CESifo WORKING PAPERS

9127
2021

June 2021

Demand Estimation Using Managerial Responses to Automated Price Recommendations

Daniel Garcia, Juha Tolvanen, Alexander K. Wagner

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 9127*

Demand Estimation Using Managerial Responses to Automated Price Recommendations

Abstract

We provide a new framework to identify demand elasticities in markets where managers rely on algorithmic recommendations for price setting, and apply it to a dataset containing bookings for a sample of mid-sized hotels in Europe. Using non-binding algorithmic price recommendations and observed delay in price adjustments by decision makers, we demonstrate that a control-function approach, combined with state-of-the-art model selection techniques, can be used to isolate exogenous price variation and identify demand elasticities across hotel room types and over time. We confirm these elasticity estimates with a difference-in-differences approach that leverages the same delays in price adjustments by decision makers. However, the difference-in-differences estimates are more noisy and only yield consistent estimates if data is pooled across hotels. We then apply our control-function approach to two classic questions in the dynamic pricing literature: the evolution of price elasticity of demand over time as well as the effects of a transitory price change on future demand due to the presence of strategic buyers. Finally, we discuss how our empirical framework can be applied directly to other decision-making situations in which recommendation systems are used.

JEL-Codes: L130, L830, D120.

Keywords: big data, causal inference, machine learning, revenue management, price recommendations, demand estimation.

Daniel Garcia
Department of Economics
University of Vienna / Austria
daniel.garcia@univie.ac.at

Juha Tolvanen
Department of Economics
University of Vienna / Austria
juha.tolvanen@univie.ac.at

Alexander K. Wagner
VCEE
University of Vienna / Austria
alexander.k.wagner@univie.ac.at

June 2, 2021
Forthcoming in *Management Science*
The authors thank the corporate sponsor for making the data available for this project. We are particularly grateful to Hannes Ullrich, three anonymous referees, the department editor, and the associate editor for very useful comments. We also thank audiences at the University of Vienna, the 2018 CESifo Network Conference on the Economics of Digitization, and the 2019 Industrial Organization Ausschuss for comments. William Burton provided excellent research assistance. The computational results presented have been computed in part using the Vienna Scientific Cluster (VSC). Garcia gratefully acknowledges funding from FWF Single Project "Understanding Consumer Search".

1 Introduction

Revenue management provides a toolkit for sellers to optimally price goods with uncertain demand and fixed capacity. By combining rigorous statistical methods with business applications, the discipline has become a resounding success and is now a standard tool in multiple industries. In the classical framework, firms adjust prices over time in response to the arrival of new consumers with uncertain demand, taking into account the opportunity cost of selling an additional unit (Talluri and van Ryzin, 2004; Phillips, 2005). In addition to accurate demand forecasting, optimal pricing also requires precise knowledge of the marginal effect a (transitory) price change has on the final quantity sold of a given item. In this paper, we provide a novel approach to causally estimate price elasticities of demand that can be used in static as well as dynamic settings. The estimation approach is particularly well suited for environments which require controlling for a large number of potentially confounding variables. We demonstrate its strength in an application to dynamic hotel room pricing, using a dataset that contains all bookings, prices and algorithmic price recommendations for 9 European hotels for a period of 14 months. Our application is directly relevant for revenue management in the hospitality industry, one of the largest sectors in the worldwide economy.[1] The approach can also be applied to an array of decision problems that rely on recommendation systems outside our specific application.

Unlike standard ways of estimating demand, our method leverages demand shocks rather than cost shocks. This difference is important especially in the service and financial sectors in which demand shocks are frequent while cost shocks are rare and often hard to identify. Consequently, estimating demand is notoriously difficult in these industries. The problem is exacerbated by managers increasing prices in periods of high demand, inducing a positive correlation between prices and sales (Athey, 2017; Bajari et al., 2015). In the hotel industry,

[1] According to Eurostat statistics (https://ec.europa.eu/eurostat/), EU27 citizens spent in 2018 about 124 billion euros or 1.7% of their total consumption expenditures on accommodation services.

our main focus, this correlation has been widely documented using market-level and hotel-level data (e.g. Enz and Canina, 2010; Cho et al., 2018). Our approach offers an alternative to cost shock-based methods in situations where those cost shocks are not available and it produces plausible negative elasticities when traditional methods fail to do so. It can therefore be used in practice, allowing revenue managers to set prices more accurately, as well as by researchers who require accurate demand elasticities as parameters in their models of competition and industry-level welfare.

Our identification strategy rests on the ability to observe prices adjusting with delay to demand shocks. Intuitively, holding demand constant on its new, post-shock level, the difference in sales before and after the price change informs us of the price elasticity of demand. Such variation is abundantly available in industries which rely on price recommendation systems for pricing decisions (cf. the accommodation, transportation, car rental, and other inventory management industries). We show in our application to hotel room pricing that hoteliers follow the algorithmic recommendations very closely but do so with a delay, thereby generating exactly the type of variation required by our estimation approach.

The above identification strategy suggests a simple difference-in-differences estimator which compares sales in a period after the recommended price changed but before the actual price change has been implemented to sales in a period following the actual price change. As a first step, we implement this simple method and show, using both real-world and simulated data, that it produces accurate demand estimates when there are enough purchasing events within a short interval around the time of the actual price change. However, the frequency of purchasing events in our sample is relatively low which leads to noisy estimates. We conclude that the difference-in-differences estimator is only appropriate for pricing homogeneous products that are purchased frequently and in relatively large quantities. In contrast, the estimates are too inaccurate for pricing hotel rooms whose demand varies by room type, location, and across seasons.

We remedy the low efficiency of the difference-in-differences estimator by applying a high dimensional control function approach for causal inference developed recently in the machine learning literature (Belloni et al., 2014). It allows us to use all of the relative variation on prices and price recommendations, and all of the purchasing events while flexibly controlling for a large number of potential omitted variables. This approach provides accurate estimates even at the level of a single room type for a single season. Furthermore, compared to structurally modeling demand and the response of hotel managers to recommendations, the method is less susceptible to functional misspecification because it allows for a high dimensional set of competing functional forms. The results from the control function approach agree with the difference-in-differences results on the pooled data which makes us also confident in its ability to measure price elasticities accurately at the room type level. We further demonstrate the robustness and accuracy of the results with simulations even for the case in which the model is misspecified.

Our approach is related to recent attempts to obtain causal inference on managerial decisions in various industries, such as ride-sharing (Cohen et al., 2016; Castillo, 2019), staffing (Mani et al., 2015) and loan markets (Costello et al., 2020); each of these papers exploits a specific institutional or technological feature which generates exogenous variation in the implemented policy. In contrast, our estimation approach uses a 'behavioral reaction' of managers that can be applied potentially in many other contexts in which decisions are based on algorithmic recommendations. For example, it could be used to estimate the expected marginal value of inventory in retail applications where an inventory management system recommends restocking times and quantities to a local store manager (Conlon and Mortimer, 2013; Cachon et al., 2019);[2] see Section 3.2 for more examples.

In our application, we extend our estimation framework and investigate two aspects of inter-temporal demand important for optimal pricing in dynamic settings. First, we estimate how much the demand elasticity for hotel rooms varies over the booking horizon. Note that if the demand elasticity was constant over different booking horizons, then revenue-maximizing prices would tend to decrease as the day of arrival approaches, because the opportunity cost of an unsold room is decreasing over time. Instead, our results reveal that prices are roughly constant, suggesting increasing markups. In particular, the price sensitivity (semi-elasticity) of consumers is highest for very early bookers and lowest for those buying between 21 and 50 days before the date of stay. Overall, our findings are consistent with the idea that consumers, who are more price sensitive, search early on. Our results also corroborate

[2] Exogenous, behavioral delays in the time it takes a manager to follow a recommendation to restock their wares can be used to estimate the marginal effect of an extra unit of stock on the probability of running out of stock. As long as stock-outs are relatively short and demand continuous over time, lost demand during a stock-out can be estimated using the number of units sold before and after the stock-out. Alternatively, one can also use sales data from similar stores which did not run out of stock.

evidence from airline tickets (Williams, 2018), and help reconciling the profile of constant or increasing prices with decreasing opportunity costs.

Second, we measure to what extent hotel customers delay their purchase when facing a higher price. To answer this question, we follow ideas in Li et al. (2014), who estimate strategic consumer behavior in the market for airline tickets, and include lagged prices to identify the effect of variation in past prices on current demand.[3] We show that past prices have lasting effects on current demand, but these effects vary across hotels. For one of the hotels in our sample, high past prices are likely to discourage searchers, reducing current demand, while for another hotel the pattern is reversed.

In summary, we provide in this paper a new reduced form regression approach to identify demand elasticities in markets with dynamic demand and a limited access to observable supply shocks. In an application to hotel-room pricing, we then study the variation in demand elasticity as the day of arrival approaches and the impact of strategic consumers on optimal pricing. An advantage of our approach is that it provides, in contrast to other methods, reliable estimates for relatively small sample sizes and circumvents the need for experimental variation (e.g. Moon et al., 2018; Nambiar et al., forthcoming) or structural modeling (e.g. Li et al., 2014; Cho et al., 2018). A key innovation of our approach is that it leverages recommended rates produced by an advanced pricing algorithm and the associated behavioral responses of managers to this kind of information.

1.1 Contribution to the Literature

Our contribution to the literature is twofold. First, we provide a simple yet powerful approach to estimate demand leveraging on recommendation data that is very often available to decision makers. The approach combines machine-learning algorithms with modern causal-inference methods (cf. Belloni et al., 2014, 2016; Chernozhukov et al., 2015). The only paper we are aware of that uses similar ideas for identifying demand is the contemporaneous work by Castillo (2019). We both use a reduced-form approach, isolating short-run price changes which are arguably uncorrelated with both demand and supply. Castillo (2019) exploits

[3] There are a number of important differences between hotel room and airline pricing with regards to strategic consumer behavior. Hotels for instance rarely use markdowns to price discriminate across consumers based on the time of the day or the day of the week in which consumers search. From an applied revenue management perspective, our empirical applications provides useful insights for optimal pricing in the hotel industry.

discontinuities in the specific price setting algorithm of a ride-hailing company to identify customers' willingness-to-pay. Instead, we use sluggish price adjustment by managers as the source of exogenous variation. The 'behavioral response' can be applied in a broad range of settings, from estimating the marginal impact of an additional worker in a retail store to the foregone profits due to stock-outs as discussed in Section 3.2.

Our second main contribution concerns an application to hotel-room pricing (cf. Section 5). Extending our estimation approach, we address two questions relevant for optimal dynamic pricing: how does the price-sensitivity of demand change over time and what is the effect of a price change on behavior of forward-looking consumers? To the best of our knowledge, only a handful of papers attempt to estimate demand in dynamic markets. These include Li et al. (2014), Lazarev (2013) and Williams (2018) who study the market for airline tickets; Joo et al. (forthcoming) in the holiday cruise industry; and Li et al. (2018) and Cho et al. (2018) who, like us, focus on dynamic pricing in the hotel industry. Most of them resort to structural modeling to estimate dynamic demand.[4]

In the first application, we recover time-varying demand elasticities for the biggest hotels in our sample. In a related work, Williams (2018) develops a structural model and identifies the consumer demand elasticity over time as well as the welfare implications of dynamic pricing by combining posted prices of airline tickets and flight availability data from aircraft seat maps. Lazarev (2013) uses a high-frequency data set on airfares to study the welfare effects of dynamic pricing in a structural model which also allows for individuals to learn about their travel needs by delaying purchases. They both find that earlier consumers are more price sensitive. An exception is the recent work by Joo et al. (forthcoming) who estimate demand for holiday cruises, using a cost shock for identification, and find that demand becomes *more elastic* as the departure date approaches.

The second application concerns the role of forward-looking consumers. Whereas the classical dynamic pricing literature assumed that consumers act myopically, there has been a growing interest in the theoretical literature on strategic, or forward-looking, behavior of consumers in dynamic settings. Our empirical results are hence not only of interest for

[4] Li et al. (2018) is the only paper, other than the present article, to obtain credible estimates of dynamic demand for hotels using reduced-form methods. Reduced-form methods are undoubtedly more useful for practitioners. Indeed, both Cho et al. (2018) and Li et al. (2014) acknowledge that they use structural methods because in their applications reduced-form estimates are likely to be contaminated by unobserved demand shocks.

practitioners but can also inform theoretical work on optimal dynamic pricing.[5] To our knowledge, there are very few papers that try to identify the presence of strategic consumers, most of them rely on structural identification using data from the airline industry. In terms of magnitudes, our results are broadly consistent with those of Li et al. (2014) and suggest that, for the biggest hotel in our sample, approximately one third of consumers is willing to delay their purchasing decisions in the hope of obtaining a better price in the future.

