LASSO regression, introduced by Tibshirani (1996), is widely recognized as the gold standard for variable selection (Bhadra et al. 2019). The elastic net, a hybrid of ridge regression and LASSO regression proposed by Zou and Hastie (2005), is also a popular technique for variable selection. We select the optimal parameters for LASSO and elastic net using a five-fold cross-validation, and the identified competitors correspond to the minimum cross-validated mean squared errors.

We also adopt two more recent methods for variable selection. One method is the Graphical Gaussian Model (GGM) (Yuan and Lin 2007), which is a probabilistic graphical model designed to capture conditional independence relationships between a set of Gaussian random variables (e.g., hotel demand and price variables). The other method is the Sequence LASSO, which is an extension based on the LASSO regression. In particular, Chernozhukov et al. (2021) introduced a sequential estimation procedure to estimate a system of high-dimensional sparse regressions and to determine the appropriate penalty level for LASSO by controlling the aggregated errors within the system.

Finally, we evaluate the effectiveness of our method in comparison to industry practice, which typically identifies competitors based on geographical proximity and brand tiers. We identify a focal hotel's competitors as hotels that have the same brand tier and are located within 5 miles in the Washington region or 100 miles in the Amarillo region as competitors. However, in the simulation

study using random hotel prices, we do not consider brand tiers as these hotels are not associated with real brands and thus do not have brand tiers.

Simulation Results

The main findings of our simulation analysis are visually presented in Figures 3-5 and further detailed in Tables W3-W4. To evaluate the performance of our proposed CSIS method against alternative methods, we adopt three widely used metrics: area under the ROC curve (AUC), balanced accuracy, and F1-score. These metrics score between 0% and 100%, with higher values indicating superior performance. Overall, our simulation results highlight the superior performance of the CSIS method when compared to alternative methods such as LASSO, elastic net, sequence LASSO, GMM, and the proximity-and-brand-based rule. This performance advantage holds across all performance metrics as shown in Figure 3.

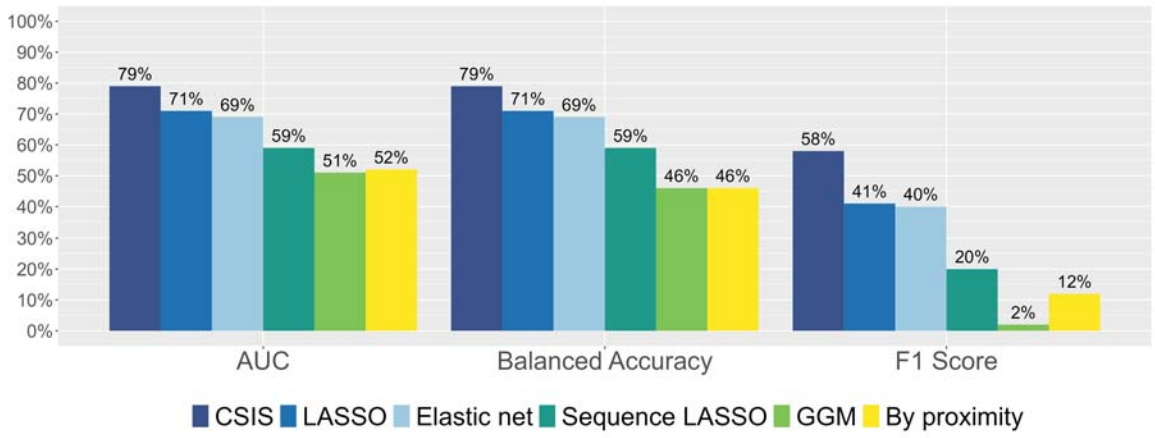
Our results also reveal valuable insights into the boundary conditions that influence the performance of our method. Figures 4(a) and 5(a) show that the number of true competitors has a limited impact on method performance when actual prices are used for simulations. However, for random prices, all methods tend to perform better when the competitive set is smaller. Figures 4(b) and 5(b) indicate that the CSIS method consistently outperforms other methods across different data-generating processes. This finding suggests that the CSIS method is robust to spatial misspecification. In Figures 4(c) and 5(c), the performance of the CSIS method notably improves as the time window becomes longer. Specifically, the CSIS method exhibits superior performance in longer time windows but may be surpassed by other methods when the time window is limited to 30 days.[2] Figures 4(d) and 5(d) indicate that the performance of all methods improves as the magnitude of price coefficients increases. In Figures 4(e) and 5(e), while the CSIS method demonstrates a significant advantage over other methods when dealing with a small number of hotels, their performance gap narrows as the number of hotels in a market increases. Finally, the CSIS method consistently outperforms other methods across regions and segments as shown in Figures 4(f) and 5(f)-5(g).

Finally, we explore the performance of the CSIS method in situations where a manager has

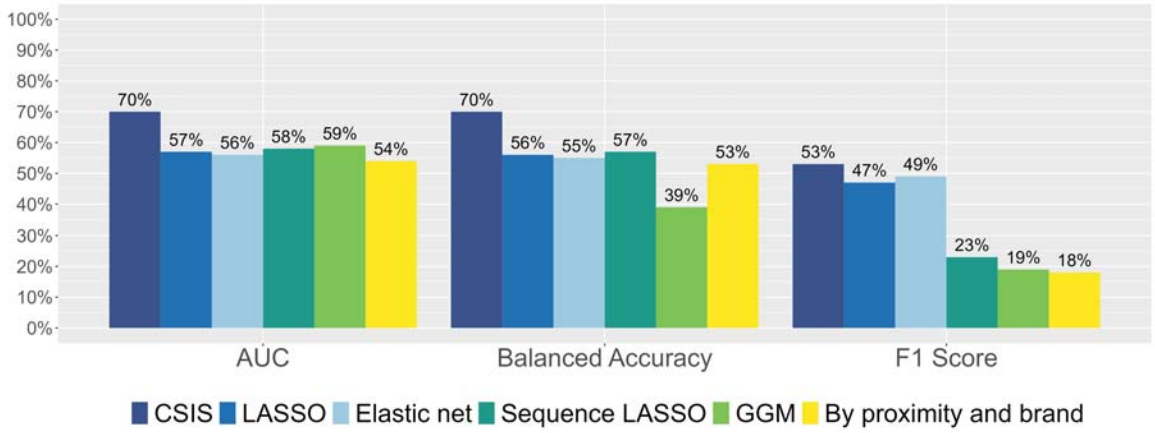
[2] Given the impractical shortness of 30 days, we include these simulation datasets when reporting the overall effects while excluding them when reporting performance under other boundary conditions.

Figure 3: Overall Simulation Results

(a) Random Prices



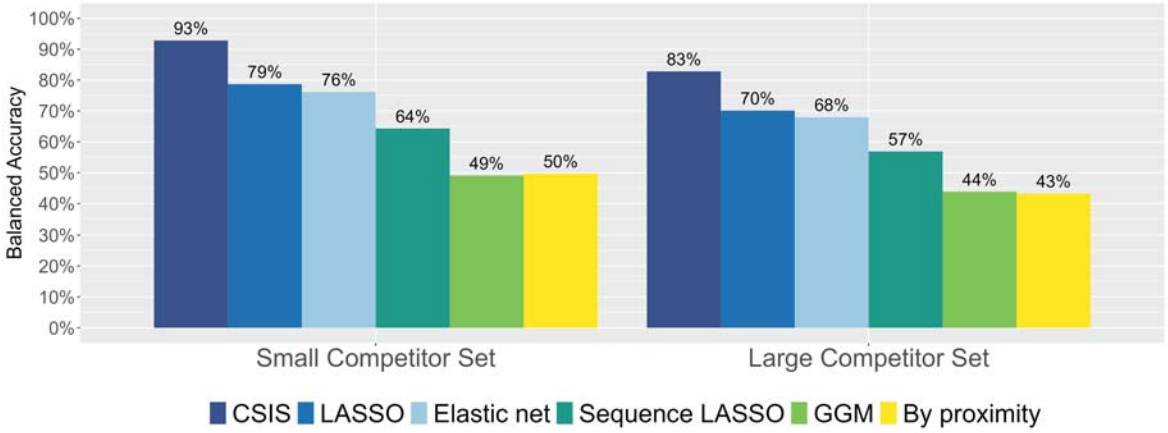
(b) Actual Prices



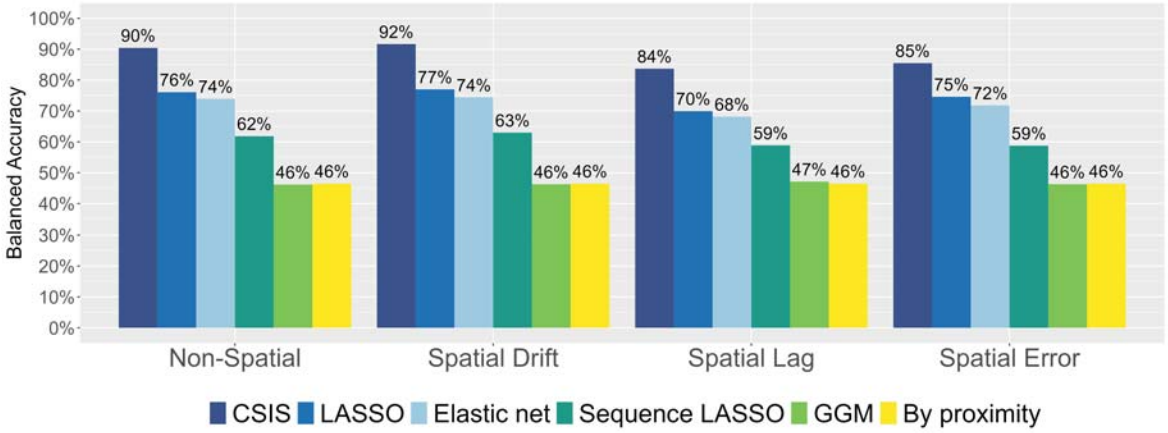
prior knowledge about the competitive set. Specifically, we considered two scenarios where we applied the CSIS method to our 1,536 simulated datasets with actual prices. We consider scenarios where the manager of a focal hotel believes Hotel A is a competitor. In the first scenario, this prior knowledge is correct and the price of Hotel A does influence the simulated demand for the focal hotel. In the second scenario, this knowledge is incorrect and the price of Hotel A does not influence the simulated demand for the focal hotel. When applying the CSIS method, we consistently include the price variable of Hotel A in our model, regardless of whether the manager's prior knowledge is correct. Our simulation results in Table 6 indicate that the CSIS method with correct prior knowledge outperforms the others, with the CSIS method without prior knowledge following behind. On the other hand, the CSIS method with incorrect prior knowledge performs the least accurately.