2 Data and Institutional Background

Our data set contains more than 5 million observations of hotel-room pricing information and the corresponding universe of about 60 thousand bookings, all aggregated at the daily level. This high-resolution proprietary data set was made available by our corporate sponsor, whose identity is withheld by request. The corporate sponsor is based in Europe and provides a number of revenue management services, including price recommendations, to a large number of independent hotels. The rate and booking data comes from 9 different hotels, eight located at resort destinations and one (hotel 6) located in an urban area. Our data contains bookings and prices for each hotel over a period of about 14 months. For each potential arrival date we observe the flow of bookings, actual rates charged by the hotel, and rates recommended for every room by the revenue management company. The actual rate can be thought of as an index price set by the hotel. It is is modified by pre-specified and channel-specific discounts or fees and is then pushed to each channel in which the product is offered. We also the actual rate in our analysis because both the revenue management system (RMS) and the hotelier use this variable as the main instrument for price optimization.[6] In addition, we observe for each booking the unique customer identifier, date of booking, date of arrival, length of stay,

[5] Various strategies have been proposed recently to accommodate strategic behavior by consumers in the theoretical literature. Some models study how frequent discounts (Cachon and Feldman, 2015) or randomized pricing policies (Chen et al., 2018) can counteract strategic behavior of customers. There is also a number of recent articles in operations research that complements the literature by providing novel robust approaches to dynamic pricing with strategic consumers (e.g. Besbes and Lobel, 2015; Caldentey et al., 2016; Chen and Farias, 2018).

[6] The total price paid can include, on top of channel-specific discounts, also promotional discounts (e.g. 'book x nights but pay only x-1') as well as extra charges (e.g. for room service or spa visits). As shown in Table 1, the average total price paid by customers is a relatively stable share of the average actual price of a product. For details on prominent distribution channels, specific discounts and contractual settings relevant in the hotel industry, see for example Hunold et al. (2018) who study the effects of best price clauses previously used by some online travel agencies (OTAs). Note also that rooms that are sold through long-term contracts between the hotel and a reseller, who may be free to set their own prices, are not included in data.

and the total price paid by a customer as can be seen from Table 1.

For our empirical analysis, we construct an unbalanced panel from the data. We define a product $i = 1, \ldots, N$ as a combination of an arrival date and a room type in a given hotel $h = 1, \ldots, 9$. Possible arrival dates span over at least 396 consecutive days for each hotel. Across all hotels, we have 75 room types which we often aggregate into four larger categories: single, standard, superior, and suite. For each product, we observe daily bookings, recommended and actual rates for all possible booking dates $t = 0, 1, \ldots, T_i$ with $T_i \leq 365$ being the maximum number of days before arrival in our sample for product i. If a customer buys multiple nights at a hotel, she buys, according to our definition of a product, multiple products from this hotel.

2.1 Bookings

In Table 1 we present summary statistics on bookings, actual rates set by the hotelier, and recommended rates from the revenue manager's pricing algorithm for each room category. Mean daily actual rates vary substantially across room categories, ranging from 82 euros for *single* rooms to 211 euros for *suites*. The most popular category is the *standard* room, both in terms of availability and occupancy. With regards to the duration of stay, approximately 50% of the bookings involve a single night but the mean length of stay exceeds three nights. Clearly, there is a lot of seasonal variation in the industry, especially at resort destinations. We observe demand peaks in the Winter and Summer holiday seasons and relatively low demand in Spring and Fall.

There is substantial dispersion in both the flow of bookings and the actual rates for the same room over time. Figure 1 shows the number of bookings as a function of days before arrival for the two largest hotels (ID 6 and 175). Most bookings occur within 30 days of arrival but there is a long right tail in each hotel. Given the booking pattern, it is not surprising that hotels change actual rates more often as the day of arrival approaches.[7] Table 2 summarizes the distribution of actual rate changes over the booking horizon of a product. Over the complete booking horizon, hoteliers change the actual rate of a product on average

[7] The majority of transactions of a hotel is realized through online travel agencies (OTAs) or through a hotel's website but we do not have information about the specific booking channel used in a transaction. As mentioned above, we do not consider prices for specific booking channels in our analysis to estimate demand because we have direct access to actual rates.

once a month. Because most bookings and price changes happen close to the arrival date, the median consumer can expect to see a price change every two weeks.[8] Conditional on the actual rate changing, the median absolute adjustment amounts to about 8 euros per night.

2.2 Price Recommendations

A crucial input to our analysis are recommended rates provided by the revenue management system (RMS). We are not aware of any other study that exploits price recommendations provided to managers to estimate demand. The RMS which produces recommended rates for the hotel manager works as follows. The revenue management company offers a RMS service with the objective to maximize the client's revenue through optimized pricing. It is incentivized to provide the best possible performance because the fee hotels pay for the service is benchmarked against expected revenue increases generated by the service. For price optimization, the revenue management company has access to all bookings from a hotel on a real-time basis, as well as additional demand-related information such as local variation in weather conditions, events, public holidays, hotel reputation, competitor prices, etc.

In addition, the revenue manager and the hoteliers are regularly exchanging information about local demand conditions. Clients are trained to transmit private information about local demand shocks to the revenue management company. The company combines all relevant information and feeds it into their proprietary pricing algorithm to produce a rate recommendation for each product of the hotel.[9] The hotelier decides every day whether to log into the system. If she does, she observes the current rate recommendation for each room and then decides which rates to update and by how much .[10] After manual confirmation by the hotelier, updated actual rates are pushed from the RMS to the hotel's property management

[8] This calculation assumes that the arrival of potential customers is proportionate to the number of bookings. In contrast to the setting in Li et al. (2014), in which strategic consumers can predict the direction of the price change of airline tickets, it is extremely difficult for the average potential customer to predict the direction or the magnitude of a change in the price of a product in our setting, without observing the recommendation. Rate updates are not serially correlated and they are fairly symmetrically distributed (see Figure 2).

[9] We do not have access to the company's proprietary pricing algorithm. A common concern in competition policy is that algorithmic pricing can be used to coordinate with competitors (Miklós-Thal and Tucker, 2019). We believe that coordination through algorithmic pricing is not much of a concern in our context. Given the highly fragmented nature of the industry, the number of hotels which use recommendations from our corporate sponsor is very small ($\leq 1\%$) compared to the size of the industry in the considered regions.

[10] In addition to recommended rates, the hotelier has also access to additional information on the RMS platform, such as prices of local competitors, own performance metrics, etc.

system and distribution channels (mostly OTAs).

Although the hotels in our data are typical for their respective regions, with about 50 rooms per hotel they are relatively small by international standards. Most of them are family run which means that the manager's job description includes many responsibilities beyond revenue management. From private communications with the revenue management company we learned that managing prices takes only a small fraction of a hotelier's weekly work schedule. In Section 3.4, we provide a detailed account of the behavioral response of hotel managers to rate recommendations. There, we show that hotel managers follow algorithmic recommendations quite closely, but do so with a delay. This will be an important ingredient of our identification approach.

3 Empirical Framework

A common challenge encountered in market studies is to find sources of exogenous variation in prices that allow to identify demand. In our setting, the statistical inference problem is further complicated because we are interested in price variation over time for the same physical object. An instrument, therefore, should vary not only across arrival dates (e.g. shocks to capacity) but also across booking dates for a given arrival date and these shocks should be unanticipated (e.g. news shocks about future costs). We circumvent this problem by leveraging rate recommendations, which allows us to isolate price changes that are plausibly uncorrelated with contemporaneous demand shocks.

3.1 Demand Estimation

In each booking period $t = 0, 1, \ldots, T$, with $t = 0$ representing the date of arrival, the manager at hotel h sets a price for product $i = 1, \ldots, N$ (e.g. a standard room for March 20) according to

$$p_{hit} = g^h(r_{hit}, X_{hit}, \omega_{hit}) + \nu_{hit}, \tag{1}$$

where r_{hit} is the price recommendation of the revenue manager, X_{hit} is a vector of observable characteristics, with a typical element X^j_{hit}, $j = 1, 2, \ldots, J$, ω_{hit} is the information observed by the hotelier but unobserved by the econometrician, and ν_{hit} is the price innovation. The characteristics X_{hit} can contain information that depends on both the booking period t and the arrival-date-room-type combination. Examples include available capacity at booking time t for room i and information about the seasonality of demand for hotel h.

The revenue manager forms recommendations according to

$$r_{hit} = R(X_{hit}, \Omega_{hit}), \tag{2}$$

where Ω_{hit} is information observed by the revenue manager but not by the econometrician. Key for our identification strategy is that the revenue manager receives all demand-related information that the hotel manager observes, in other words that ω_{hit} is measurable with respect to Ω_{hit}. Given some room characteristics X_{hit}, we assume that the recommendation is a sufficient statistic for the information held by the revenue manager Ω_{hit}. That is, $\Omega_{hit} \mapsto R(X_{hit}, \Omega_{hit})$ has a measurable inverse for all X_{hit}. Denote this inverse by $f^h(r_{hit}, X_{hit}) + \lambda_{ht} = \Omega_{hit}$, where λ_{ht} captures hotel and time (days aheads) specific effects on demand; this includes the arrival rate of potential consumers to hotel h as a function of days before arrival t which we assume to be known to the revenue manger.

We assume that the number of bookings for a given product i that occur t days before arrival can be written as

$$Q_{hit} = \lambda_{ht} + p_{hit}\eta^h + X'_{hit}\beta^h + \tilde{\Omega}_{hit} + \epsilon_{hit}, \tag{3}$$

where η^h is the parameter of interest, $\tilde{\Omega}_{hit} = \Omega_{hit} - \lambda_{ht}$, and ϵ_{hit} is an error term.[11] Given the structure of the model, the two assumptions that guarantee identification of η^h are then the following.

Asumption 1 (Exogeneity)

$$\mathbb{E}\left[\epsilon_{hit} \mid X_{hit}, \Omega_{hit}\right] = 0$$

Asumption 2 (Conditional Independence)

$$\mathbb{E}\left[\epsilon_{hit}\nu_{hit} \mid X_{hit}, r_{hit}\right] = 0$$

As shown below, we do not need to include the price in the exogeneity assumption because it is determined by the recommendation and observables. The second assumption is guaranteed, for instance, if ν_{hit} captures only variations in the availability of the hotelier to

[11] In our empirical analysis, see Section 4, we also include a model with a multiplicative structure.

make changes in actual rates and local cost shocks, but does not include information about demand shocks. For this to hold it is sufficient that any information that the hotelier obtains about demand is incorporated in the revenue management system.[12] It then follows that we can directly estimate the parameter η^h from

$$Q_{hit} = \lambda_{ht} + p_{hit}\eta^h + X'_{hit}\beta^h + f^h(r_{hit}, X_{hit}) + \epsilon_{hit}, \tag{4}$$

as now

$$\begin{aligned} \mathbb{E}\left[p_{hit}\epsilon_{hit}\right] &= \mathbb{E}\left[\mathbb{E}\left[p_{hit}\epsilon_{hit} \mid r_{hit}, X_{hit}\right]\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[(g^h(r_{hit}, X_{hit}, \omega_{hit}) + \nu_{hit})\epsilon_{hit} \mid r_{hit}, X_{hit}\right]\right] = 0. \end{aligned} \tag{5}$$

The last last equality holds, because of the two identifying assumptions, the fact that $g^h(r_{hit}, X_{hit}, \omega_{hit})$ is measurable with respect to (r_{hit}, X_{hit}) and the identical equality $(\Omega_{hit}, X_{hit}) \equiv (f^h(r_{hit}, X_{hit}), X_{hit})$ where f^h is measurable.

In the empirical implementation of our framework in Section 4, we allow $f^h(r_{hit}, X_{hit})$ to vary across different room-type-arrival month combinations, days before arrival t, and also across weekdays in which hoteliers update rates. The latter variable allows us to pick up manager-specific variation, e.g. arising from her schedule at work.[13] Finally, notice that η^h in eq. 4 can be interpreted as the marginal impact of the price on the probability that a booking occurs on a given date. It can be thought of as an estimate of the derivative of the residual demand curve of a given product (hotel-room-type-arrival-date combination), which contains information about the responsiveness of consumers to a price change as well as an implicit conduct parameter measuring market competitiveness. In Appendix C, we extend our framework and study heterogeneity in price elasticities by allowing η^h to vary across different dimensions such as booking weekdays and seasons of arrival.

3.2 Generality of empirical framework

The above estimation approach is neither restricted to the particular institutional setting nor the pricing application we study empirically in this paper. It can be adapted to a variety of

[12] See Section 3.4 for a more detailed discussion of the validity of the identifying assumptions.

[13] An alternative specification would be a control function approach, using a two-step estimation method (price on recommendation and controls and quantity on these residuals). In our case, however, this is not necessary because we can confidently rule out any direct effect of the recommendation on demand.

setups that use recommendation systems, or decision support systems (DSS), more generally. The fundamental idea is to use plausibly exogenous variation in how recommendations are implemented behaviorally by decision makers to identify the marginal impact of the choice variable (e.g. prices) on the outcome variable (e.g. sales). In the following, we present two examples that fit well in our framework.

Staffing. A number of papers have estimated the effect an additional worker has on profits by constructing a predictive model of staffing behavior and then exploiting observed departures from it as the exogenous variation in staff levels (e.g. Mani et al., 2015; Fisher et al., 2018). While such methods may work well in some cases, it is generally difficult to assess the extent to which the econometrician's model is misspecified and the observed departures truly exogenous. Nevertheless, many organizations adjust their workforce at different locations based on their *own* demand predictions. This adjustment is rarely perfect because managerial inattention and worker availability can generate deviations from the ideal level of staff given the predicted demand. Our framework can be directly applied to identify the marginal effect of an additional worker on sales if the realized number of workers does not perfectly match the firm's own prediction of the optimal staffing level. To be more specific, let p_{it} denote the staff level at location i at time t, and let the rest of the variables in eq. 1 be as before, with r_{it} denoting the ideal level of staff given the predicted demand. Provided that there is some exogenous variation how the realized level of staff is determined, conditional on the ideal level, one may obtain a causal estimate of the marginal impact of a worker on sales by estimating the analogue of eq. 4.

Inventory management. One of the areas in which DSS have been employed heavily is inventory management (e.g. Shang et al., 2008; Van Donselaar et al., 2010). As above, let r_{it} denote the restocking recommendation from the DSS and p_{it} its realized flow. Our framework can be used to estimate the marginal effect of an additional stocked unit on sales. Holding excess inventory is costly and therefore the causal impact of stock on revenue is a critical component of optimal inventory management.[14] This can be directly identified using the model described in eq. 4.

Finally, it is also illustrative to consider the limits of our framework. For example, consider a loan officer who has to assess the creditworthiness of a client with the help of a DSS. Suppose

[14] See Conlon and Mortimer (2013) and Cachon et al. (2019) for recent work on the same question.

a researcher has data on these decisions and wants to to estimate the marginal return of funds allocated to credit. In case the officer's decision differs from the DSS's recommendation because her decision is based on additional soft information about the client (Liberti and Mian, 2008), rather than being a by-product of managerial inattention, then our critical assumption is violated and our approach cannot be used. For this reason, we provide in Section 3.4 a wealth of evidence suggesting that managerial inattention is the main source of variation in our hotel sample.

3.3 Machine Learning and Model Selection

The recent literature on statistical inference with machine learning has shown that proper model selection can have important effects on estimated demand elasticities (e.g. Chernozhukov et al., 2015; Dube et al., forthcoming; Athey and Imbens, 2019). Applied to our setting, these insights mean that correctly selecting f^h can be highly important for guaranteeing unbiased estimates of η^h. Especially, if the revenue manager is not perfectly informed about the hotelier's objective function, it becomes critical to correctly control for possible confounds. Even conditional on the revenue manager's recommendations there may remain demand-related confounds that are correlated with both the price and frequency of bookings. For example, a combination of fixed costs and credit constraints may lead the hotel manager to price more aggressively during off-season to guarantee enough bookings to remain solvent. Consequently, the relationship between the actual rate and the recommended rate could vary in a way that is correlated with the intensity of demand, violating our identifying assumption.[15]

Furthermore, the response of the hotel manager to the price recommendation may depend both on its magnitude as well as its timing. Allowing for flexible controls for weekday-month combinations, the days before arrival, a third-order polynomial of the recommended prices and its interactions with both the weekday and arrival month dummies already implies more than 500 potential parameters. When the median number of bookings in a hotel is less than 7500, informative inference on the effect of price on demand while estimating all of these parameters is challenging. We solve this problem by applying the "post-double-selection"

[15] Using a simulated dataset, we provide in Appendix B evidence that the estimates from a simple linear model are severely biased.

method from Belloni et al. (2014) to select the approximately right set of controls that leads to consistent estimates of the price elasticity of demand. We demonstrate in Appendix B that the nuisance parameter problem is indeed binding for simulated datasets comparable to the ones used in our application; in particular, estimates of the price semi-elasticity of the linear model are shown to be downward biased.

For the application of the post-double-selection method, and with a slight abuse of notation, let J, with a typical element j, be the index set of possible controls, $\tilde{X}^j_{hit}$, that includes a flexible polynomial of the recommended price and its interactions with weekday and month-of-arrival dummies. As a first stage, we use a standard LASSO to estimate

$$p_{hit} = \tilde{X}'_{hit}\alpha^h_1 + \xi_{hit}, \tag{6}$$

where ξ_{hit} is the error term. Denote the set of controls X^j_{hit} selected by the LASSO by $J_{h1} \subset J$. Intuitively, these controls are correlated with the price per night and thus are potentially important confounding factors. We then run another LASSO to estimate

$$Q_{hit} = \tilde{X}'_{hit}\alpha^h_2 + \nu_{hit}, \tag{7}$$

where ν_{hit} is the error term. For this equation the LASSO picks a potentially different set of controls $J_{h2} \subset J$. This estimation stage selects covariates that are important in the equation of interest and hence helps to reduce the residual variance as well as potentially captures confounds that are strongly correlated with bookings and only weakly correlated with the price.

The post-double-selection estimator $\hat{\eta}^h$ for hotel h is then given by the least squares estimator obtained by regressing Q_{hit} on p_{hit} and the control terms selected in the previous stages, X^j_{hit}, with $j \in J_{h1} \cup J_{h2}$:

$$(\hat{\eta}^h, \hat{\gamma}^h) = \underset{\eta^h\in\mathbb{R},\, \gamma^h\in\mathbb{R}^p}{\text{argmin}} \left\{ \frac{\sum_{t=0}^{T}\sum_{i=1}^{N}\left(Q_{hit} - p_{hit}\eta^h - \tilde{X}'_{hit}\gamma^h\right)^2}{N+T+1} : \gamma^h_j = 0, \forall j \notin J_{h1} \cup J_{h2} \right\}.$$

The vector γ^h contains the regularized regression coefficients for the control variables. Belloni et al. (2014) show the consistency and asymptotic normality of this estimator and also demonstrate that the standard errors for the last stage yield asymptotically correct confidence sets. They also provide an algorithm for estimating the penalty loadings for the two LASSOs in the selection stages.[16]

[16] In terms of computational resources, the complete estimation procedure took more than 8000 CPU hours

3.4 Validity of the Identifying Assumptions

In addition to controlling correctly for f^h, identification relies on two behavioral assumptions on the relationship between the hotel manager and revenue manager. First, we assume that the price recommendation is a sufficient statistic for the information held by the revenue manager. Second, we assume that the variation in actual rates, conditional on observable product characteristics and recommended rates, is due to factors orthogonal to variation in demand. In addition to the information about the institutional relationship between the revenue manager and hotelier presented in Section 2, we present in this subsection evidence in support of these two identifying assumptions. To do so, we perform a deeper analysis of the recommendation data from the pricing algorithm as well as the behavior of the hotelier when making pricing decisions.

The first assumption can be recast as a communication game between the revenue manager and the hotel manager. In this game, the recommendation sent by the revenue manager is the message and the actual rate is the action of the hotelier. If the incentives of the revenue manager are sufficiently aligned with those of the hotel manager, a truth-telling equilibrium, in which the recommendation is the best predictor for the current demand shock, exists.[17] More realistically, however, incentives of the two contracting parties may not be perfectly aligned. In comparison to the hotelier, the revenue manager may for instance weigh revenue more heavily than profits and may not face negative consequences from excessive price variation. If this was the case, it would suffice that the revenue manager would suffer an additional reputational or moral cost if she distorted her recommendation excessively (see e.g. Kartik, 2009). This would allow for a monotone signaling equilibrium and the invertibility assumption would still hold. In Appendix B, we report on a simulated model in which recommendations are systematically biased; our estimates remain valid despite the bias.

To further evaluate the role of price recommendations in the formation of actual rates, we construct a large LASSO prediction model, which does not include recommended rates, to predict the actual rates. We then compare the residuals from this model to (i) the same LASSO prediction model which includes recommended rates as a predictor and (ii) to the raw

for the data set including all 9 hotels (using 9 supercomputing nodes each containing 256GB of memory and 2 INTEL Xeon processors).

[17] For instance, in Crawford and Sobel (1982), as the bias of the sender converges to zero, the most informative equilibrium becomes perfectly informative.

difference between the actual rates and the recommended rates. The resulting distributions of residuals from the three models are shown in Figure 3. Although, observable characteristics clearly help in predicting the actual rate, the recommended rate alone does a much better job than our large LASSO model *without* recommended rates. The figure also shows that the LASSO predictions improve considerably when the recommended rate is added as a control. This improvement implies that the hotelier finds revenue manager's recommendations highly relevant when setting prices. Furthermore, the high frequency at which the price recommendation matches the actual rate demonstrates clearly that a hotel manager often copies the price recommended by the revenue manager perfectly. This strongly indicates that a hotel manager often follows the revenue manager's advice.[18]

Regarding the second identifying assumption, we show that the residual variation in the difference between actual and recommended rates is unlikely to be related to variation in demand. To do so, we first consider the variation in the frequency of updates on different weekdays in both the actual rate and the recommended one. Figure 4 plots the distribution of rate updates in both the actual rate and the recommended rate for every day of the week in the two biggest hotels in our data (ID 6 and 175). The distribution of updates in the actual rate has clear gaps that are likely related to the hotel manager's work flow. For example, the manager of hotel 6 probably works on other things than pricing on Thursdays and Sundays and updates most frequently on Tuesdays and Saturdays. The recommendation updates are much more evenly distributed over the week and their frequency does not seem to be related to the peaks in the hotelier's distribution of rate updates. Similar patterns can be seen for all hotels in Table 3.

Note that updating the prices for all arrival dates within the next two months in the median hotel requires to manually update 300 different rates. It is hence reasonable for a hotel manager to economize on this dimension by concentrating the workload on certain weekdays. Observed updating behavior is also in line with the work schedule and responsibilities of hotel managers in our data given in Section 2 (see also Footnote 18).

[18] A cursory inspection of Figure 3 might lead to the conclusion that the recommended rate is doing a better job at predicting the actual rate than the LASSO that includes the recommended rate. This, of course, is not possible. The LASSO prediction has much less mass on the absolute errors in the 15-30 euro range. In terms of mean squared error, these mistakes are much more expensive than the errors in the 5-10 euro ranges where the LASSO has considerably more mass than the simple recommended rate. Notice, however, that the LASSO that does *not* include the recommendation has more mass on all but the smallest errors than either of the two other predictions and hence performs considerably worse than them.

The timing and magnitude of deviations between the actual and recommended rates, conditional on an update, can be rationalized by standard models of price adjustment (see e.g. Reis, 2006; Alvarez et al., 2011). In these models, managers have only partial information about demand and incur some observation cost for reassessing the optimal price. Managers trade off the benefits of a more precise price with these observation costs and chooses to pay these costs only when the average expected update is sufficiently high. From this it follows that the realized average update will be roughly independent of time. The posited behavior matches exactly the pattern we see in our data. The scatter plot in Figure 8 shows the magnitude of all rate updates as a function of time since the last update. The locally smooth best predictor of the magnitude of the price update as a function of days since the last update is roughly constant, and is the same for price increases and decreases and across different hotels. Similarly, consistent with models with observation costs, the correlation between recommendations and prices is fairly constant over the booking horizon.[19]

Price-setting heuristics of hotel managers. As a final piece of evidence, we show that part of the variation in prices is consistent with satisficing behavior by managers (Simon, 1956).[20] Hotel managers often resort to behavioral heuristics in room pricing, because keeping a large number of prices up to date is a perpetual and daunting task. A particularly common behavior of hotel managers observed in our data is to accept all rate recommendations for a given room in a given week, i.e. implement the recommended rate in every single night of the week. With approximately 22,000 such events in our data, they account for 52.2% of all rate updates by hotel managers. It is unlikely that these events are driven by private information of a hotel manager. Indeed, the data shows that on days in which we observe a hotel manager 'accepting all' recommendations for a room for a given arrival week, we typically also observe dozens of other unrelated rate updates, often for completely different arrival dates. This can be seen in Figure 6, where we plot the total number of actual rate changes (in log scale) on booking days when the hotelier copies all recommendations for at least one whole week, and on booking days when none of the arrival weeks fully match the recommendation. It can

[19] The idea of fixed costs for adjustment as well as the use of heuristics in price setting have a long tradition in economics and management.