Figure 4: Simulation Results by Boundary Conditions (Random Prices)

(a) Number of True Competitors



(b) Data Generating Processes



(c) Length of Time Periods

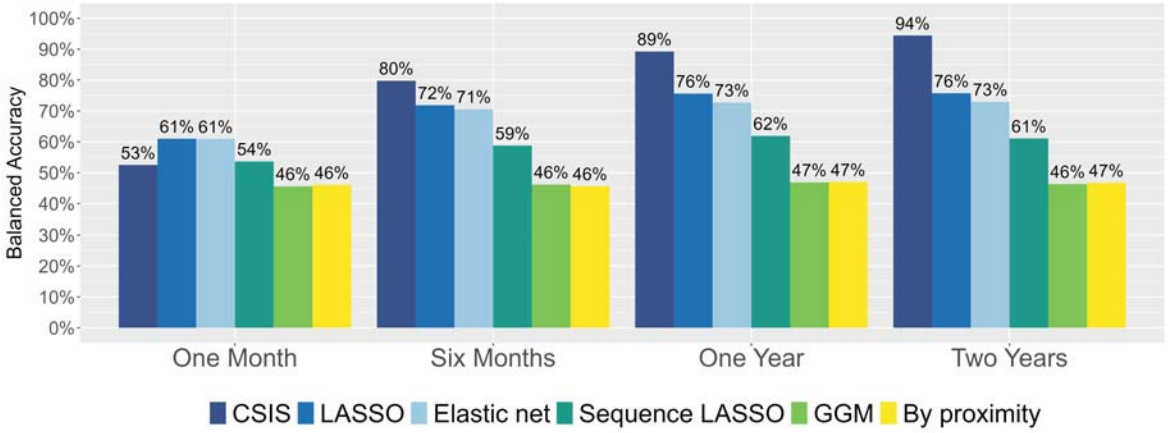
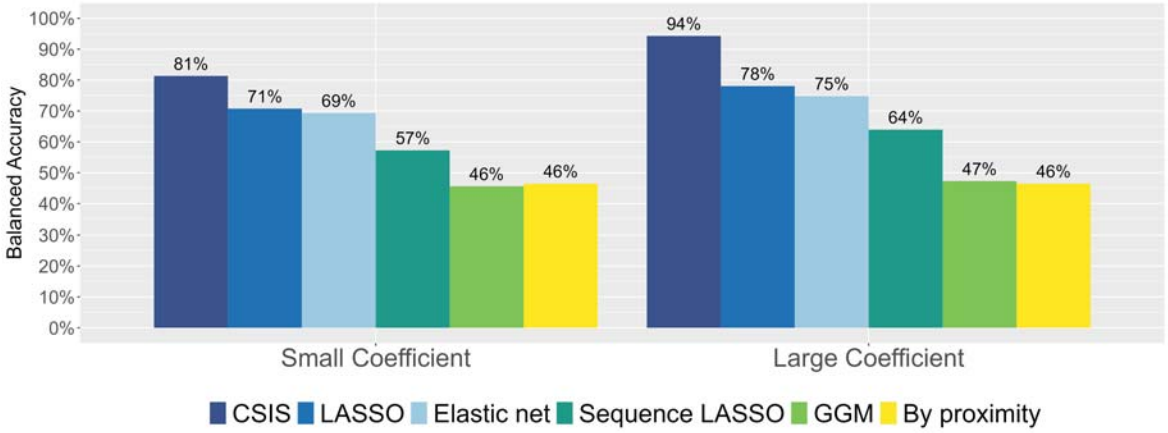
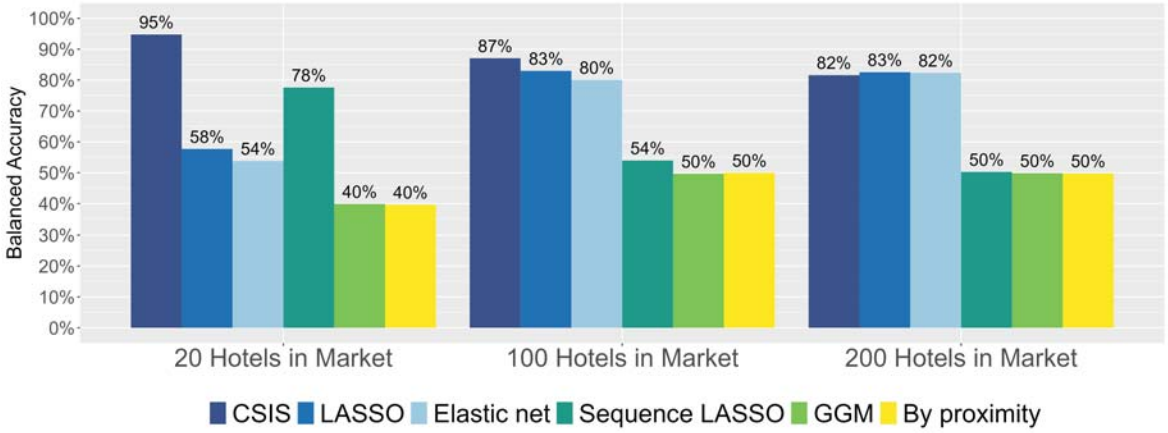


Figure 4: Simulation Results by Boundary Conditions (Random Prices)

(d) Average Magnitude of Price Coefficients



(e) Number of Hotels in the Region



(f) Geographical Regions

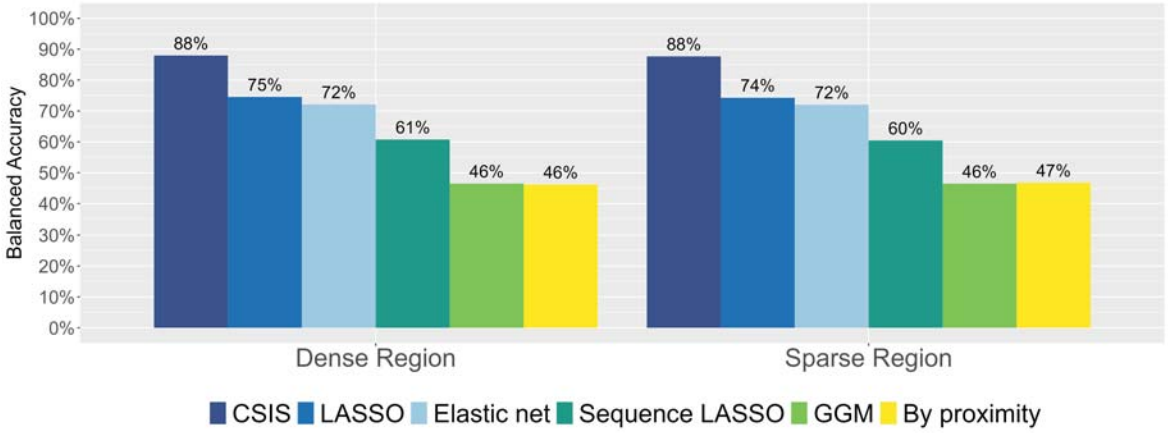
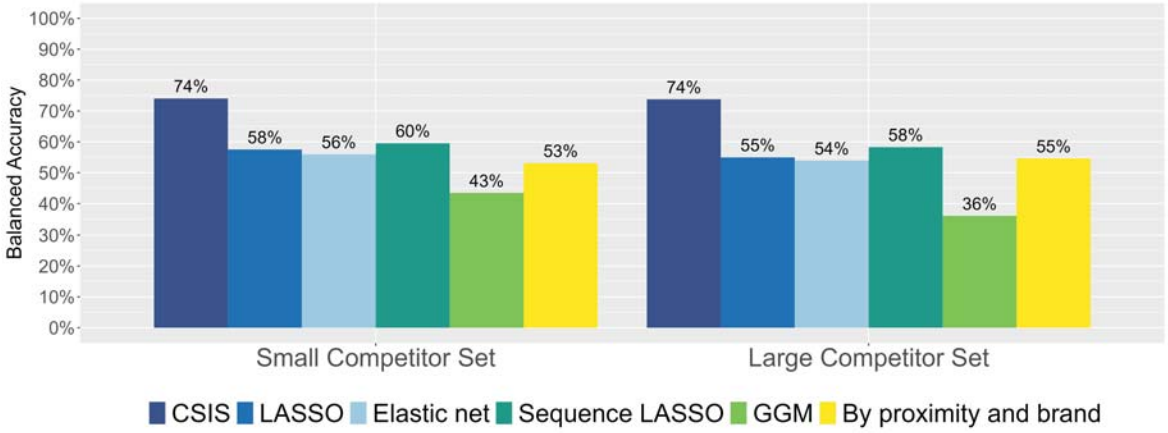
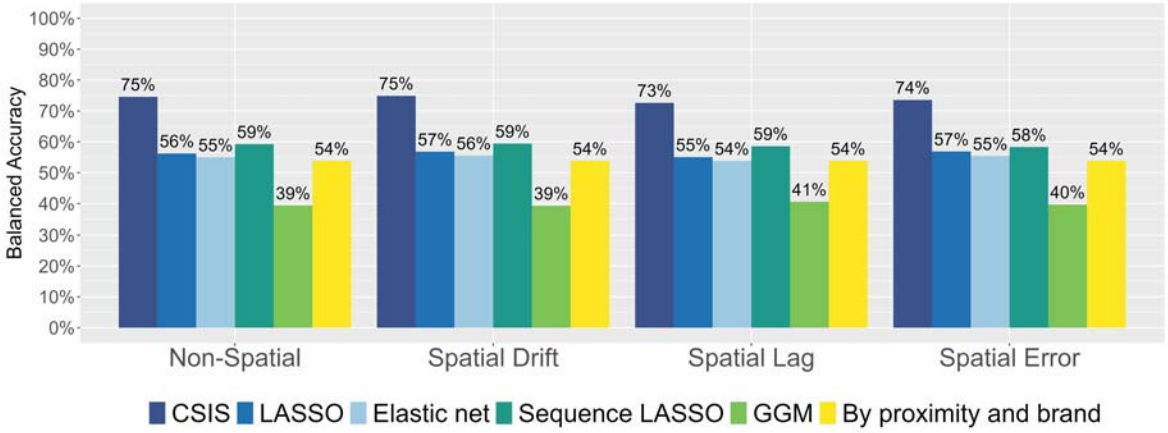


Figure 5: Simulation Results by Boundary Conditions (Actual Prices)

(a) Number of True Competitors



(b) Data Generating Processes



(c) Length of Time Periods

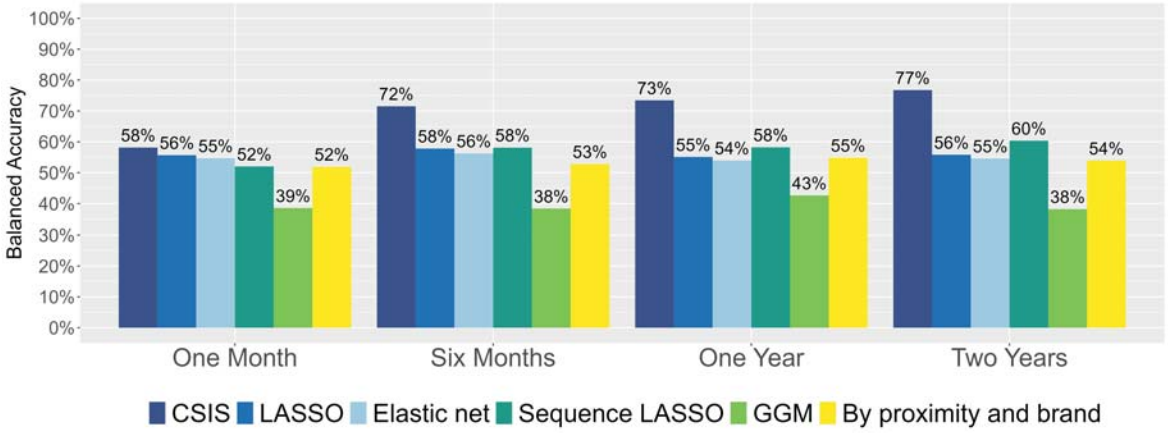
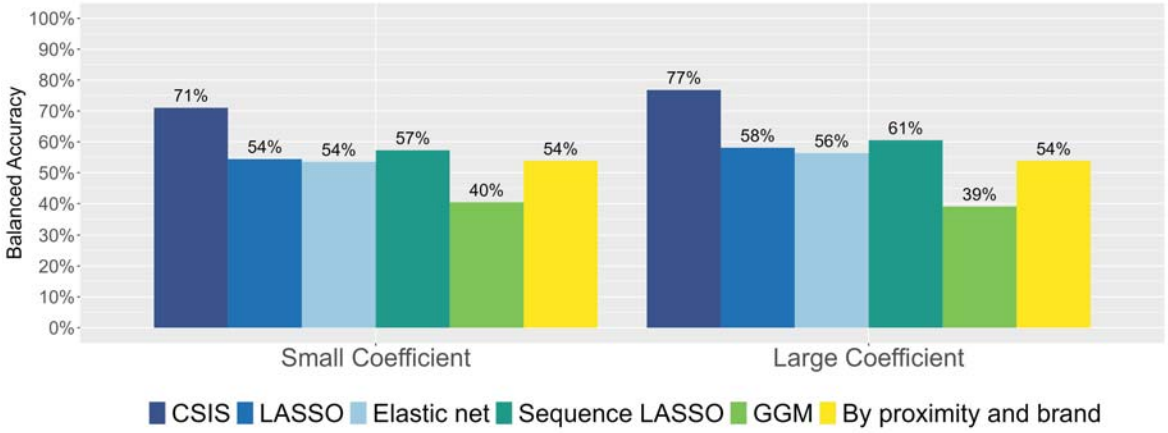


Figure 5: Simulation Results by Boundary Conditions (Actual Prices)

(d) Average Magnitude of Price Coefficients



(e) Number of Hotels in the Region

(f) Geographical Regions

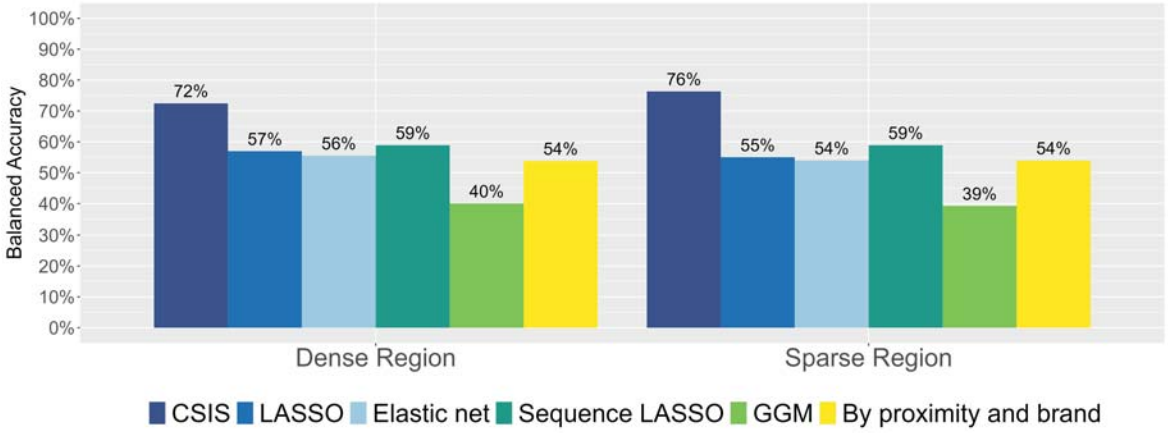
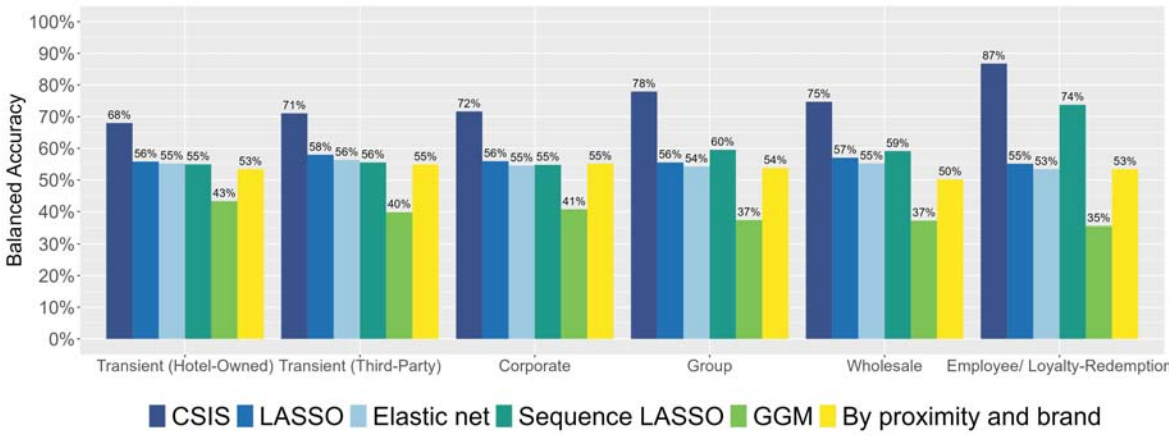


Figure 5: Simulation Results by Boundary Conditions (Actual Prices)

(g) Market Segments



Hence, our findings suggest that incorporating prior knowledge can be advantageous when the knowledge is correct, but it may lead to a disadvantage when the knowledge is incorrect.

Table 6: Simulation Results with Managers' Prior Knowledge

Method	AUC (%)	Balanced Accuracy (%)	F1-Score (%)
CSIS (without prior knowledge)	70.3	70.0	52.6
CSIS (with correct prior knowledge)	78.5	78.3	65.6
CSIS (with incorrect prior knowledge)	67.0	65.3	49.1

Overall, our simulation results provide robust evidence in support of the effectiveness of the CSIS method in identifying true competitors and market structures.

Empirical Application

In this section, we present an empirical application of the CSIS method and summarize the identified competitive sets across six market segments in the Washington and Amarillo regions.

Competitive Sets

Table 7 reports the average size of identified competitive sets the hotels have. The results reveal a significant level of variability, with the average number of competitors ranging from .2 to

45.1 depending on the market and the segment. The segment-level results also demonstrate a high level of face validity. For example, the employee/loyalty-redemption segment has a small number of competitors in both regions. This could be due to that redemption and employee bookings may often be focused on specific hotel properties and chains. Similarly, smaller segments such as wholesale also tend to have smaller competitive sets since wholesale booking agents may focus on a few properties that they have relationships with.

Table 7: Summary Statistics of Competitive Set Sizes

Segment	Total Number of Hotels	Size of Competitive Sets						
		Mean	SD	5%	25%	50%	75%	95%
Washington, D.C.								
Transient (hotel-owned channels)	125	12.3	18.7	.2	2	5	11	57.8
Transient (third-party channels)	115	3.5	4.6	0	1	2	4	15.3
Corporate	120	45.1	23.6	7.9	25.8	45	64	80.1
Group	65	11.2	8.7	.2	5	10	15	26.6
Wholesale	27	2.1	2.2	0	.5	1	3.5	5.7
Employee/loyalty-redemption	94	.2	.5	0	0	0	0	1
Amarillo, TX								
Transient (hotel-owned channels)	101	17	12.2	3	6	15	27	37
Transient (third-party channels)	42	3.9	3.8	0	1	3	5	10.9
Corporate	69	11.5	8.4	1.4	6	10	15	27.4
Group	2	.5	.7	.1	.2	.5	.8	.9
Wholesale	3	2	0	2	2	2	2	2
Employee/loyalty-redemption	31	4.1	2.2	2	3	4	5	8

Interestingly, the competitive set sizes for individual guests booking through hotel-owned channels are around 12 and 17 on average, while the competitive set sizes for individual bookings through third-party channels are smaller, with a mean of 3.5 in the Washington region and 3.9 in the Amarillo region. This suggests that, for a focal hotel, third-party channels tend to concentrate price competition among fewer competitors than hotel-owned channels, where the competition is more diffuse. In addition, hotel properties targeting the corporate segment in the Washington region tend to have large competitive sets, indicating the fierce competition in the corporate segment given the urban location. In contrast, the Amarillo region has a much smaller average competitive set size in the corporate segment, indicating that competition in this region is less intense.

Table 8 presents an overview of the percentage of common competitors in different segments in the Washington and Amarillo regions. The results of our analysis indicate that merely 35% and 27% of competitors in the Washington and Amarillo regions, respectively, are present in common across market segments. This emphasizes the importance for hotel managers to gain a comprehensive understanding of their competitive landscape in each market segment.

Table 8: Percentage of Common Competitors Across Market Segments

Common Competitors	Washington, D.C.	Amarillo, TX
Across 2 Segments	29.04%	24.86%
Across 3 Segments	5.29%	1.71%
Across 4 Segments	.78%	0%
Across 5 Segments	.06%	0%
Across 6 Segments	0%	0%
Total	35.17%	26.57%

Finally, we visualize the competition matrices for all segments and regions in Figure W2. The colored cell in column i and row j represents hotel j is a competitor for a focal hotel i. Thus, column i represents all identified competitors for a focal hotel i. A blank cell suggests that there is no identified competitive relationship. Notably, the competition matrices reveal two critical patterns. First, competition among hotels is almost always asymmetric, with clear asymmetry observed about the diagonals of the matrices. This finding highlights the need to incorporate competitive asymmetry in models to better comprehend market competition. Second, the intensity of competition varies significantly across not only market segments and geographical regions but also individual hotels. Even within the same segment and region, certain hotels face a higher number of competitors while others compete with only a few or none at all. These observations underscore the complex nature of hotel competition and highlight the importance of considering varying levels of competition in any analysis of market competition.

Characteristics of Identified Competitors

To gain a better understanding of the identified competitors, we summarize the distance between each focal hotel and its identified competitors, as well as their brand tier, in Table 9. Our

analysis of the Washington region reveals that for all segments, 50% of the competitors are situated within a 10-mile radius of the focal hotel. However, we also discovered that some competitors are located more than 20 miles away from the focal hotel, and many of these distant competitors could be hotels within the same chain as the focal hotel. This finding is important, as it is not usual for hotel managers in densely populated regions like Washington to consider distant competitors.

Table 9: Characteristics of Identified Competitors

	Distance from Competitors (miles)					Brand of Competitors (%)	
	5%	25%	50%	75%	95%	Same brand tier	Adjacent and same brand tier
Washington, D.C.							
Transient (hotel-owned channels)	1.3	5.1	8.9	12.3	20.3	34.7	72.8
Transient (third-party channels)	1.3	6.5	9.8	13.1	21.4	42.9	71.4
Corporate	.8	4.2	8.3	11.7	18.5	34.1	82.6
Group	.5	2.2	7.7	10.8	15.3	35.6	83.6
Wholesale	.4	1.3	4.3	8.8	14.0	48.2	87.5
Employee/loyalty-redemption	.3	8.8	10.9	14.2	21.0	18.2	68.2
Amarillo, TX							
Transient (hotel-owned channels)	2.6	54.5	116.0	148.5	250.0	35.2	70.9
Transient (third-party channels)	2.7	41.4	131.2	147.0	197.4	28.7	61.6
Corporate	2.2	30.0	108.8	146.2	199.6	26.9	73.8
Group	0	0	0	0	0	100	100
Wholesale	0.3	36.2	143.9	143.9	143.9	33.3	100
Employee/loyalty-redemption	1.3	5.8	141.4	144.9	198.1	36.2	84.3

In the Amarillo region, the median distance between a focal hotel and its competitors is over 100 miles, much farther than in the Washington region. Our results also suggest that the geographical boundaries of competition in sparsely populated regions can encompass vast areas. We find that a significant portion of the identified competing hotels belong to the same brand tier as the focal hotel. On average, approximately 60% to 100% of the identified competing hotels are in adjacent or the same brand tiers as the focal hotel. However, our findings underscore that hotel managers who solely focus on competing hotels of similar or adjacent brand tiers may disregard a significant number of relevant competitors whose pricing strategies can also affect the demand for the focal hotel.

In Figure 6, we provide several examples of the competitive sets identified by our method and those identified by proximity and brand tier in both regions. The degree of overlap between the competitive sets identified by the two methods could vary across different focal hotels. We select focal hotels shown in the maps from the transient (hotel-owned channels) segment. This particular segment has the largest total number of hotels and the highest revenue shares in both regions, making it more representative and important than other segments.