[20] Simon (1962, p.10) argued that "...price setting involves an enormous burden of information gathering and computation that precludes the use of any but simple rules of thumb as guiding principles. Through detailed study of pricing by multiple interviews throughout the firm we begin to get a picture of the informational and computational constraints that hedge in the pricing process and give it form.'

be seen from Figure 6 that hoteliers adjust considerably more rates on days in which they copy recommendations. This indicates that following algorithmic recommendations is much less costly for the hotelier than coming up with a large number of optimal prices herself. We will present a number of difference-in-difference estimations in Section 4.1, which leverage the timing of these events, to obtain a credible quasi-experimental benchmark for our main specification.

To illustrate the weekly updating pattern of actual rates by hotel managers, we first run a simple regression (see column 1 in Table 5) which predicts changes in actual rates by changes in recommended rates, and hotel-room type and arrival month fixed effects. We group all observations from booking dates that share the same room type and the same arrival week; we refer to these observations as 'peer' products. We then regress the residuals from the first regression on the average change in actual rates of other peer products as well as the average change in prices of non-peers with the same booking date. One would expect any private information about demand that is not already incorporated in the recommendation in the first stage to be correlated with demand for other products and hence induce a strong correlation between the non-peer prices and the residual from the first stage. According to our results in column 2 in Table 5, a 1 euro increase in the prices of all peer products is associated to a 0.96 euro increase in the price of a given product, while a 1 euro increase in the prices of non-peer products is associated with an increase of only 4 cents in the price of that same product. The adjusted R^2 in the residual regression is 0.96. Together, these results imply that almost all of the residual variation in the actual rate is captured by the *average* actual rate for that calendar week and room type. Hence, hotel managers update mostly rates for one calendar week at a time for a given room type using similar adjustment heuristics for all arrival days in that week. At the same time, spillovers across arrival weeks and room types are marginal at best.

4 Empirical Results

Based on our demand estimation framework, we now report results from various specifications of the following model

$$Q_{hit} = \lambda_{ht} + p_{hit}\eta^h + X'_{hit}\beta^h + f^h(r_{hit}, X_{hit}) + \epsilon_{hit}. \tag{8}$$

In this model, the dependent variable Q_{hit} equals one if there is a booking for a room-type-arrival-date combination, or product, i (e.g. standard room for March 21st) at hotel h, t days before arrival. The average probability of a booking on a given day is 2.4%, and only very rarely do we observe multiple customers purchasing the same room-type-arrival-date combination on the same booking date. The main regressors are the actual rate, or price, p_{hit}, and the rate recommended by the revenue manager r_{hit}. We allow the recommended rate to have a non-linear impact on the actual rate and allow for different hotels to adjust rates differently. We allow the LASSO to pick from a set of controls that includes month- and weekday-of-arrival fixed effects, their interactions with a third-order polynomial of the recommended rate, booking horizon fixed effects and all interactions between the month of arrival and the weekday of arrival. The strength of the double selection method is that it allows for highly flexible forms of the potentially infinite dimensional 'nuisance parameter' f^h with finite amount of data while still delivering consistent estimates of η^h.

We start by estimating the relationship between actual rates and daily bookings using a double selection (DS) procedure in which recommended rates are completely omitted. The results are summarized in column 2 in Table 6 where we report the seventh (L), fifth (M) and third (H) highest semi-elasticities out of our nine hotels. Omitting recommended rates yields often *positive* or *very small* coefficients of price on quantity, suggesting endogeneity despite the large number of seasonal controls. Even those hotels for which we would estimate negative price elasticities would be able to charge markups ranging from 100-200% of their price, i.e. implying prices inconsistent with profit maximization. We then consider a simple fully linear specification without double selection, where we include only actual rates, recommended rates, and room-type fixed effects on the right-hand side. The results can be found in column 1 of Table 6. More hotels exhibit now a negative relation between price and quantity but there is still huge dispersion across hotels.[21]

Next, we run a double selection regression in which we include the same large set of potential controls as in column 2 of Table 6, including a flexible polynomial of the recommended rate and interactions between seasonal characteristics and recommendations. These interactions capture the (possibly) non-linear response of hotels to information and are consistent with the model described in the previous section. Results from this (preferred) specification

[21] This is consistent with the results from the simulated model, where OLS estimates are biased in the presence of sufficiently many confounders, see Appendix B.

are summarized in column 3.[22] The median hotel faces a 1% drop in bookings should they increase their price by 3 euros. Translated into markups, these estimates suggest that the median hotel's variable profit from an additional purchase is approximately 55% of its price.[23] This median masks substantial variation across hotels. For example, hotels 175 and 6 are facing much more elastic residual demand functions, as we show below. Similar results are obtained using both a log-log and a logit specification (see columns 4 and 5 in Table 6). The first allows for the baseline arrival rate to affect demand multiplicatively rather than additively, while the second appropriately handles the large number of days with zero bookings in our data set.

Leveraging the richness of our data, we can also compute hotel-room-type specific semi-elasticites. For each hotel, we estimate the following model

$$Q_{hit} = \lambda_{ht} + \sum_{\rho} p_{hit}\eta^{\rho}\mathbb{1}_{i\in\rho} + f^{h}(r_{hit}, X_{hit}) + X'_{hit}\beta^{h} + \epsilon_{hit}, \tag{9}$$

with index ρ denoting the room-type associated with room i and $\mathbb{1}$ is the indicator function. Figure 7 plots the distribution of estimated coefficients, weighted by the booking rate of each room type. As expected, most coefficients are negative and statistically significant. The median coefficient is quite close to the median of the distribution of semi-elasticities across all hotels. Taken together, these results imply markups of around 50% of the price, and hence significant market power for the median room type. Based on these estimates, we can also perform a back-of-the-envelope calculation for the potential losses associated with sub-optimal price setting. The average hotel in our data has some 2,000 active rates at every point in time. If each of these rates are off by 7 euros (approximately 5% of the average rate), the daily loss in profits from managerial inattention would add up to 40 euros, hardly justifying the required investment in time and effort.[24]

[22] We experimented with a version of this model that includes an even larger set of potential fixed effects and that is estimated using the method suggested in Kock and Tang (2019). The results are quantitatively comparable and available from the authors on request.

[23] The hotel's variable profits are given by $(p-c)q(p)$, where p is the price, q is the quantity sold as a function of price, and c is the marginal cost. The first-order condition for optimizing this with respect to the price can be written as $(p-c)\psi = -1$, where $\psi = \frac{1}{q}\cdot\frac{\mathrm{d}q(p)}{\mathrm{d}p}$ is the semi-elasticity of demand which according to our preferred model equals approximately -0.0038. The average expenditure, p, is approximately 477 euros. These computations disregard cross-price effects which multi-product firms such as these hotels may want to take into account in their pricing; see Section 6 and D for more details.

[24] This number is computed by simply performing the Taylor approximation of the profit function $(p-c)Q(p)$ around the optimal price p^*. Using our linear demand estimates, the average loss per room is $\Delta\Pi \approx$

For more advanced dynamic pricing features, we can also investigate heterogeneity in price elasticities beyond the hotel-room type. As an illustration, we provide in Appendix C price elasticity estimates for different bookings days of the week as well arrival seasons which may both depend on the type of consumer, e.g. holiday or business traveler, a hotel faces.

4.1 Difference-in-Differences Estimations Using Accept-All Events

An alternative, more direct approach to exploit the variation in our data is to run a simple difference-in-differences estimation. To this end, we restrict our analysis to 'accept-all' events, that is, events in which a hotelier updates all rates for a given week and room type by following the recommendation perfectly. We restrict our data to 2-day windows around these events observing the prices and demand before and after the event. As the hotelier is exactly matching the recommended rate for a whole week of arrival dates, these events are unlikely to be driven by new information, and should rather be interpreted as a useful behavioral heuristic busy managers fall back on when updating rates. In particular, our identification assumption is that the exact timing of the actual rate update is uncorrelated with contemporaneous, hotel-specific demand shocks within the 2-day window. Because the recommendation was already in place on the first day, it is natural to assume that, had the manager decided to make a change for this particular week the day before, she would have chosen the same actual rate. In other words, the difference between the two rates was due to managerial inattention. This arguably quasi-experimental variation allows us to apply a difference-in-differences logic comparing demand and actual rate changes over those two days, using similar rooms in other hotels without an 'accept-all' event on that same day for the same arrival week as the control group.

Denote by Δ the first-difference operator on a given variable, such that $\Delta Q_{hit} = Q_{hit} - Q_{hit-1}$. Let δ_{hit} be a dummy variable that takes the value 1 if hotel h has an 'accept-all' event on date t for room i. Our regression equation is

$$\Delta Q_{hit} = \eta \delta_{hit} \Delta p_{hit} + \theta \Delta p_{hit} + \beta \delta_{hit} + \mu_{it} + \epsilon_{hit}, \tag{10}$$

where μ_{it} is a room-category, arrival-date combination fixed effect. The marginal effect of a 1 euro increase in the actual rate on the probability of observing a booking is captured by

$\frac{1}{2}(\Delta p)^2 \Pi''(p)$. Since $\Pi''(p) = 2Q'(p) \approx 2 \times 10^{-4}$, multiplying by the number of rooms yields a loss of about 40 euros.

$\eta + \theta$.

Table 7 summarizes the results. The estimated coefficient for the change in price across consecutive days, θ, is very small and statistically insignificant. In particular, the coefficient corresponds to an implausibly low semi-elasticity of -0.045 p.p., and is very precisely estimated (s.e. 0.05). That is, we find no significant correlation between changes in actual rates and contemporaneous quantity changes at the product level for those price changes that do not exactly match the recommendation. Instead, we find a sizeable and statistically significant effect of the interaction term: the estimated coefficient of the difference-in-differences equation, η, translates into a semi-elasticity of -1.2 p.p., and we cannot reject the null hypothesis that the estimated coefficient equals the estimated semi-elasticity for the median room (see Figure 7).

The treatment group for the difference-in-differences regression above requires that the hotel manager updates rates for a full week of arrival dates for a given room type *and* matches the recommended price while doing so. A natural question here is whether the event of updating a full week of prices for the same room type will introduce enough exogenous variation in the prices in form of pricing mistakes and delays to yield reasonable estimates even if price recommendations are unavailable. When we run the same model as in the first column of Table 7 but do not require that all prices in the treatment group match the recommendation after the price change, the estimated elasticity becomes positive (the estimated coefficient is 0.045 with a standard error equal to 0.025). We conclude that a sizable number of simultaneous prices changes, without controlling for the recommended prices, is *not* enough to control for the endogeneity resulting from hotel manager's information about demand shocks.

The lack of power of the above difference-in-difference approach using accept-all events might seem surprising at first, especially since there are 819 hotel-booking day pairs on which the hotel manager copied the recommendation for a whole week. However, given the low average booking rate, very few of the day pairs around the copying event have contemporaneous bookings, explaining the low power (see also simulation results in B). As the variation in prices in this model is very likely to be exogenous, we think that the results presented here serve as an excellent test for whether the estimates from the double selected model in the previous section are close to the truth.

The direct applicability of this difference-in-differences approach in real-world applications is however quite limited as the method requires data sets with very large numbers of *bookings* and cannot be applied at the hotel-level at least for the medium-sized hotels in our sample (because our hotel-level estimates exhibit considerable heterogeneity). Furthermore, the difference-in-differences approach provides only a local average treatment effect around the days the hotel manager copied recommendations. If the sporadic updating behavior is driven by adjustment costs, hotel managers are adjusting prices more frequently when the payoff from doing so is high. Consequently, such estimates are going to be applicable only in times of relatively high demand. Lastly, the difference-in-difference model requires panel data while our control-function approach can be applied to cross-sectional data as well.

4.2 Why Standard Instrumental Variables Do Not Work

The classical approach to demand estimation involves the use of instrumental variables. Roughly speaking, three types of instruments have been used in the literature: cost shifters, characteristics of other products and prices in other markets (Hausman instruments). In the hotel industry, there are no obvious cost shifters that move frequently enough to help identify demand when firms engage in dynamic pricing. Similarly, characteristics of other competing hotel rooms vary only over the very long run and, thus, cannot be used as instruments for these types of datasets. The only remaining potential candidate is to leverage prices in other markets to capture supply shocks. The key identifying assumption is that, conditional on product characteristics, the contemporaneous demand shock of hotels h and $\tilde{h}$ are uncorrelated if they are located in different markets. This assumption is likely to hold in markets where demand is driven by income shocks or taste parameters but perhaps less likely to hold in the hotel industry. In particular, weather shocks are strongly correlated across local markets.