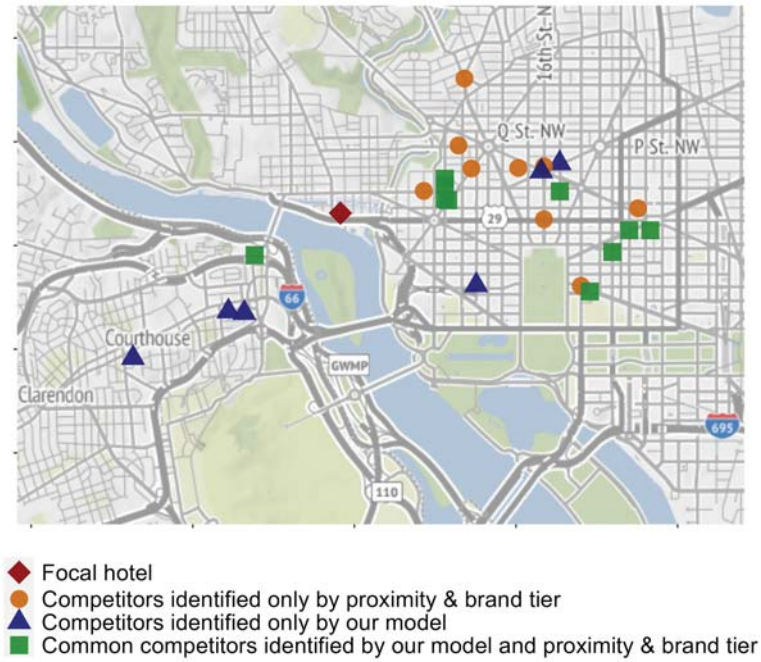
In Figure 6(a), for instance, the focal hotel is a Ritz-Carlton hotel, a luxury Marriott property. Compared to nearby competing hotels identified based on proximity and brand tier, our method identifies luxury and upscale Marriott hotels as competitors (blue triangles and green squares). In Figure 6(b), the focal hotel is a Residence Inn, an upscale Marriott property. Our method identifies two properties far away from the focal hotel (a green square and a blue triangle) - both of which are also Residence Inn. These findings reinforce the idea that while proximity and brand tier rules can sometimes identify relevant competitors, they may miss out on others. Therefore, our method results could serve as a useful complement to factors such as proximity and brand tier.

One particularly interesting observation is that hotels in the same chain may still compete with each other even when they are much farther away from the focal hotel. This could be explained by customers accessing the hotel chain's website or mobile app and choosing among the chain's properties. Loyalty programs offered by the chains also generally lead customers to their channels, which could explain such competition. Our conversations with hotel managers revealed that while they were aware of such possibilities arising from customers' usage of hotel-owned channels, they were not always able to identify the specific properties that competed with their focal hotel. This is where our method's strength lies – in its ability to identify such competitors.

The focal hotel in Figures 6(c) is a Holiday Inn Express in the Amarillo region, which is an upper midscale hotel by InterContinental Hotels Group (IHG). Our results suggest that it competes with quite a few IHG hotels a little farther away but not with the Comfort Suite next to it. In Figure 6(d), the focal hotel is a Baymont Inn & Suites, a midscale hotel by Wyndham group. Our results suggest that it competes with three economy and midscale Wyndham properties and a Country Inn & Suites, which are also identified by the proximity and brand tier rule.

Figure 6: Examples of Identified Competitive Sets

(a) Washington, D.C. (High Overlap)



(b) Washington, D.C. (Low Overlap)

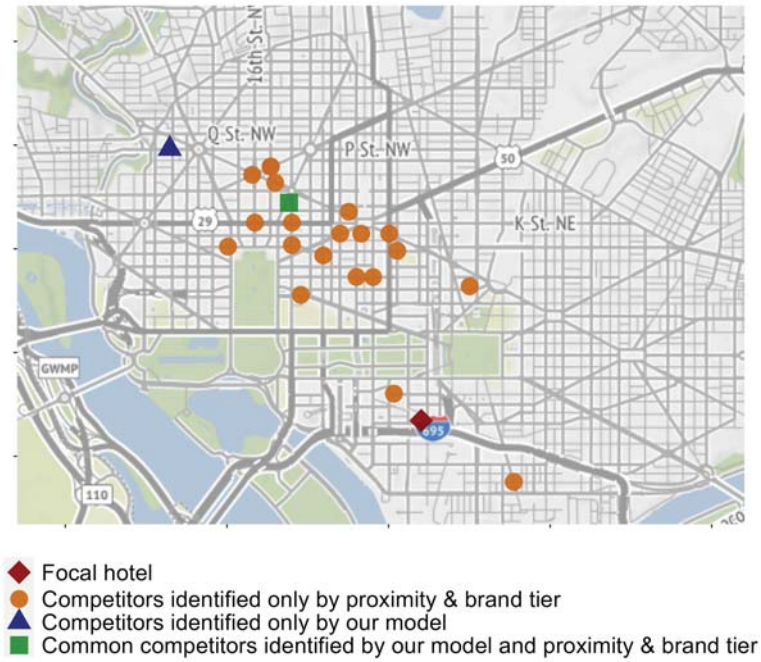
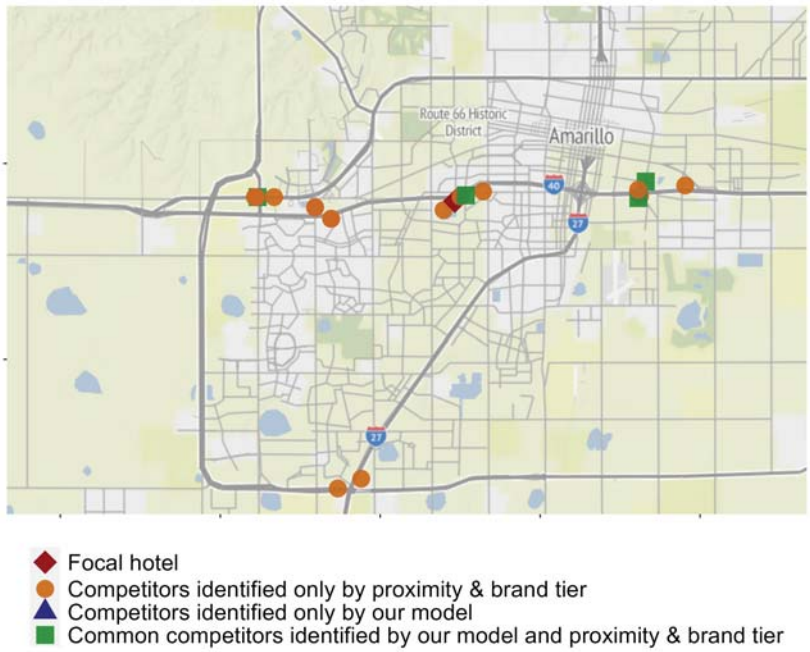


Figure 6: Examples of Identified Competitive Sets (Continued)

(c) Amarillo, TX (High Overlap)

(d) Amarillo, TX (Low Overlap)



Additional Analysis

This section presents additional analysis to enhance our understanding of hotel competition. First, the sales trends presented in Figure 2 reveal seasonal variations that may play an important role in shaping competitive sets. By understanding hotel price competition across seasons (peak versus off-peak), hotel managers can make informed decisions that enable them to effectively compete in the marketplace during different seasons. Based on Figure 2, we consider the four months from April to July as the peak season, characterized by high demand for hotel rooms, and the four months from November to February as the off-peak season, when demand is low.

Table W5 in Web Appendix shows that in both markets, competitive sets are different across peak and off-peak seasons, with the average overlap of competitive sets between the two seasons ranging from 12% and 100%. Furthermore, the competitive set size is usually larger during peak seasons than during off-peak seasons. This change is expected because of higher demand for hotel rooms, which triggers price competition among competitors seeking to attract price-sensitive customers. During peak season, customers are more likely to compare prices and search for the best deals, which can further drive wider price competition among hotels. Moreover, hotels may have higher operational costs during peak season, such as increased staffing and maintenance costs, which can put pressure on hotels' pricing strategies.

Also, while our main analysis examines competition within the same market segment, it is also possible that hotels compete across different segments. For example, bookings on hilton.com (hotel-owned channels) might be influenced by bookings made through expedia.com (third-party channels). To dive deeper into this broader competition, we expand our analysis to include price variables from all hotels in all segments in the demand function for each hotel i in segment s when using the CSIS method. Table W6 summarizes how competitors are spread across six segments. The findings indicate that the transient segment through hotel-owned channels is most important in influencing demand across other segments, with corporate and third-party transient segments also showing notable impact. Conversely, the wholesale segment appears to have the least influence on competition.

Managerial Implications

Our research focuses on identifying the most relevant competitors of a given hotel within a specific geographical region. Our proposed CSIS method, when applied to a system of demand functions both with and without spatial dependence, has demonstrated exceptional performance in this task. With many hotel properties being owned by asset managers, individual owners, or hotel chains, accurately identifying the competitive set has important implications.

The first implication involves competitive benchmarking, which is of great interest to property owners and asset managers. Determining whether a property is providing returns comparable to hotels in its competitive set involves comparing revenue per available room (RevPAR), occupancy rates, and contribution to operating profit and expense (COPE revenue) across competitive hotels and different segments the hotel serves. Our results show that the competitive set can vary significantly across various segments for a focal hotel. Therefore, accurately identifying the competitive sets for each segment is vital for benchmarking purposes and developing strategies to enhance performance in each segment that the focal hotel serves.

The second implication of accurately identifying a hotel's competitors involves developing effective pricing and revenue management strategies. Identifying the true competitors enables revenue managers to adjust their own pricing and revenue strategies in reaction to those of their competitors. Generally, revenue managers operate at the property level to setting prices. For smaller chains, general managers at the property level may provide input to the chain-level revenue managers on likely competitors. These property-level managers use various criteria, such as proximity, facilities and amenities, hotel size, and unique features of the hotels in identifying competitive sets. However, price competition is a crucial dimension that can only be identified using advanced methods such as ours using historical data. This information can be extremely valuable in complementing other information that managers use to accurately calibrate their competitive sets.

In particular, managers use revenue management systems to automatically pull competitor pricing data from APIs that allow for automated data exchange, data feeds from third-party channels, direct hotel booking systems, and sometimes resorting to manual data entry. This data is then analyzed using revenue management systems to identify pricing patterns and trends among

competitors and forecast future demand, occupancy levels, and room capacity allocation across segments. Managers at the property level adjust their room rates and provide discounts and promotions in real time to optimize revenue and stay competitive. Inaccurate identification of competitors can lead to misinterpretation of market conditions and inaccurate pricing decisions, resulting in lost revenue, missed opportunities, and reduced profitability. For example, including too many competitors can lead to data overload, causing challenges in data collection and analysis. This, in turn, wastes resources and time. It may also expose hotels to the risk of undercutting or overpricing when irrelevant competitors are considered in pricing decisions.

Finally, competitor identification has a direct impact on resource allocation decisions for hotels. When facing intense competition in a specific segment, hotels can allocate additional resources to that segment, enhancing amenities and increasing marketing efforts. Omitting key competitors, however, can result in an incomplete understanding of the competitive landscape, thereby causing hotels to overlook opportunities or threats. For instance, if Hotel A is a competitor for Hotel B in the corporate segment but is omitted from Hotel B's competitive set, Hotel B may fail to respond to Hotel A's strategies, like expanding conference spaces. This could potentially lead to a loss of customers from Hotel B to Hotel A.

Conclusions

Our study addresses three key issues associated with competitor identification in a geographical region. First, we demonstrate the effective performance of the CSIS method across different spatial specifications. Second, the CSIS method can capture asymmetric competition, a common feature of hotel markets, and tackle the challenge of high dimensionality resulting from large competitive sets. Finally, the CSIS method offers remarkable computational efficiency because it does not require the estimation of a full model for all competitors. We conduct extensive simulation analyses to demonstrate the accuracy and superiority of our inference about hotel market structure. Using data from hotels across various U.S. geographical regions, the empirical application of our method identifies various patterns of hotel competition across geographical densities, market segments, and peak/off-peak seasons.

Our method also opens up opportunities to meld econometric approaches for identifying market structure based on price coefficients with machine learning-based mapping approaches, which are becoming common in the marketing literature (e.g., Gabel, Guhl, and Klapper 2019; Netzer et al. 2012; Ringel and Skiera 2016). By incorporating the implicit spatial dependence among brands or products based on various product features and social media mentions, the CSIS method can effectively account for the latent spatial structures present in these applications. This makes our method highly adaptable and widely applicable beyond its initial scope. Furthermore, the CSIS method is versatile enough to identify competitors in markets with far more businesses than the hotel industry, such as restaurants, automotive services, and e-commerce platforms. This is because the method has been successfully applied in fields like genomics and neuroscience, which deal with more than thousands of variables.

Our research has several limitations that require further investigation. First, we examine the price competition among hotels and use fixed effects to control for other hotel characteristics such as online word-of-mouth and traffic conditions. However, scholars interested in exploring other dimensions of firm competition, such as the quality of service, can adapt our method to include those dimensions in the future. Second, utilizing daily sales and price data may not be the only approach for identifying competitors. Ideally, future researchers may use consumer-level data that describe consumers' consideration and decision-making processes, such as consumer search, for their analysis. Third, we acknowledge that another source of endogeneity may be hotel managers' own understanding of competition when determining pricing strategies, which may correlate with hotel demand and lead to endogeneity bias. While we address it using instrumental variables, future studies may overcome this limitation by modeling hotel managers' competitive considerations. Also, we focus our analysis on competition among branded hotels of hotel groups, but there are other accommodation businesses, such as independent hotels, Airbnb-type guest houses, bed and breakfasts, and lodges, that also operate in the hotel markets. Although many of these businesses may not compete with branded hotels in market segments such as Corporate and Group, they may still impact competition in the Transient segments. Future research can provide a more comprehensive understanding of hotel competition by including data covering all types of accommodation businesses in the hotel market. In addition, our competitive analysis is based on

pre-defined market segments; however, future research may gather consumer-level reservations to identify the competitive boundaries that represent consumers' decision-making between booking channels. Finally, our method does not estimate the entire demand model. Future research may conduct a counterfactual equilibrium analysis of optimal locations for hotel entry if the entire demand model is estimated.

References

Banerjee, Sudipto, Bradley P Carlin, and Alan E Gelfand (2014). *Hierarchical Modeling and Analysis for Spatial Data*. CRC press.

Barut, Emre, Jianqing Fan, and Anneleen Verhasselt (2016), Conditional Sure Independence Screening. *Journal of the American Statistical Association*, 111(515), 1266–1277.

Baum, Joel AC and Theresa K Lant (2003). Hits and Misses: Managers' (Mis)categorization of Competitors in the Manhattan Hotel Industry. In *Geography and Strategy*. Emerald Group Publishing Limited.

Bhadra, Anindya, Jyotishka Datta, Nicholas G Polson, and Brandon Willard (2019), Lasso Meets Horseshoe: A Survey. *Statistical Science*, 34(3), 405–427.

Bradlow, Eric T, Bart Bronnenberg, Gary J Russell, Neeraj Arora, David R Bell, Sri Devi Duvvuri, Frankel Ter Hofstede, Catarina Sismeiro, Raphael Thomadsen, and Sha Yang (2005), Spatial Models in Marketing. *Marketing Letters*, 16(3), 267–278.

Bronnenberg, Bart J and Vijay Mahajan (2001), Unobserved Retailer Behavior in Multimarket Data: Joint Spatial Dependence in Market Shares and Promotion Variables. *Marketing Science*, 20(3), 284–299.

Brunsdon, Chris, Stewart Fotheringham, and Martin Charlton (1998), Geographically Weighted Regression. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47(3), 431–443.

Bucklin, Randolph E and V Srinivasan (1989). *Determining Cross-price Elasticities and Product-market Structure Through the Measurement of Consumer Preference Structures*. Graduate School of Business, Stanford University.

Chernozhukov, Victor, Wolfgang Karl Härdle, Chen Huang, and Weining Wang (2021), Lasso-Driven Inference in Time and Space. *The Annals of Statistics*, 49(3), 1702–1735.

Conlon, Christopher and Julie Holland Mortimer (2021), Empirical Properties of Diversion Ratios. *The RAND Journal of Economics*, 52(4), 693–726.

Davis, Peter (2006), Spatial Competition in Retail Markets: Movie Theaters. *The RAND Journal of Economics*, 37(4), 964–982.

DellaVigna, Stefano and Matthew Gentzkow (2019), Uniform Pricing in U.S. Retail Chains. *The Quarterly Journal of Economics*, 134(4), 2011–2084.

DeSarbo, Wayne S, Simon J Blanchard, and A Selin Atalay (2017). A New Spatial Classification Methodology for Simultaneous Segmentation, Targeting, and Positioning (STP Analysis) for Marketing Research. In *Review of Marketing Research*, pp. 75–103. Routledge.

Duan, Jason A and Carl F Mela (2009), The Role of Spatial Demand on Outlet Location and Pricing. *Journal of Marketing Research*, 46(2), 260–278.

Elrod, Terry, Gary J Russell, Allan D Shocker, Rick L Andrews, Lynd Bacon, Barry L Bayus, J Douglas Carroll, Richard M Johnson, Wagner A Kamakura, Peter Lenk, et al. (2002), Inferring Market Structure from Customer Response to Competing and Complementary Products. *Marketing Letters*, 13(3), 221–232.

Fan, Jianqing and Runze Li (2001), Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties. *Journal of the American statistical Association*, 96(456), 1348–1360.

Fan, Jianqing and Jinchi Lv (2008), Sure Independence Screening for Ultrahigh Dimensional Feature Space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(5), 849–911.

Fan, Jianqing, Richard Samworth, and Yichao Wu (2009), Ultrahigh dimensional feature selection: beyond the linear model. *The Journal of Machine Learning Research*, *10*, 2013–2038.

Farrell, Joseph and Carl Shapiro (2010), Antitrust Evaluation of Horizontal Mergers: An Economic Alternative to Market Definition. *The B.E. Journal of Theoretical Economics*, *10*, 1–41.

Gabel, Sebastian, Daniel Guhl, and Daniel Klapper (2019), P2V-MAP: Mapping Market Structures for Large Retail Assortments. *Journal of Marketing Research*, 56(4), 557–580.

Gerardi, Kristopher S and Adam Hale Shapiro (2009), Does Competition Reduce Price Dispersion? New Evidence from the Airline Industry. *Journal of Political Economy*, 117(1), 1–37.

Govind, Rahul, Rabikar Chatterjee, and Vikas Mittal (2018), Segmentation of Spatially Dependent Geographical Units: Model and Application. *Management Science*, 64(4), 1941–1956.

Gowrisankaran, Gautam, Aviv Nevo, and Robert Town (2015), Mergers When Prices are Negotiated: Evidence from the Hospital Industry. *American Economic Review*, 105(1), 172–203.

Granados, Nelson, Alok Gupta, and Robert J Kauffman (2012), Online and Offline Demand and Price Elasticities: Evidence from the Air Travel Industry. *Information Systems Research*, 23(1), 164–181.

Green, Cindy Estis and Mark V. Lomanno (2016). Demystifying the Digital Marketplace: Spotlight on the Hospitality Industry. Technical report, Kalibri Labs.

Hastie, Trevor, Robert Tibshirani, Jerome H Friedman, and Jerome H Friedman (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Volume 2. Springer.

Hausman, Jerry A (1996). Valuation of New Goods under Perfect and Imperfect Competition. In *The Economics of New Goods*, pp. 207–248. University of Chicago Press.

Hofstede, Frenkel Ter, Michel Wedel, and Jan-Benedict EM Steenkamp (2002), Identifying Spatial Segments in International Markets. *Marketing Science*, 21(2), 160–177.

Houde, Jean-François (2012), Spatial Differentiation and Vertical Mergers in Retail Markets for Gasoline. *American Economic Review*, 102(5), 2147–82.

InterVISTAS (2007, December). Estimating Air Travel Demand Elasticities. Technical report, International Air Transport Association.

Jank, Wolfgang and PK Kannan (2005), Understanding Geographical Markets of Online Firms Using Spatial Models of Customer Choice. *Marketing Science*, 24(4), 623–634.

Joo, Mingyu, Dinesh K Gauri, and Kenneth C Wilbur (2020), Temporal Distance and Price Responsiveness: Empirical Investigation of the Cruise Industry. *Management Science*, 66(11), 5362–5388.

Kannan, PK and Susan M Sanchez (1994), Competitive Market Structures: A Subset Selection Analysis. *Management Science*, 40(11), 1484–1499.

Lee, Seul Ki (2015), Quality Differentiation and Conditional Spatial Price Competition among Hotels. *Tourism Management*, *46*, 114–122.

Lee, Thomas Y and Eric T Bradlow (2011), Automated Marketing Research Using Online Customer Reviews. *Journal of Marketing Research*, 48(5), 881–894.

Li, Jun, Serguei Netessine, and Sergei Koulayev (2018), Price to compete ... with many: How to identify price competition in high-dimensional space. *Management Science*, 64(9), 4118–4136.

Madhok, Roy and Theresa Doherty (2020). Evolving Hotel Segmentation. In *Hospitality Revenue Management*, pp. 61–83. Apple Academic Press.

Mohammed, Ibrahim, Basak Denizci Guillet, and Rob Law (2014), Competitor Set Identification in the Hotel Industry: A Case Study of A Full-Service Hotel in Hong Kong. *International Journal of Hospitality Management*, *39*, 29–40.

Netzer, Oded, Ronen Feldman, Jacob Goldenberg, and Moshe Fresko (2012), Mine Your Own Business: Market-Structure Surveillance through Text Mining. *Marketing Science*, 31(3), 521–543.

Nevo, Aviv (2000), Mergers with Differentiated Products: The Case of the Ready-to-Eat Cereal Industry. *The RAND Journal of Economics*, 395–421.

Nevo, Aviv (2001), Measuring Market Power in the Ready-to-Eat Cereal Industry. *Econometrica*, 69(2), 307–342.

Pinkse, Joris, Margaret E Slade, and Craig Brett (2002), Spatial Price Competition: A Semiparametric Approach. *Econometrica*, 70(3), 1111–1153.

Ringel, Daniel M and Bernd Skiera (2016), Visualizing Asymmetric Competition Among More Than 1,000 Products Using Big Search Data. *Marketing Science*, 35(3), 511–534.

Rossi, Peter E, Greg M Allenby, and Rob McCulloch (2005). *Bayesian Statistics and Marketing*. John Wiley & Sons.

Rossi, Peter E, Greg M Allenby, and Rob McCulloch (2012). *Bayesian Statistics and Marketing*. John Wiley & Sons.

Russell, Gary J (1992), A Model of Latent Symmetry in Cross Price Elasticities. *Marketing Letters*, 3(2), 157–169.

Schwartz, Zvi and Timothy Webb (2022), Resource Similarity, Market Commonality, and Spatial Distribution of Hotel Competitive Sets. *Journal of Hospitality & Tourism Research*, 46(4), 724–741.

Schwartz, Zvi, Timothy Webb, and Jing Ma (2021), Hotel Analytics: The Case for Reverse Competitive Sets. *Cornell Hospitality Quarterly*, 19389655211036656.

Shugan, Steven M. (2014). Market Structure Research. In Russell S Winer and Scott A Neslin (Eds.), *The History of Marketing Science*, Chapter 6, pp. 129–164. World Scientific.

Smith, Adam N, Peter E Rossi, and Greg M Allenby (2019), Inference for Product Competition and Separable Demand. *Marketing Science*, 38(4), 690–710.

Staiger, Douglas and James H. Stock (1997), Instrumental Variables Regression with Weak Instruments. *Econometrica*, 65(3), 557–586.

STR (2021). Tourism After Lockdown: Hotel Guests Expect Normal In The New Normal. `https://str.com/data-insights-blog/tourism-after-lockdown-hotel-guests-expect-normal-in-the-new-normal`.

Thomadsen, Raphael (2005), The Effect of Ownership Structure on Prices in Geographically Differentiated Industries. *RAND Journal of Economics*, 908–929.

Tibshirani, Robert (1996), Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267–288.

Wöber, Karl W (2002). *Benchmarking in Tourism and Hospitality Industries: The Selection of Benchmarking Partners*. CABI.