To investigate the performance of this instrumental variable (IV) approach, we use the 'triple selection' IV algorithm from Chernozhukov et al. (2015) in which we include as instruments prices of hotels located in other regional markets. The algorithm allows us to select both the relevant exogenous controls as well as instruments. In particular, we estimate

$$p_{hit} = \sum_{\tilde{h}\in\mathcal{H}\setminus\{h\}} p_{\tilde{h}it}\gamma^h_{\tilde{h}} + X'_{hit}\gamma^h_x + u_{hit}, \tag{11}$$

$$Q_{hit} = p_{hit}\eta^h + X'_{hit}\beta^h + \epsilon_{hit}, \tag{12}$$

where $\mathcal{H}$ is the set of 9 hotels in our data and X_{hit} is the same set of controls as in the double selected regressions above. In other words, we use for each hotel all of the other hotels as instruments. In addition, we perform a similar exercise in which we use only hotel 6, which is located in a large city, as an instrument for prices of the other hotels which are located in resort destinations. This gives the best chance for the IV regression model to yield consistent estimates. We also estimate IV model specifications which include price recommendations as controls.

Table 8 shows that the above IV model yields mostly positive and highly heterogeneous coefficients. This suggests that the instruments are endogenous or, put differently, that the identifying assumption of demand shocks for hotels in different markets does not hold. The endogeneity is so strong that the estimates are biased even when controlling for recommended rates.

One might argue that the main driver of correlation in demand shocks across hotels in our application is weather. Since weather forecasts are not very reliable more than 15 days ahead, excluding the last 15 days before arrival would remove any endogeneity resulting from common weather shocks. Results from the IV model, excluding the last 15 days before arrival, are reported in the last two columns of Table 8. Only if we include recommended rates as a control, the IV regression is able to identify a negative price coefficient (see column 4). This illustrates that price recommendations contain information about common demand shocks which is not related to only weather across hotel locations. Overall, we conclude that Hausman-type instruments are unlikely to provide credible estimates due to unobserved and correlated demand shocks in the hotel industry.

5 Applications: Dynamic Components of Demand

We now demonstrate our empirical framework by investigating two essential features of demand in the hospitality industry. First, we perform a heterogeneity analysis by studying variation in price sensitivity of consumers across different time horizons which is crucial for inter-temporal price discrimination. Second, we introduce lagged prices in our main regression framework to investigate whether, and to which extent, current demand is affected by previous price changes. The idea is that if some consumers would be willing to delay their purchase in the hope of a better price in the future, lagged prices would be positively correlated with current demand.

5.1 Time-Varying Price Elasticities

Williams (2018) shows in an application of dynamic pricing in the airline industry that the benefits of inter-temporal price discrimination depend crucially on the temporal pattern of consumers' price sensitivity. Since the average opportunity cost (conditional on not having sold the room) is decreasing over time, the optimal price should fall as long as the semi-elasticity of demand remains constant. To investigate inter-temporal price sensitivity of customers for our hotel data set, we run our double-selected baseline specification for two hotels (6 and 175) separately for different booking horizons.[25] Results suggest that the estimated price semi-elasticity is highest for early bookings and lowest for late bookings as shown in Table 10. Our results are consistent with the intuition that consumers who search early on are relatively more price sensitive. This finding contrasts with recent work by Joo et al. (forthcoming) who find that the price sensitivity increases over time for holiday cruises.

We can also use these results to decompose the observed price into the markup and the opportunity cost (containing both marginal operating costs and foregone future opportunities), disregarding potential cross-substitution patterns across products and forward-looking consumers. This decomposition is depicted in Figure 9. Prices in hotel 6 are remarkably constant over time, while for hotel 175 prices are increasing slightly with the arrival day approaching. Because opportunity costs fall over time, this pattern is prima facie inconsistent with profit maximization. Constant prices can only be rationalized if the hotel 6 faces a decreasing price-elasticity over time, which is what we find: our estimated price elasticity is decreasing and the implied optimal markups are increasing, see blue dashed line in Figure 9. This suggests that properly accounting for demand heterogeneity, such as temporal differences in demand of consumers, is key for dynamic pricing policies in our setting.

5.2 Forward-Looking Consumers

Suppose that some consumers are willing to delay their purchasing decision and instead continue searching for a better price. We refer to these consumers as strategic, whereas the rest of consumers are assumed to be myopic. Strategic consumers, who delay their purchase waiting for better future prices, introduce a dynamic component in demand. Previous prices

[25] Specifically, we distinguish between very early bookings (>100 days ahead), early bookings ($51-100$ days ahead), late bookings ($21-50$ days ahead) and very late bookings (≤ 20 days ahead).

now have an impact on the probability that a strategic consumer who considered purchasing in the past remains active in the market.

More precisely, a price hike in the past has two distinct effects on the current demand of strategic consumers as we show in the Electronic Companion (section A). First, an increase in the price leads some consumers to strategically delay purchases in the hope of obtaining a better price in the future. Second, some consumers who would have otherwise waited will exit the market in response to a price hike. The second effect is relatively important in our setting because prices seem to follow a martingale, in the sense that the best predictor for tomorrow's price is today's price (unless one has access to price recommendations).[26] We refer to the sum of these two opposing mechanisms as the 'market-size effect' of a price hike. Empirically, we consider the following model:

$$Q_{hit} = \lambda_{ht} + p_{hit}\eta^h + p_{hit'}\eta^h_\tau + f^h(r_{hit}, X_{hit}) + X'_{hit}\beta^h + \epsilon_{hit}, \tag{13}$$

where $t' = t + \tau$ for some $\tau > 0$. The coefficients of interest are η^h and η^h_τ, which capture the contemporaneous and dynamic demand elasticity respectively. The identifying assumption is now that $(\Omega_{hit'}, \varepsilon_{hit'})$ for $t' > t$ are conditionally independent from ν_{hit} given (X_{hit}, r_{hit}). Provided that r_{hit} is a sufficient statistic for current *and* future demand, the assumption is satisfied.

Results from the double selection model for hotels 175 and 6 are summarized in column 2 of Table 9. We find that there is a persistent negative effect of lagged prices on contemporaneous demand for hotel 175, while this effect is positive for hotel 6. The magnitude of these effects is slightly bigger than a half of the effect of contemporaneous prices, suggesting that the proportion of customers willing to delay purchases is substantial for both. Taken at face value, these results suggest that there exists significant heterogeneity in the response of customers to price changes across different hotels in our sample.

In addition, price changes in the past are known to change the composition of potential customers (Li et al., 2014). When prices are increasing (decreasing), the fraction of consumers who wait for future price reductions is smaller (larger, respectively). Consequently, the price elasticity depends mostly on the behavior of myopic consumers when prices are increasing and

[26] Intuitively, this can be seen from Figure 2 which plots the distribution of changes in actual rates. Price hikes and markdowns are distributed similarly and the median change in either direction is 8 euros per night or 22 euros on an average booking; see also Figure 9.

on strategic consumers when they are decreasing. To quantify these effects, we run the model described in equation (8) interacting the actual rate with an indicator variable that equals one if the actual rates has increased and zero otherwise. Formally, restricting the sample to periods in which $\Delta := p_{hit} - p_{hit'} \neq 0$ for $t' = t + 1$, we estimate

$$Q_{hit} = p_{hit}\eta^h + p_{hit}\mathbb{1}_{\Delta>0}\eta_+^h + p_{hit}\mathbb{1}_{\Delta<0}\eta_-^h + X_{hit}'\beta^h + f_h(r_{hit}) + \theta_+\mathbb{1}_{\Delta>0} + \theta_-\mathbb{1}_{\Delta<0} + \lambda_{ht} + \epsilon_{hit}^D, \tag{14}$$

where θ_+ and θ_- are the regression coefficients associated with the level effects and λ_{ht} is a hotel specific arrival rate. The dynamics in our hotel-room demand estimates are generally consistent with findings by Li et al. (2014) in the air-travel industry. Results are reported in column 3 of Table 9. Again, these hotels present different patterns. In hotel 175, in periods of increasing prices, demand becomes slightly less elastic. In hotel 6, instead, we find that the responsiveness of demand to a price change is lowest when prices are falling, although the magnitude of this effect is also small. Coupled with the results above, they suggest that patient consumers may not be more price sensitive than their myopic counterparts in our sample. A possible explanation is that patient consumers have a higher willingness-to-pay for this specific hotel and simply optimize the timing of their purchase.

6 Conclusion

Recommendation systems that support pricing, inventory and production decisions are becoming ubiquitous with the rise of machine learning and artificial intelligence. In this paper, we have provided a novel approach leveraging historical recommendation data to isolate exogenous variation in pricing decisions which can be used in static as well as dynamic settings. Our approach credibly identifies the causal impact a price change has on demand as long as algorithmic price recommendations are not perfectly passed on to final prices. It also allows for counterfactual analysis in environments where such inference was not possible with traditional methods, such as instrumental variable approaches.

Applying our general framework to pricing in the hotel industry, we provide evidence that hotel managers seem to face significant fixed costs for adjusting actual rates. Consequently, they integrate the information contained in price recommendations sluggishly and only partially into realized prices. We then show how the discrepancies between the recommendations and realized prices can be used to identify hotel- and room-type-level demand elasticities. Our paper adds to the empirical literature on dynamic pricing also because we can leverage our high-frequency pricing and booking data to identify a subset of the data that exhibits quasi-experimental variation in prices. A simple difference-in-differences estimation applied to this subset validates our more general hotel-level estimates.

Our method can be used to measure dynamic patterns in demand that have critical implications for optimal pricing. In our application to hotel pricing, we have provided evidence for strategic behavior by a significant fraction of consumers and established the existence of substantial heterogeneity in the price sensitivity of the average consumer as a function of the number of days before arrival. In comparison to structural models, our reduced form approach can be easily adapted to account for the dependence of the price elasticity on other variables, by estimating the interactions between the variable of interest and prices. A typical example is estimating cross-price elasticities for substitutes. Note that we have not included cross-price effects of substitute products for the elasticity estimations in the main text because they seem negligible in our application (see D for details). For other applications, however, extending the estimation approach and the underlying identification strategy to incorporate cross-price elasticities more generally would be a valuable exercise which we leave for future research.

Finally, our approach only requires transaction-level data and, thus, represents an inexpensive alternative to both structural modeling and price experimentation which are often outside the reach of all but the largest and technologically most advanced corporations. It extends easily to many environments in which managers face a large number of small decisions, daily prices in our setting, and have access to algorithmic recommendations.

References

Alvarez, Fernando E., Francesco Lippi, and Luigi Paciello, "Optimal Price Setting with Observation and Menu Costs," *Quarterly Journal of Economics*, 2011, *126* (4), 1909–1960.

Athey, Susan, "Beyond Prediction: Using Big Data for Policy Problems," *Science*, 2017, *355* (6324), 483–485.

_ **and Guido W. Imbens**, "Machine Learning Methods That Economists Should Know About," *Annual Review of Econonomics*, 2019, *11*, 685–725.

Bajari, Patrick, Denis Nekipelov, Stephen P. Ryan, and Miaoyu Yang, "Machine Learning Methods for Demand Estimation," *American Economic Review: Papers & Proceedings*, 2015, *105* (5), 481–485.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen, "Inference on Treatment Effects after Selection among High-Dimensional Controls," *Review of Economic Studies*, 2014, *81* (2), 608–650.

_ **,** _ **, and** _ , "Post-Selection Inference for Generalized Linear Models with Many Controls," *Journal of Business & Economic Statistics*, 2016, *34* (4), 606–619.

Besbes, Omar and Ilan Lobel, "Intertemporal Price Discrimination: Structure and Computation of Optimal Policies," *Management Science*, 2015, *61* (1), 92–110.

Cachon, Gérard P. and Pnina Feldman, "Price Commitments with Strategic Consumers: Why it Can be Optimal to Discount more Frequently... Than Optimal," *Manufacturing & Service Operations Management*, 2015, *17* (3), 399–410.

Cachon, Gérard P, Santiago Gallino, and Marcelo Olivares, "Does adding inventory increase sales? Evidence of a scarcity effect in US automobile dealerships," *Management Science*, 2019, *65* (4), 1469–1485.

Caldentey, René, Ying Liu, and Ilan Lobel, "Intertemporal Pricing Under Minimax Regret," *Operations Research*, 2016, *65* (1), 104–129.

Castillo, Juan Camilo, "Who Benefits from Surge Pricing?," 2019. working paper.

Chen, Hongqiao, Ming Hu, and Jiahua Wu, "Intertemporal Price Discrimination Via Randomized Pricing," 2018. working paper.

Chen, Yiwei and Vivek F. Farias, "Robust Dynamic Pricing with Strategic Customers," *Mathematics of Operations Research*, 2018, *43* (4), 1119–1142.