Yang, Yi, Kunpeng Zhang, and PK Kannan (2022), Identifying Market Structure: A Deep Network Representation Learning of Social Engagement. *Journal of Marketing*, 86(4), 37–56.

Ye, Fei, Qian Xia, Minhao Zhang, Yuanzhu Zhan, and Yina Li (2022), Harvesting Online Reviews to Identify the Competitor Set in a Service Business: Evidence from the Hotel Industry. *Journal of Service Research*, 25(2), 301–327.

Yuan, Ming and Yi Lin (2007), Model Selection and Estimation in the Gaussian Graphical Model. *Biometrika*, 94(1), 19–35.

Zou, Hui and Trevor Hastie (2005), Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2), 301–320.

Web Appendix

Identifying Competitors in Geographical Markets Using the CSIS Method

Xian Gu

Assistant Professor in Marketing

Kelley School of Business

Indiana University, Bloomington

HH 2100, 1309 E. 10TH St., Bloomington, IN 47405

Tel: 812-856-1073

xiangu@iu.edu

P. K. Kannan

Dean's Chair in Marketing Science

Robert H. Smith School of Business

University of Maryland

3445 Van Munching Hall, College Park, MD 20742

Tel: 301-405-2188

pkannan@umd.edu

Web Appendix A: Comparison of Methods

Table W1: Comparison between CSIS and Penalized Likelihood Methods

	CSIS	Penalized Likelihood Methods
Basic Idea	CSIS is a screening method that selects candidate variables based on their relevance to the outcome variable, conditional on any other variables that are known to be relevant.	The penalized likelihood methods are primarily designed to maintain the overall predictive power of a model while reducing the number of predictors.
Estimation Specifics	CSIS computes a score for each candidate variable based on their relevance to the outcome variable, which is derived by maximizing the marginal likelihood function for each candidate variable conditional on any other variables that are known to be relevant to the outcome. Candidate variables are ranked based on their relevance scores, and those exceeding a threshold value are retained. To determine this threshold, CSIS computes relevance scores based on randomly permuted values for each candidate variable, similar to the importance scores in Random Forest or Gradient Boosting Machine (GBM) models. Subsequently, we set the threshold at the 99th percentile of the relevance scores, representing the maximum value of coefficients achievable under pure noise.	The penalized likelihood methods minimize the model's root mean square error (RMSE) with a penalty term based on the magnitudes of coefficients. This aims to shrink the coefficients of less relevant variables towards zero, thereby reducing the value of the objective function and creating a simpler model.
Model Objective	CSIS aims to select candidate variables that are relevant to the outcome variable based on their conditional maximum marginal likelihood estimators. CSIS does not compromise on variable selection to maintain the overall model's predictive power.	The penalized likelihood methods aim to maximize the overall model predictive power across all variables.
Variable Selection	CSIS has a desirable "sure screening" property: as the number of observations tends towards infinity (in our case, as $T \to \infty$), the probability of the CSIS retaining all relevant variables in the model converges to one.	The penalized likelihood methods may overlook a more important variable with a larger coefficient in favor of multiple smaller coefficients contributing to higher overall predictive power.
Correlated Predictors	CSIS computes the conditional maximum marginal likelihood estimators for each candidate variable independently and thus does not require trading off variables even in the case of high correlation.	In cases of high correlation among predictors, the penalized likelihood methods may arbitrarily choose one variable, leading to the exclusion of other important variables from the model.

Web Appendix B: Spatial CSIS Method

B1: Modeling Spatial Dependence

To account for the spatial dependence among hotels, we employ a spatial structure for the price coefficients in the demand model (Bradlow et al. 2005).

$$\log q_{sit} = \alpha_{si} + \beta_{s,ii} \log p_{sit} + \sum_{j \neq i} \left(\tilde{\beta}_{s,ij} + \rho_{si} w_{s,ij} \right) \log p_{sjt} + z_t' \gamma_{si} + \epsilon_{sit} \tag{W1}$$

In this equation, ρ_{si} is a scaling parameter, and $w_{s,ij}$ is a spatial weighting function. This functional form is flexible because it allows the empirical data to determine whether the spatial dependence influences demand estimation ($\rho_{si} \neq 0$) or not ($\rho_{si} = 0$). We adopt the widely used Gaussian kernel weighting function (Brunsdon, Fotheringham, and Charlton 1998):

$$w_{s,ij} = w_s(d_{ij}) = \exp\left(-\frac{d_{ij}^2}{2h_{si}^2} \right), \ \forall i, j \tag{W2}$$

where $d_{ij} = \|\mathbf{L}_i - \mathbf{L}_j\|$ denote the driving distance between hotel i and j, and $\mathbf{L} = (\mathbf{L}_1, \ldots, \mathbf{L}_N)$ denotes a vector of geographical coordinates (latitudes and longitudes) of the N hotels. The weighting function satisfies desirable properties: (1) $w(0) = 1$, (2) $\lim_{d \to \infty} w(d) = 0$, and (3) $w(\cdot)$ is a monotone decreasing function for positive real numbers.

Since the strength of spatial dependence may vary across hotels and market segments, we allow the scaling parameter, ρ_{si}, and the kernel bandwidth, h_{si}, to be hotel-and-segment-specific. Following previous literature on spatial models (Brunsdon, Fotheringham, and Charlton 1998), we minimize the sum of squared errors to determine the optimal parameter values:

$$\{\rho_{si}^*, h_{si}^*\} = \underset{\rho_{si}, h_{si}}{\arg\min} \left\{ \sum_{t=1}^{T} \left[\log q_{sit} - \log \hat{q}_{sit}(\rho_{si}, h_{si}) \right]^2 \right\} \tag{W3}$$

where $\log \hat{q}_{sit}$ represents the fitted value of $\log q_{sit}$ from Equation W1. This optimization is performed prior to the competitor identification and with all price variables retained in the model.

B2: Spatial CSIS

While the optimization problem in the spatial CSIS method remains the same as in Equation 4, we have an alternative condition that the price coefficients of relevant competitors. That is,

$\tilde{\beta}^*_{s,ij} = \beta^*_{s,ij} - \rho_{si} w_{s,ij}$, exceed the threshold value $\tilde{\delta}_{si} = \kappa \tilde{\lambda}_{si,\tau}$, where $\tilde{\lambda}_{si,\tau}$ is the $\tau-$quantile of the values $\left\{ \left| \breve{\beta}^{(k)}_{s,ij} - \rho_{si} w_{s,ij} \right|, \ \forall j \neq i \right\}_{k=1}^{K}$. Thus, the alternative set of identified competitors for hotel i in segment s, $J^{c,sp}_{si}$, is given by:

$$J^{c,sp}_{si} = \left\{ j \neq i : \left\{ \left| \tilde{\beta}^*_{s,ij} \right| \geqslant \tilde{\delta}_{si} \right\} \right\} \tag{W4}$$

This criterion enables spatial factors to influence the identification of the competitive sets. We refer to this method as the "spatial CSIS" method. It is worth noting that the spatial CSIS method can incorporate only one type of spatial dependence (spatial drift), and estimating spatial parameters can be challenging when there are many potential competitors. Given the original CSIS method's robustness to different types of spatial dependence structures and its computational efficiency, we focus on the original CSIS method in our paper.

B3: Results

Figure W1 shows the overall simulation results for the spatial CSIS method across all simulated datasets, alongside the results of other methods used in our simulation analysis. Across all boundary conditions, the original CSIS method performs better than the spatial CSIS method on average. We also examine different values of the spatial parameter, ρ_{si}, and find that the spatial CSIS method may slightly outperform the original CSIS method when ρ_{si} is around -0.1 and -0.2 in our simulation.

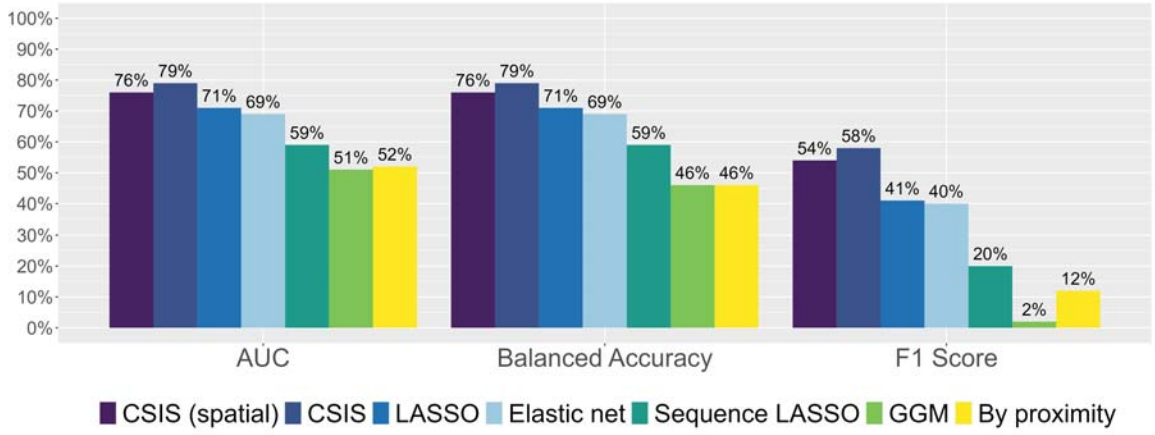
In the empirical application, the spatial CSIS method and the original CSIS method produce comparable results, as presented in Table W2.

Table W2: Percentage of Common Competitors Identified Using the Original and Spatial CSIS Methods

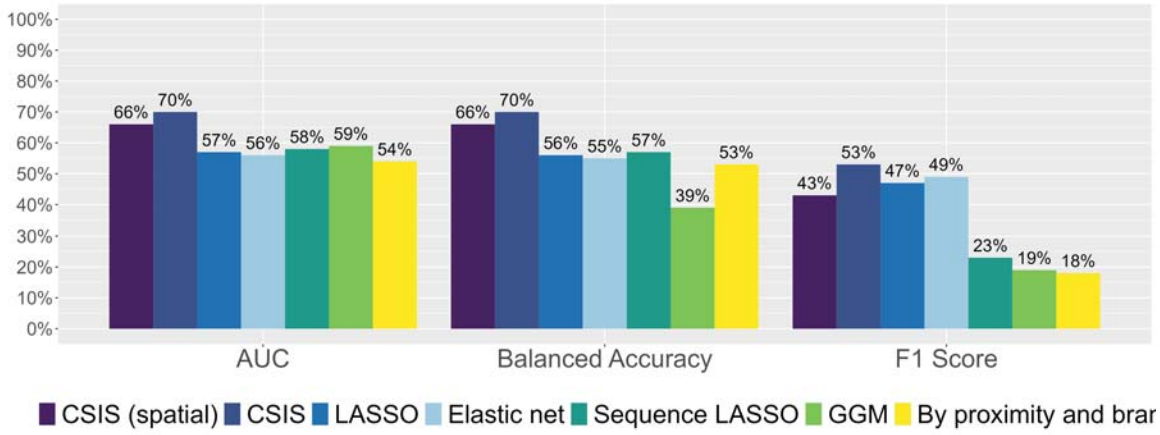
	Washington, D.C.	**Amarillo, TX**
Transient (hotel-owned channels)	97.7%	98.2%
Transient (third-party channels)	94.6%	90.2%
Corporate	97.2%	94.9%
Group	96.8%	0%
Wholesale	87.3%	100%
Employee/loyalty-redemption	100%	100%

Figure W1: Overall Simulation Results for Spatial CSIS

(a) Random Prices



(b) Actual Prices



Web Appendix C: Validation by Simulation Analysis

Table W3: Simulation Results (Random Prices)