Chernozhukov, Victor, Christian Hansen, and Martin Spindler, "Valid Post-Selection and Post-Regularization Inference: An Elementary, General Approach," *Annual Review of Econonomics*, 2015, *7* (1), 649–688.

Cho, Sungjin, Gong Lee, John Rust, and Mengkai Yu, "Optimal Dynamic Hotel Pricing," 2018. working paper.

Cohen, Peter, Robert Hahn, Jonathan Hall, Steven Levitt, and Robert Metcalfe, "Using Big Data to Estimate Consumer Surplus: The Case of Uber," 2016. NBER working paper.

Conlon, Christopher T. and Julie Holland Mortimer, "Demand Estimation Under Incomplete Product Availability," *American Economic Journal: Microeconomics*, 2013, *5* (4), 1–30.

Costello, Anna M, Andrea K Down, and Mihir N Mehta, "Machine+Man: A field experiment on the role of discretion in augmenting AI-based lending models," *Journal of Accounting and Economics*, 2020, (101360).

Crawford, Vincent P and Joel Sobel, "Strategic information transmission," *Econometrica*, 1982, pp. 1431–1451.

Donselaar, Karel H. Van, Vishal Gaur, Tom Van Woensel, Rob A.C.M. Broekmeulen, and Jan C. Fransoo, "Ordering Behavior in Retail Stores and Implications for Automated Replenishment," *Management Science*, 2010, *56* (5), 766–784.

Dube, Arindrajit, Jeff Jacobs, Suresh Naidu, and Siddharth Suri, "Monopsony in Online Labor Markets," *American Economic Review: Insights*, forthcoming.

Enz, Cathy A and Linda Canina, "Competitive Pricing in European Hotels," in "Advances in Hospitality and Leisure," Emerald Group Publishing Limited, 2010, pp. 3–25.

Fisher, Marshall, Santiago Gallino, and Serguei Netessine, "Setting Retail Staffing Levels: A Methodology Validated with Implementation," 2018. working paper.

Hunold, Matthias, Reinhold Kesler, Ulrich Laitenberger, and Frank Schlütter, "Evaluation of Best Price Clauses in Online Hotel Bookings," *International Journal of Industrial Organization*, 2018, *61*, 542–571.

Joo, Mingyu, Kenneth C. Wilbur, and D. Gauri, "Temporal Distance and Price Elasticity: Empirical Investigation of the Cruise Industry," *Management Science*, forthcoming.

Kartik, Navin, "Strategic Communication with Lying Costs," *Review of Economic Studies*, 2009, *76* (4), 1359–1395.

Kock, Anders Bredahl and Haihan Tang, "Uniform Inference in high-dimensional dynamic panel data models with approximately sparse fixed effects," *Econometric Theory*, 2019, *35* (2), 295–359.

Lazarev, John, "The Welfare Effects of Intertemporal Price Discrimination: An Empirical Analysis of Airline Pricing in US Monopoly Markets," 2013. working paper.

Li, Jun, Nelson Granados, and Serguei Netessine, "Are Consumers Strategic? Structural Estimation from the Air-Travel Industry," *Management Science*, 2014, *60* (9), 2114–2137.

_ , Serguei Netessine, and Sergei Koulayev, "Price to Compete... With Many: How to Identify Price Competition in High-Dimensional Space," *Management Science*, 2018, *64* (9), 4118–4136.

Liberti, Jose M and Atif R Mian, "Estimating the effect of hierarchies on information use," *Review of Financial Studies*, 2008, *22* (10), 4057–4090.

Mani, Vidya, Saravanan Kesavan, and Jayashankar M Swaminathan, "Estimating the impact of understaffing on sales and profitability in retail stores," *Production and Operations Management*, 2015, *24* (2), 201–218.

Miklós-Thal, Jeanine and Catherine Tucker, "Collusion by algorithm: Does better demand prediction facilitate coordination between sellers?," *Management Science*, 2019, *65* (4), 1552–1561.

Moon, Ken, Kostas Bimpikis, and Haim Mendelson, "Randomized Markdowns and Online Monitoring," *Management Science*, 2018, *64* (3), 1271–1290.

Nambiar, Mila, David Simchi-Levi, and He Wang, "Dynamic Learning and Pricing with Model Misspecification," *Management Science*, forthcoming.

Phillips, Robert Lewis, *Pricing and Revenue Optimization*, Stanford University Press, 2005.

Puller, Steven L and Lisa M Taylor, "Price discrimination by day-of-week of purchase: Evidence from the US airline industry," *Journal of Economic Behavior & Organization*, 2012, *84* (3), 801–812.

Reis, Ricardo, "Inattentive producers," *Review of Economic Studies*, 2006, *73* (3), 793–821.

Semenova, Vira, Matt Goldman, Victor Chernozhukov, and Matt Taddy, "Orthogonal machine learning for demand estimation: High dimensional causal inference in dynamic panels," 2017. working paper.

Shang, Jennifer, Pandu R Tadikamalla, Laurie J Kirsch, and Lawrence Brown, "A decision support system for managing inventory at GlaxoSmithKline," *Decision Support Systems*, 2008, *46* (1), 1–13.

Simon, Herbert A., "Rational Choice and the Structure of the Environment," *Psychological Review*, 1956, *63* (2), 129–138.

_ , "New Developments in the Theory of the Firm," *American Economic Review*, 1962, *52* (2), 1–15.

Talluri, Kalyan T. and Garrett J. van Ryzin, *The Theory and Practice of Revenue Management*, Springer, 2004.

Williams, Kevin, "Dynamic Airline Pricing and Seat Availability," 2018. working paper.

Tables and Figures (intended for main text)

Table 1: Summary Statistics of Rates and Bookings by Room Category

	Single	Standard	Superior	Suite
Actual rate	81.56	160.53	173.94	210.61
Actual rate (conditional on booking)	81.34	159.35	168.88	198.75
Total price per night	69.79	125.97	130.27	174.38
Recommended rate	81.25	161.08	176.20	215.05
Bookings per room	0.31	0.36	0.33	0.26
Duration of stay (days)	1.92	2.25	2.70	2.77

Notes: Mean values are reported by room category. Actual rate is the baseline price charged by the hotel, total price is the final price paid by the customer for one night, recommended rate is the price provided by the pricing algorithm, and booking per room is the mean utilisation of capacity, $\frac{1}{T}\sum_t \left[\frac{1}{N}\sum_h b_{h,t}/n_h\right]$, with $b_{h,t}$ the number of bookings in the given room category at hotel h at time t and n_h the capacity of hotel h in that room category.

Table 2: Actual Rate Updates over Time

	Days before arrival			
	100–0	50–0	30–0	14–0
Single	0.0251	0.0367	0.0444	0.0553
Standard	0.0465	0.0717	0.0918	0.1130
Suite	0.0189	0.0262	0.0322	0.0393
Superior	0.0347	0.0498	0.0616	0.0769

Notes: Fraction of days with an actual rate update by the hotel manager for different intervals of days before arrival. Data includes only products for which we observe $T \geq 100$ days before arrival.

Table 3: Distribution of Actual Rate Updates

Hotel ID	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
6	0.19	**0.24**	0.11	**0.04**	0.12	0.25	0.04
10	0.19	**0.06**	**0.27**	0.09	0.07	0.16	0.17
11	0.17	**0.47**	0.02	0.04	0.18	0.12	**0.00**
23	0.14	0.14	0.19	0.12	**0.24**	0.10	**0.08**
30	0.22	0.10	0.17	**0.25**	0.16	0.06	**0.03**
131	0.14	0.15	**0.16**	**0.07**	0.16	0.17	0.16
175	**0.22**	0.16	0.12	0.19	0.19	0.08	**0.04**
192	0.12	0.16	**0.30**	**0.04**	0.14	0.11	0.13
208	0.07	0.31	0.07	0.08	**0.36**	**0.03**	0.06

Notes: Numbers in bold indicate the day with maximal or minimal density of actual rate updates for each hotel. Rate updates for each hotel sum to 1, rounding errors may apply. Data includes only products for which we observe $T \geq 100$ days before arrival.

Table 4: Pass-Through of Recommended Rate into Actual Rate

	Dependent variable	
	$\sum_{s=t}^{t+9} \Delta$ actual rate (s)	$\sum_{s=t}^{t+9} \lvert\Delta$ actual rate $(s)\rvert$
Δrecommended rate (t)	0.101***	
	(0.001)	
$\lvert\Delta$recommended rate $(t)\rvert$		0.140***
		(0.002)
R^2	0.025	0.041
N	193,844	193,844

Notes: All regressions are estimated *conditional on a change in the recommended rate.* Pass-through of recommended rate into actual rate, is the sum of changes in the actual rate over the next ten days, i.e. instantaneous change together with the cumulative variation within the following nine days, regressed on the change in recommended rate (without a constant). Standard errors in parentheses. Data includes only products for which we observe $T \geq 100$ days before arrival. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Hotel Managers' Updating Behavior and Product Similarity

	Dependent variable	
	Δactual rate	Residual from column 1
Δrecommended rate	0.936***	
	(0.073)	
Days before arrival	-0.005	
	(0.009)	
Mean Δactual rate peers		0.964***
		(0.013)
Mean Δactual rate non-peers		0.036***
		(0.013)
Hotel-room type fe	Yes	–
Arrival month fe	Yes	–
$\bar{R}^2$	0.26	0.96
N	131,699	117,469

Notes: All regressions are estimated *conditional on a change in the actual rate.* A 'peer' of a product is defined as a room in a hotel on a given arrival date that is otherwise identical to the product in question but has its arrival date within a calendar week from the product being considered. For example, on the booking date January 2 for a double room in Hotel 6 on the arrival date of June 1 (Thursday) the Mean Δactual rate peers would be the average change in actual rate between January 1 and January 2 for double rooms in Hotel 6 for all arrival dates in May 29 (Monday) - May 31 (Wednesday) and June 2 (Friday) - June 4 (Sunday). The difference in the number of observations between column 1 and 2 is due to some products having no peers or all of the remaining observations being its peers. Standard errors clustered at the hotel level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Demand Semi-Elasticities in p.p.

	Model specification				
	Simple OLS	DS linear	DS linear	DS log-log	DS logit
Actual rate (L)	0.23	0.27**	−0.18***	0.04	0.05
	(0.16)	(0.11)	(0.03)	(0.18)	(0.06)
Actual rate (M)	−0.59***	−0.15	−0.54***	−0.29	−0.92***
	(0.06)	(0.15)	(0.19)	(0.29)	(0.20)
Actual rate (H)	−2.34***	−0.38***	−1.24***	−0.84***	−1.52***
	(0.10)	(0.11)	(0.36)	(0.11)	(0.24)
Recommended rate	Yes	No	Yes	Yes	Yes

Notes: The dependent variable is the probability of a booking for a given product on a given date. We run a separate regression for each hotel. The OLS results control for room type and recommended price. DS in columns 2-5 refers to double selected models. In columns 3-5 the double selection method is allowed to select from a third-order polynomial of the recommended rate and its interactions with the month of arrival and weekday of booking, room type, interactions between the weekday of arrival and arrival month, and indicators for the number of days before arrival. In column 2 the set of controls is the same but does not include the recommended rate or any of its interactions. Reported coefficients correspond to the 3rd (H), 5th (M), and 7th (L) most elastic estimate out of the 9 hotels. The semi-elasticities for the logit model in the last column are calculated using the average marginal effect of a price change divided by the average booking rate. Standard errors in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: Estimates of the Slope of the Demand Curve and Implied Demand Semi-Elasticities in p.p. from the Difference-in-Differences Specification

	Model specification			
	All observations	Excl. suites	Last 20 days	Excl. last 20 days
Δactual rate	-0.0018	-0.0027	-0.0044	-0.0012
	(0.0018)	(0.0024)	(0.0055)	(0.0018)
Δactual rate $\times$ copied all recs	-0.044^{***}	-0.047^{***}	-0.063^{***}	-0.040^{**}
	(0.017)	(0.016)	(0.17)	(0.18)
Copied all recommendations	-0.827^{***}	-0.94^{***}	-0.145	1.07^{***}
	(0.262)	(0.271)	(0.274)	(0.341)
Average booking rate	0.038	0.041	0.077	0.019
Semi-elasticity in p.p. (interaction)	$\mathbf{-1.20}$	$\mathbf{-1.22}$	$\mathbf{-0.88}$	$\mathbf{-2.22}$
Product category - time fe	Yes	Yes	Yes	Yes
N	5,886,196	4,603,198	631,864	5,254,332

Notes: The dependent variable is the change in the number of a bookings (in p.p.). Copied all recs equals one if on that booking day the hotel manager updated at least one price and all of the prices for that arrival week matched the recommendation. Note that this requires matching prices also for other room types. Fixed effects are at the level of unique combinations of room category, date of arrival and booking date. Standard errors shown in the parentheses are clustered at the hotel level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8: Triple-Selected Instrumental-Variable Regressions

	Set of instruments and booking horizon			
	All	Only hotel 6	All excluding last 15 days	
Actual rate (L)	1.6***	0.12*		
	(0.03)	(0.05)		
Actual rate (M)	0.8***	0.05	0.001	-0.5^{***}
	(0.20)	(0.04)	(0.004)	(0.02)
Actual rate (H)	-0.025	0.01		
	(0.07)	(0.20)		
Recommended rate	Yes	Yes	No	Yes
Hotel ID	All	6 excluded	175	175

Notes: The dependent variable is the number of bookings for a given product on a given date. All specifications use the three-step selection procedure introduced in the main text. We run a separate regression for each hotel. Reported coefficients correspond to the 3rd (H), 5th (M), and 7th (L) most elastic estimate out of the 9 hotels. Control variables include month and day of arrival and room-type indicators and days before arrival. In columns 1, 2 and 4, the triple selection method is allowed to select from a third-order polynomial of the recommended rate and its interactions with the month of arrival and weekday of booking, room type and fixed effects for days until arrival. Instruments are prices in other hotels. In column 2 we only use prices in hotel 6 as instrument, and we drop hotel 6 from the main analysis. In columns 3 and 4 we exclude the last 15 days. Standard errors in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 9: Double-Selected Demand Semi-Elasticities in p.p. Including Lagged Prices

	Model specification		
	Baseline	Incl. 8-day lags	Incl. interactions
Hotel 175 ($N = 827,797$)			
Actual rate	-2.75^{***}	-1.62^{***}	-3.02^{***}
	(0.12)	(0.45)	(0.14)
Lagged rate		-1.36^{**}	
		(0.45)	
Actual rate × hike			0.22^{***}
			(0.03)
Actual rate × drop			0.08^{**}
			(0.03)
Hotel 6 ($N = 204,635$)			
Actual rate	-1.49^{***}	-3.22^{***}	-1.61^{***}
	(0.22)	(0.54)	(0.22)
Lagged rate		1.63^{***}	
		(0.51)	
Actual rate × hike			0.13
			(0.07)
Actual rate × drop			0.30^{***}
			(0.08)

Notes: The dependent variable is the probability of a booking for a given product on a given date. Control variables that the double selection is allowed to select from include month and weekday of arrival, room type indicators, a third-order polynomial of the recommended rate and its interactions with the month and weekday of arrival, interactions between the weekday of booking and arrival month, and indicators for the number of days before arrival. Actual rate is the rate on the day of booking. Lagged rate is the rate 8 days before the booking. Hike (drop) equals one if the current price exceeds (is lower than) the lagged price. Standard errors in parentheses. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 10: Double-Selected Demand Semi-Elasticities in p.p. Over Time

	Days before arrival			
	≥ 100	$100-51$	$50-21$	$20-0$
Hotel 6	-1.83^{***}	-1.15^{**}	-0.31	-0.81^{**}
	(0.58)	(0.58)	(0.56)	(0.34)
Hotel 175	-2.15^{***}	-1.61^{***}	-1.41^{***}	-1.78^{***}
	(0.21)	(0.23)	(0.19)	(0.12)

Notes: The dependent variable is the probability of a booking in hotels 175 or 6 for a given product on a given date. Control variables that the double selection is allowed to select from include month and weekday of arrival, room type indicators, a third-order polynomial of the recommended rate and its interactions with the month and weekday of arrival, interactions between the weekday of booking and arrival month, and indicators for the number of days before arrival. Actual rate is the rate on the day of booking. Lagged rate is the rate 8 days before the booking. We run separate regressions for each hotel and each booking horizon. Standard errors in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

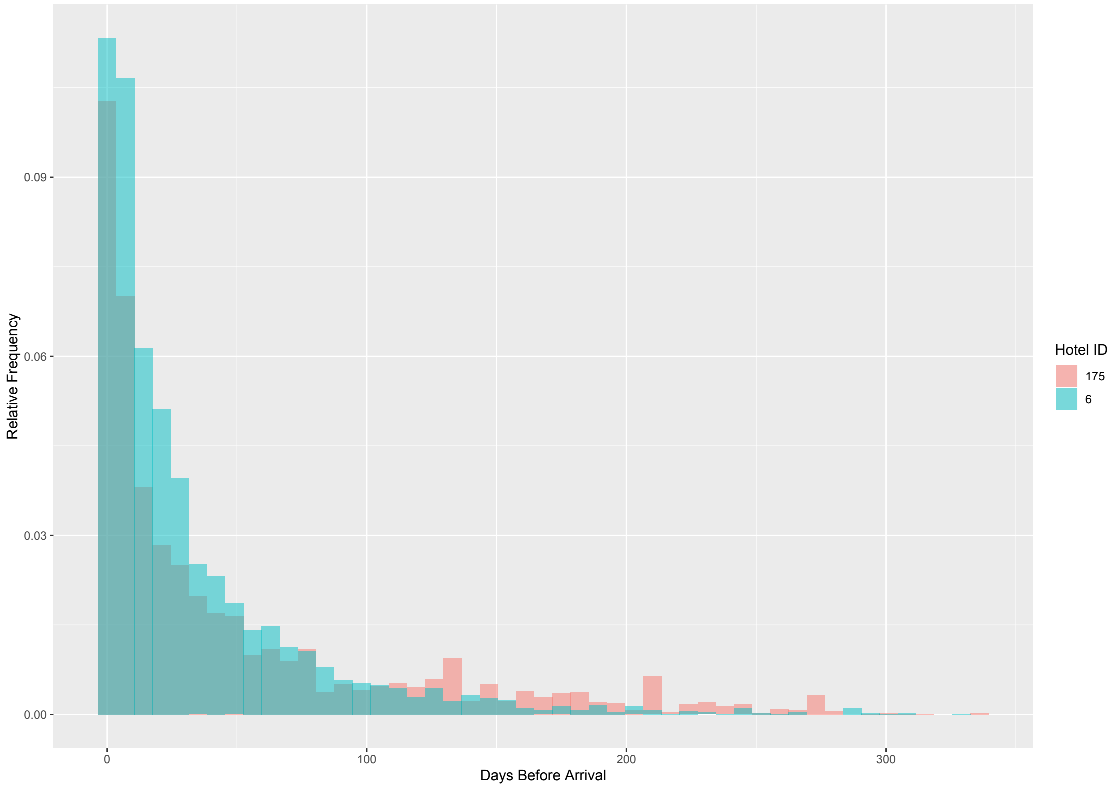
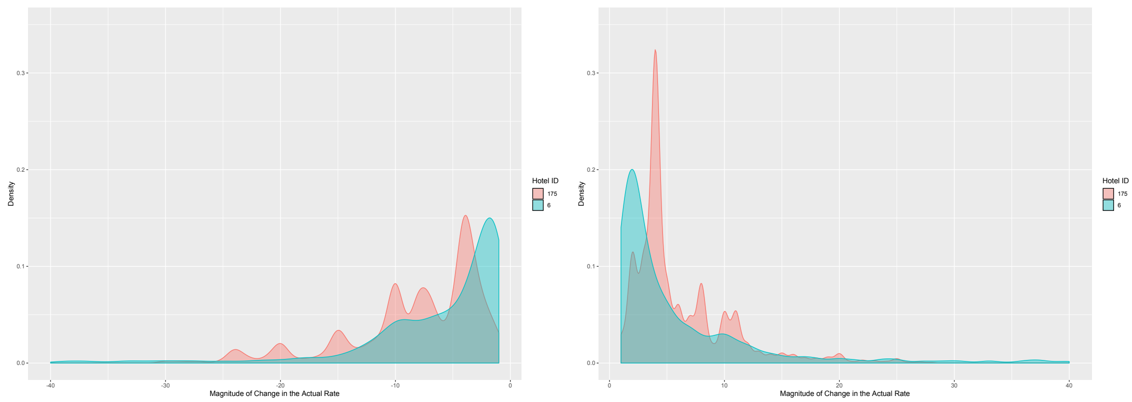


Figure 1: Arrival of Bookings as a Function of Time Before Arrival



(a) Price Decreases

(b) Price Increases

Figure 2: Empirical Distribution of Actual Rate Changes

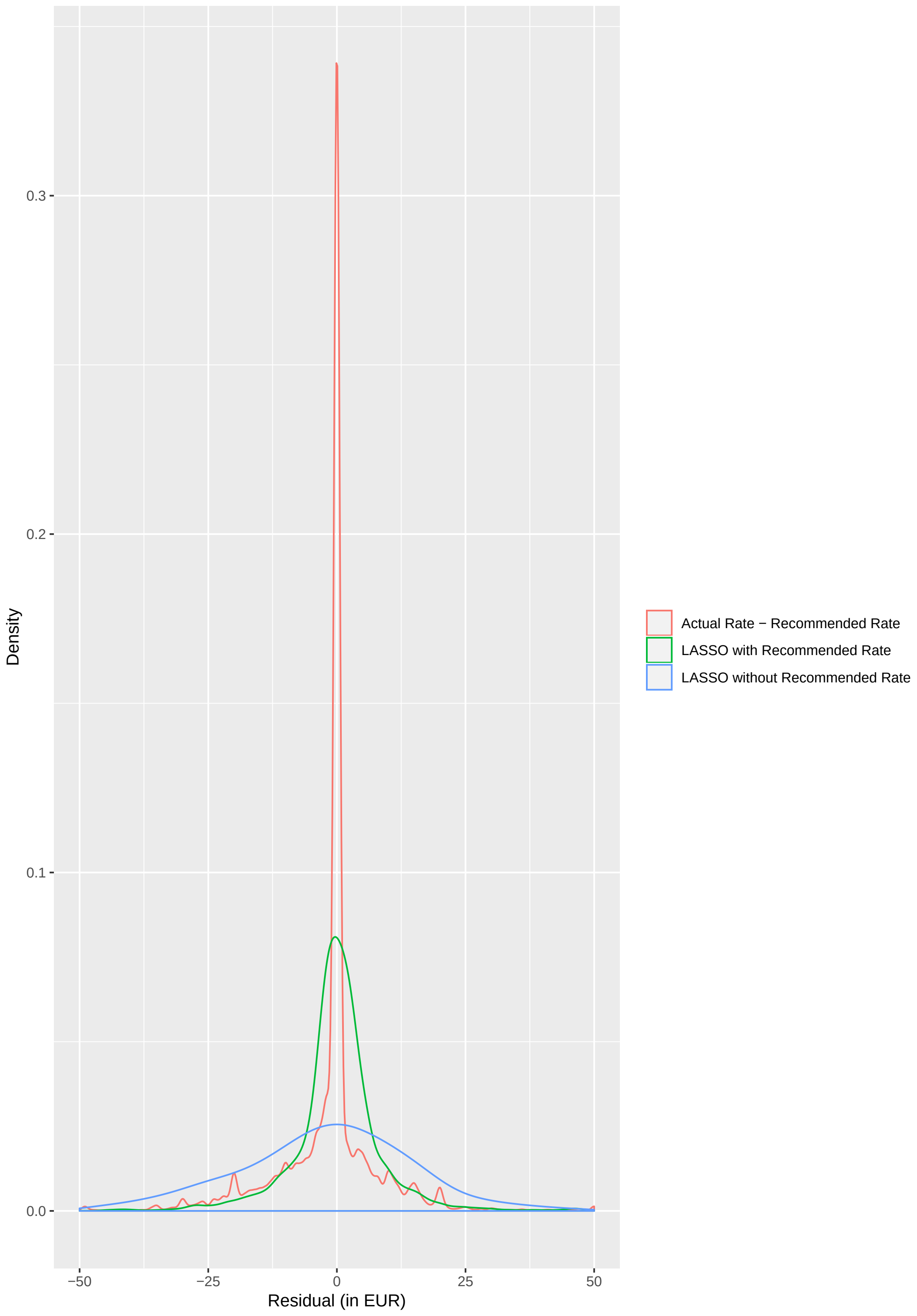


Figure 3: Difference Between Actual Rate and Recommended Rate Compared to Residuals from LASSO Predictions With and Without Recommended Rate