Method	AUC (%)	Balanced Accuracy (%)	F1-Score (%)
Overall (384 simulated datasets)			
CSIS	79	79	58.3
LASSO	71.1	71.1	41
Elastic Net	69.3	69.3	39.6
Sequence LASSO	59	58.9	20.4
GGM	51	46.3	1.9
By Proximity	52.1	46.4	12.2
Between 1 and 5 Competitors (192 simulated datasets)			
CSIS	92.8	92.8	77.2
LASSO	78.7	78.7	30.2
Elastic Net	76.2	76.2	24.8
Sequence LASSO	64.3	64.3	27.6
GGM	49.1	49.1	0.1
By Proximity	49.7	49.7	6.8
Between 15 and 20 Competitors (192 simulated datasets)			
CSIS	82.8	82.8	73.5
LASSO	70.1	70.1	61.5
Elastic Net	68	68	58.8
Sequence LASSO	57.2	56.9	21.6
GGM	52.4	43.8	2.4
By Proximity	54.5	43.2	17.2
Non-Spatial (96 simulated datasets)			
CSIS	91.7	91.7	81.6
LASSO	77	77	48.2
Elastic Net	74.4	74.4	43.7
Sequence LASSO	63.2	63	30.2
GGM	51	46.3	1.2
By Proximity	52.1	46.5	12
Spatial Drift (96 simulated datasets)			
CSIS	86.7	86.7	73.6
LASSO	73.9	73.9	45.6
Elastic Net	71.9	71.9	42
Sequence LASSO	59.8	59.5	23.6

Table W3 – continued from previous page

Method	AUC (%)	Balanced Accuracy (%)	F1-Score (%)
GGM	51	46.1	0.9
By Proximity	52.1	46.5	12
Spatial Lag (96 simulated datasets)			
CSIS	83.7	83.7	69.1
LASSO	70	70	41.8
Elastic Net	68.1	68.1	38.6
Sequence LASSO	58.9	58.9	20.9
GGM	50.1	47.1	1.5
By Proximity	52.1	46.5	12
Spatial Error (96 simulated datasets)			
CSIS	85.5	85.5	71
LASSO	74.6	74.6	45.8
Elastic Net	71.8	71.8	41.6
Sequence LASSO	59	58.7	19.4
GGM	51	46.3	1
By Proximity	52.1	46.5	12
One Month (96 simulated datasets)			
CSIS	52.5	52.5	7
LASSO	61.2	61	26.5
Elastic Net	61.1	60.9	32.8
Sequence LASSO	53.7	53.6	7.6
GGM	51.9	45.6	3.9
By Proximity	52.3	46	12.8
Six Months (96 simulated datasets)			
CSIS	79.8	79.8	61.7
LASSO	71.9	71.9	44.9
Elastic Net	70.6	70.6	41.9
Sequence LASSO	58.8	58.8	20.8
GGM	51.5	46.2	1.5
By Proximity	52.9	45.7	12.3
One Year (96 simulated datasets)			
CSIS	89.2	89.2	79.4
LASSO	75.6	75.6	46.5
Elastic Net	72.7	72.7	41.8
Sequence LASSO	62.2	61.9	26.8
GGM	50	46.9	1.1
By Proximity	51.5	47	11.6

Table W3 – continued from previous page

Method	AUC (%)	Balanced Accuracy (%)	F1-Score (%)
Two Years (96 simulated datasets)			
CSIS	94.4	94.4	85
LASSO	75.7	75.7	46.1
Elastic Net	72.9	72.9	41.7
Sequence LASSO	61.3	61.1	26.2
GGM	50.7	46.3	1.1
By Proximity	51.8	46.7	12.2
Average Price Coefficients = .5 (192 simulated datasets)			
CSIS	81.3	81.3	65.2
LASSO	70.8	70.8	43.6
Elastic Net	69.3	69.3	40.3
Sequence LASSO	57.4	57.3	16.5
GGM	50.4	45.6	1.4
By Proximity	52.1	46.5	12
Average Price Coefficients = 1.5 (192 simulated datasets)			
CSIS	94.3	94.3	85.5
LASSO	78.1	78.1	48.1
Elastic Net	74.8	74.8	43.3
Sequence LASSO	64.1	64	32.7
GGM	51.2	47.3	1
By Proximity	52.1	46.5	12
20 Hotels in the Market (128 simulated datasets)			
CSIS	94.8	94.8	94.1
LASSO	57.7	57.7	62
Elastic Net	53.8	53.8	59.5
Sequence LASSO	78.1	77.6	64.8
GGM	52.8	39.9	3.1
By Proximity	56.5	39.6	19.1
100 Hotels in the Market (128 simulated datasets)			
CSIS	87.1	87.1	71.1
LASSO	83	83	39.3
Elastic Net	80.1	80.1	34.6
Sequence LASSO	54	54	8.3
GGM	49.7	49.7	0.2
By Proximity	50	50	10.4
200 Hotels in the Market (128 simulated datasets)			
CSIS	81.6	81.6	60.8

Table W3 – continued from previous page

Method	AUC (%)	Balanced Accuracy (%)	F1-Score (%)
LASSO	82.5	82.5	36.2
Elastic Net	82.4	82.4	31.3
Sequence LASSO	50.3	50.3	0.7
GGM	49.9	49.9	0.3
By Proximity	49.8	49.8	6.6
Dense Region (192 simulated datasets)			
CSIS	88	88	75.5
LASSO	74.5	74.5	46
Elastic Net	72.1	72.1	41.8
Sequence LASSO	60.9	60.7	25.1
GGM	50.8	46.5	1.4
By Proximity	52.2	46.2	10.1
Sparse Region (192 simulated datasets)			
CSIS	87.7	87.7	75.2
LASSO	74.3	74.3	45.7
Elastic Net	72	72	41.8
Sequence LASSO	60.6	60.5	24.1
GGM	50.8	46.4	1.1
By Proximity	52	46.8	13.9

Table W4: Simulation Results (Actual Prices)

Method	AUC (%)	Balanced Accuracy (%)	F1-Score (%)
Overall (1,536 simulated datasets)			
CSIS	70.3	70	52.6
LASSO	57.2	56.1	47.3
Elastic Net	56.4	54.9	49.3
Sequence LASSO	58.2	57.2	23.5
GGM	59	39.4	18.8
By Proximity and Brand	54.3	53.4	18.3
Between 1 and 5 Competitors (768 simulated datasets)			
CSIS	74.2	74	50.7
LASSO	58.1	57.5	39.9
Elastic Net	56.9	55.9	38.7
Sequence LASSO	60.4	59.5	28.6
GGM	56.3	43.4	18.2
By Proximity and Brand	53.8	53.1	15.5
Between 15 and 20 Competitors (768 simulated datasets)			
CSIS	74	73.8	71.2
LASSO	55.3	55	65.8
Elastic Net	54.2	53.9	67
Sequence LASSO	59.4	58.3	27.8
GGM	61.2	36	20
By Proximity and Brand	55.8	54.6	22.8
Non-Spatial (384 simulated datasets)			
CSIS	74.7	74.6	61.9
LASSO	56.7	56.2	53.6
Elastic Net	55.7	55	53.3
Sequence LASSO	60.5	59.3	29.8
GGM	58.7	39.4	18.8
By Proximity and Brand	54.8	53.8	19.2
Spatial Drift (384 simulated datasets)			
CSIS	75	74.9	62.5
LASSO	57.2	56.8	54.3
Elastic Net	56	55.5	53.7
Sequence LASSO	60.5	59.4	30
GGM	58.5	39.3	18.6
By Proximity and Brand	54.8	53.8	19.2
Spatial Lag (384 simulated datasets)			

Table W4 – continued from previous page

Method	AUC (%)	Balanced Accuracy (%)	F1-Score (%)
CSIS	72.9	72.6	59.8
LASSO	55.5	55.1	52.4
Elastic Net	54.3	53.8	52.4
Sequence LASSO	59.4	58.6	27.1
GGM	58.7	40.6	19.5
By Proximity and Brand	54.8	53.8	19.2
Spatial Error (384 simulated datasets)			
CSIS	73.8	73.6	59.4
LASSO	57.5	56.9	51.1
Elastic Net	56.1	55.4	51.8
Sequence LASSO	59.2	58.3	25.9
GGM	59	39.7	19.6
By Proximity and Brand	54.8	53.8	19.2
One Month (384 simulated datasets)			
CSIS	59	58.2	27.6
LASSO	58.7	55.7	30.7
Elastic Net	59.1	54.7	38.8
Sequence LASSO	53.1	52	9.3
GGM	59.9	38.6	17.8
By Proximity and Brand	53	51.9	15.6
Six Months (384 simulated datasets)			
CSIS	71.8	71.5	57.8
LASSO	58.4	57.8	52.1
Elastic Net	56.8	56.3	52.8
Sequence LASSO	58.8	58.1	25.1
GGM	58.2	38.4	18.4
By Proximity and Brand	53.3	52.8	15.9
One Year (384 simulated datasets)			
CSIS	73.7	73.4	58.8
LASSO	55.7	55.1	48.8
Elastic Net	54.8	53.9	48.7
Sequence LASSO	59.2	58.2	23.5
GGM	59.6	42.6	17.4
By Proximity and Brand	56.3	54.9	22.2
Two Years (384 simulated datasets)			
CSIS	76.8	76.8	66.1
LASSO	56.1	55.8	57.6

Table W4 – continued from previous page

Method	AUC (%)	Balanced Accuracy (%)	F1-Score (%)
Elastic Net	55	54.6	57
Sequence LASSO	61.7	60.4	36
GGM	58.4	38.2	21.6
By Proximity and Brand	54.8	53.9	19.5
Average Price Coefficients = .5 (768 simulated datasets)			
CSIS	71.3	71	55.6
LASSO	55.2	54.4	50
Elastic Net	54.6	53.5	51
Sequence LASSO	58.8	57.3	22.1
GGM	59.5	40.4	22
By Proximity and Brand	54.8	53.8	19.2
Average Price Coefficients = 1.5 (768 simulated datasets)			
CSIS	76.8	76.8	66.2
LASSO	58.2	58.1	55.7
Elastic Net	56.5	56.3	54.6
Sequence LASSO	61	60.5	34.3
GGM	58	39.1	16.2
By Proximity and Brand	54.8	53.8	19.2
5 Hotels in the Market (640 simulated datasets)			
CSIS	84.2	83.9	78
LASSO	51.4	51.3	59.9
Elastic Net	50.8	50.7	59.6
Sequence LASSO	68.5	66.6	52.5
GGM	69.4	36.6	34
By Proximity and Brand	58.9	57.3	24.2
20 Hotels in the Market (640 simulated datasets)			
CSIS	71.1	71	60.7
LASSO	61.3	60.3	58.7
Elastic Net	58.9	57.7	58.6
Sequence LASSO	55.3	54.7	15
GGM	51.8	39	11.3
By Proximity and Brand	52.5	51.9	17.8
100 Hotels in the Market (256 simulated datasets)			
CSIS	56.3	56.2	18.7
LASSO	58.5	58.5	20.7
Elastic Net	58.9	58.6	21.2
Sequence LASSO	50.1	50.1	0.4

Table W4 – continued from previous page

Method	AUC (%)	Balanced Accuracy (%)	F1-Score (%)
GGM	49.4	49.4	1.1
By Proximity and Brand	50.2	50.2	9.9
Dense Region (960 simulated datasets)			
CSIS	72.6	72.5	60.3
LASSO	57.3	57	52.4
Elastic Net	56	55.5	52.1
Sequence LASSO	59.4	58.9	26.9
GGM	58.1	40	18.1
By Proximity and Brand	54.2	53.8	17.4
Sparse Region (576 simulated datasets)			
CSIS	76.5	76.3	62
LASSO	55.7	55	53.7
Elastic Net	54.8	54	54.1
Sequence LASSO	60.7	58.9	30.4
GGM	59.8	39.2	20.8
By Proximity and Brand	55.8	54	22.1
Segment 1: Transient-Hotel-Owned-Channels (384 simulated datasets)			
CSIS	68.2	68	49.7
LASSO	56.4	55.8	45.1
Elastic Net	56	55.2	45.7
Sequence LASSO	56.1	54.9	19.4
GGM	56.5	43.3	18.6
By Proximity and Brand	54.3	53.4	18.3
Segment 2: Transient-Third-Party-Channels (320 simulated datasets)			
CSIS	71.3	71.1	56.8
LASSO	58.2	58	53.5
Elastic Net	56.9	56.3	52.9
Sequence LASSO	56.7	55.5	20.2
GGM	59.6	39.9	20.7
By Proximity and Brand	55.5	54.9	20.8
Segment 3: Corporate (320 simulated datasets)			
CSIS	71.8	71.7	54.1
LASSO	56.7	55.9	49.4
Elastic Net	55.5	54.6	49.9
Sequence LASSO	56.4	54.8	19.7
GGM	58.7	40.7	20.2
By Proximity and Brand	56	55.1	21.4

Table W4 – continued from previous page

Method	AUC (%)	Balanced Accuracy (%)	F1-Score (%)
Segment 4: Group (128 simulated datasets)			
CSIS	78	77.9	71.6
LASSO	55.9	55.5	59.6
Elastic Net	54.7	54.3	59.4
Sequence LASSO	60.3	59.5	31.8
GGM	61.7	37.3	23.7
By Proximity and Brand	54.5	53.8	22.7
Segment 5: Wholesale (128 simulated datasets)			
CSIS	74.9	74.7	70.6
LASSO	57.5	57	61
Elastic Net	55.9	55.2	60.4
Sequence LASSO	59.5	59.1	27.3
GGM	60.4	37.2	23.6
By Proximity and Brand	51.8	50.2	8.1
Segment 6: Employee/Loyalty-Redemption (256 simulated datasets)			
CSIS	86.8	86.8	81.3
LASSO	55.3	55.1	60.6
Elastic Net	53.5	53.4	59.8
Sequence LASSO	74.1	73.7	60.6
GGM	58.7	35.5	12
By Proximity and Brand	54.8	53.4	19.3

Web Appendix D: Empirical Application

Table W5: Competitive Set Sizes During Peak and Off-Peak Seasons

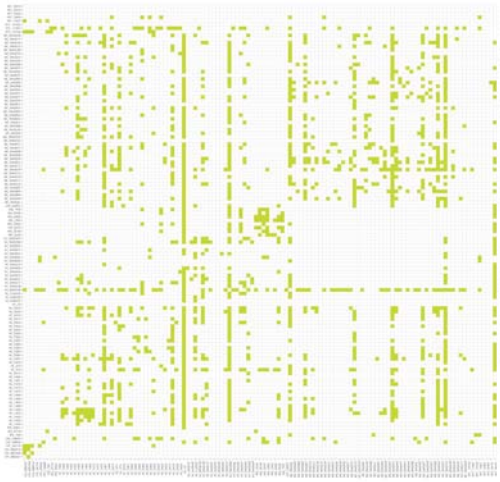
	Washington, D.C.			Amarillo, TX		
Segment	Peak Season	Off-Peak Season	Overlap Between Two Seasons	Peak Season	Off-Peak Season	Overlap Between Two Seasons
Transient (hotel-owned channels)	4.7	2.4	15.7%	6.3	2.8	32.5%
Transient (third-party channels)	1.2	0.5	44.4%	1.8	1	36.7%
Corporate	25.3	8.7	22.4%	3.9	2.1	16%
Group	5	2.9	11.5%	0.5	0.5	100%
Wholesale	0.8	0.5	34%	0.3	0.7	33.3%
Employee/Loyalty-Redemption	0.1	0.2	73.4%	2.7	1.3	54.4%

Table W6: Percentages of Inter-Segment Competitors

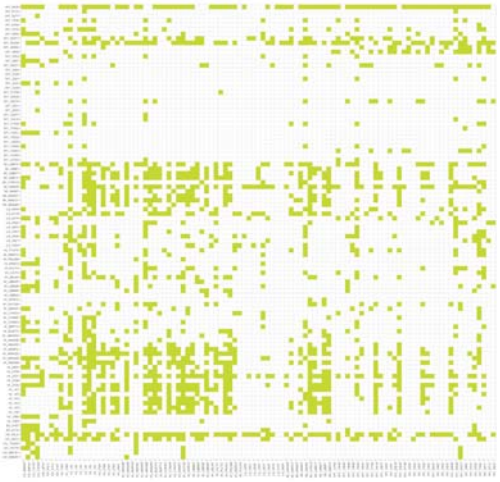
		Source of Competitors					
Seg.	Focal Segment	Seg. 1	Seg. 2	Seg. 3	Seg. 4	Seg. 5	Seg. 6
Washington, D.C.							
1	Transient (hotel-owned channels)	36%	17.8%	31.2%	11.4%	2.4%	1.3%
2	Transient (third-party channels)	36%	9.5%	32.9%	16.9%	3.5%	1.2%
3	Corporate	39.2%	16.2%	30.4%	10.6%	3.2%	.6%
4	Group	26.7%	5.6%	45.6%	16.3%	5.1%	.8%
5	Wholesale	36.8%	10.8%	33.5%	13.1%	4.3%	1.5%
6	Employee/Loyalty-Redemption	40.2%	10.5%	28.9%	16.4%	2.6%	1.5%
Amarillo, TX							
1	Transient (hotel-owned channels)	43.8%	18.1%	20.8%	2.4%	3.1%	12.1%
2	Transient (third-party channels)	42.2%	14.8%	19.6%	3.9%	2.3%	17.3%
3	Corporate	43.2%	19.2%	21.5%	1.8%	3.5%	10.8%
4	Group	30.4%	13%	17.4%	4.3%	4.3%	30.4%
5	Wholesale	49.4%	20.6%	12.9%	1.2%	3.5%	12.3%
6	Employee/Loyalty-Redemption	44.1%	14.2%	19.4%	2.4%	4.4%	15.7%

Figure W2: Graphs of Competition Matrices

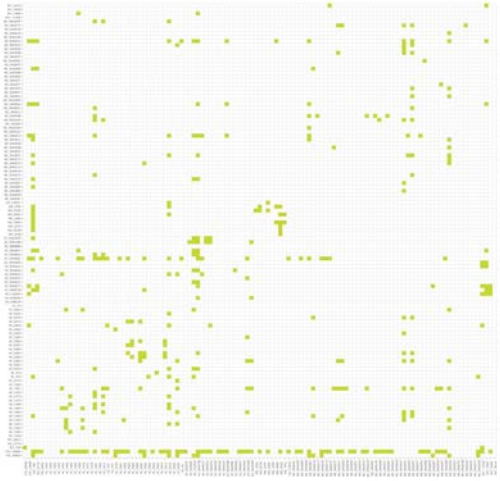
(a) Washington, D.C. - Transient (Hotel-Owned Channels)



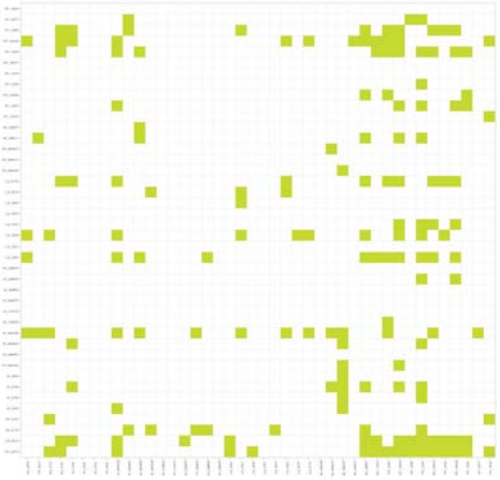
(b) Amarillo, TX - (Hotel-Owned Channels)



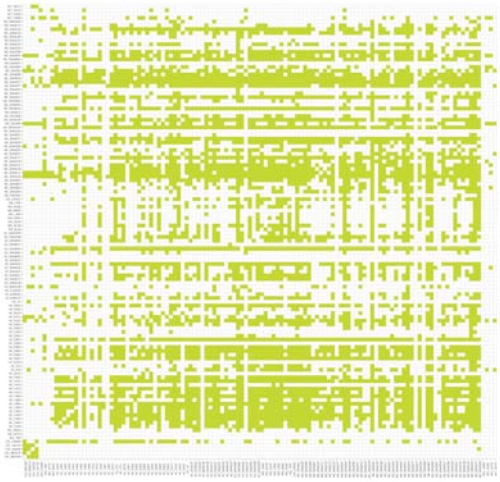
(c) Washington, D.C. - Transient (Third-Party Channels)



(d) Amarillo, TX - Transient (Third-Party Channels)



(e) Washington, D.C. - Corporate



(f) Amarillo, TX - Corporate

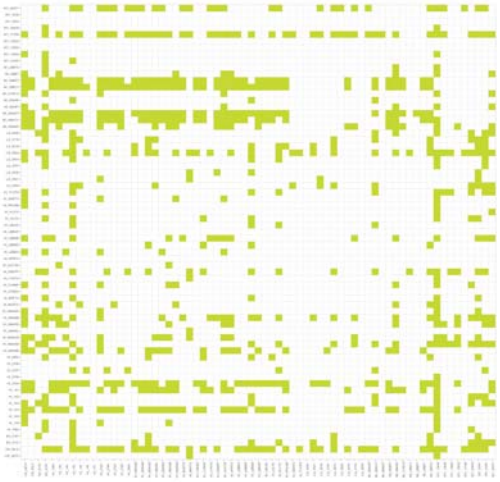
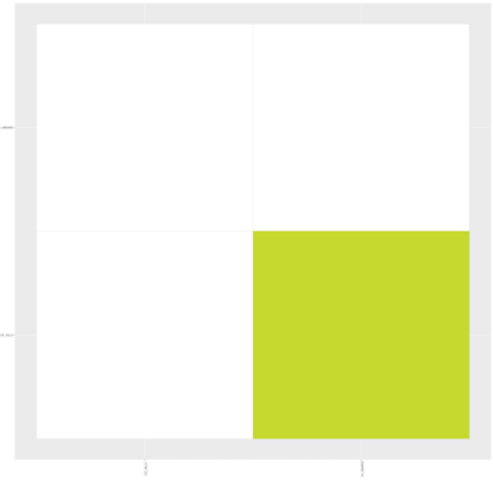


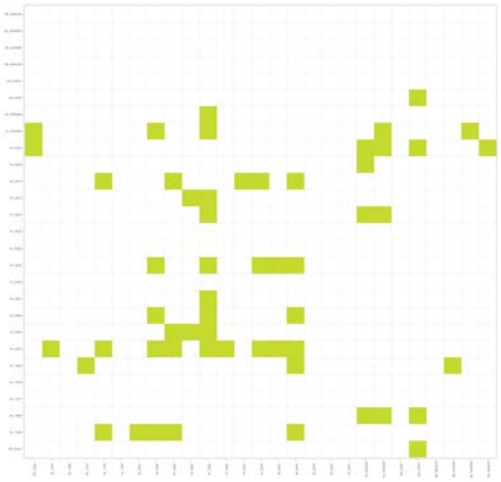
Figure W2: Competition Matrices Graphs (continued)

(g) Washington, D.C. - Group

(h) Amarillo, TX - Group

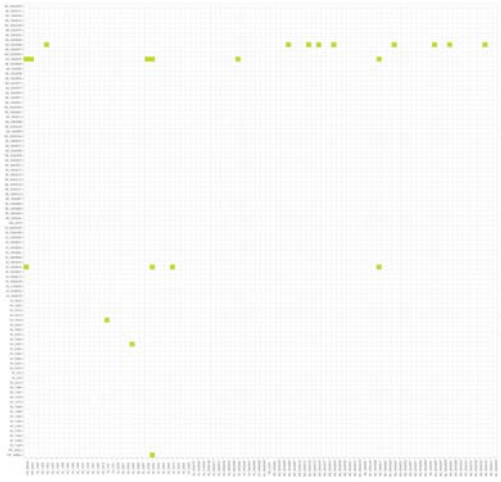


(i) Washington, D.C. - Wholesale



(j) Amarillo, TX - Wholesale

(k) Washington, D.C. - Employee/Loyalty-Redemption



(l) Amarillo, TX - Employee/Loyalty-Redemption