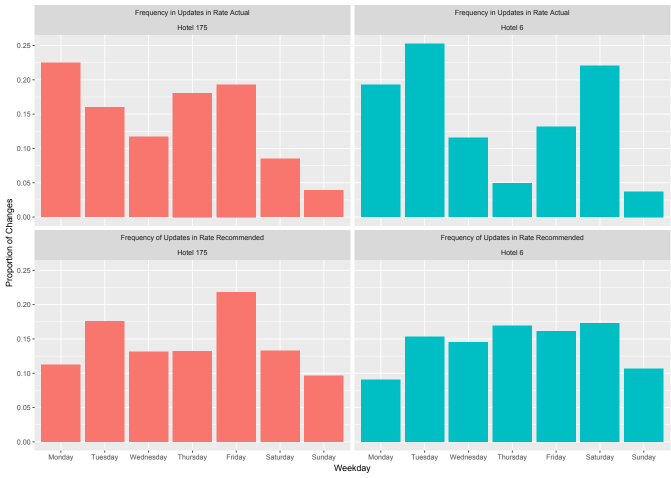


Figure 4: Frequency of Updates in Actual and Recommended Rate Across Weekdays for Hotel 6 and 175

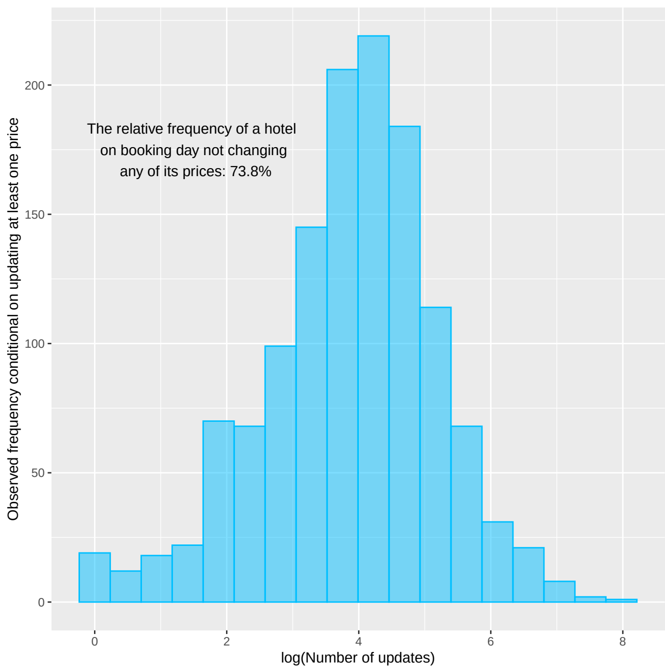


Figure 5: Distribution of Number of Changes (in Log) in Actual Rate Conditional on at Least One Rate Update

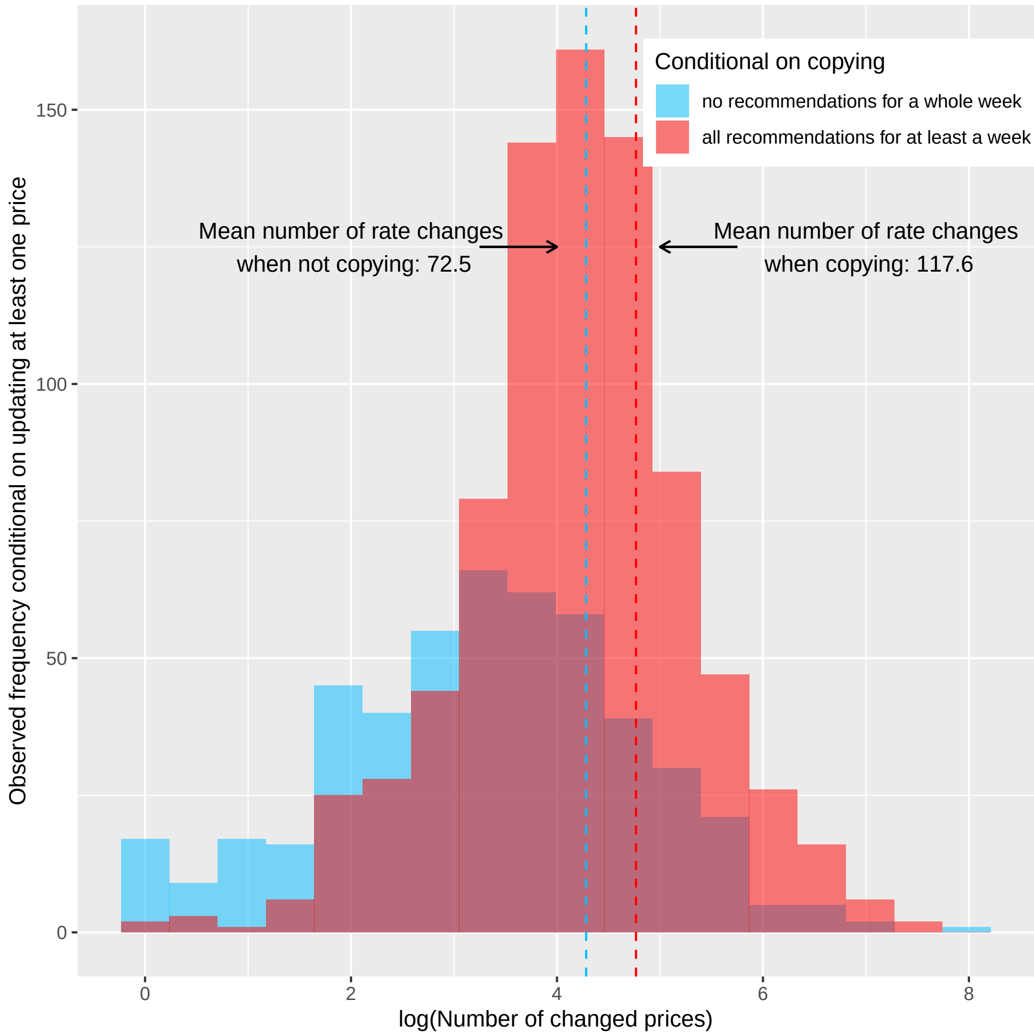


Figure 6: Number of Rate Updates (in Log) Conditional on Copying and Not Copying Recommended Rates

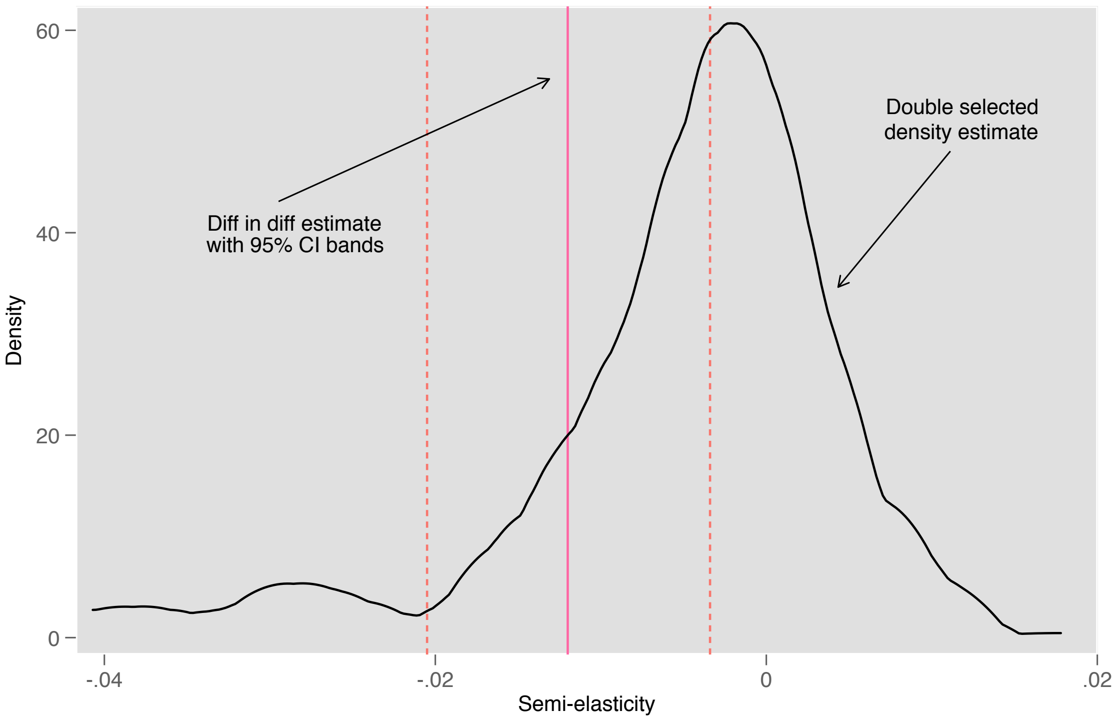


Figure 7: Distribution of Estimated Semi-Elasticities Across all Room Types and Hotels

Note: Kernel density plot (Epanechnikov kernel with bandwidth = 0.0027) of the estimated semi-elasticities across all room types and hotels using double selection, see Table 6. Although kernel density estimates exhibit some mass at positive values, only three estimates are statistically larger than zero. The figure also plots the difference-in-differences estimate reported in Table 7. The vertical line represents the difference-in-differences estimate and the dashed lines the corresponding 95% confidence intervals.

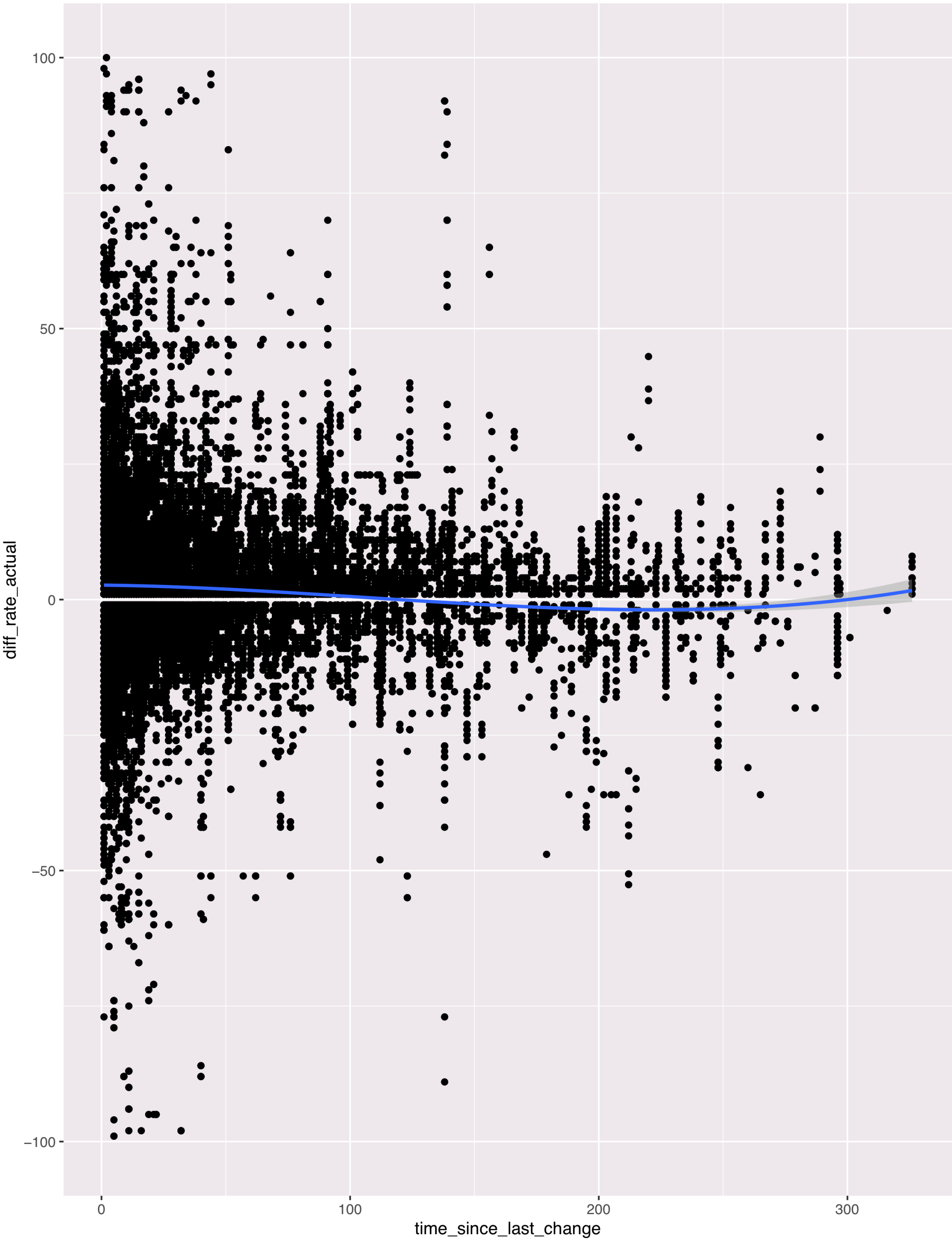


Figure 8: Changes in Actual Rate as a Function of Days Since the Last Update

Note: An observation corresponds to an update in the actual rate. The blue line plots the best locally smooth predictor of the average change as a function of days since last update.

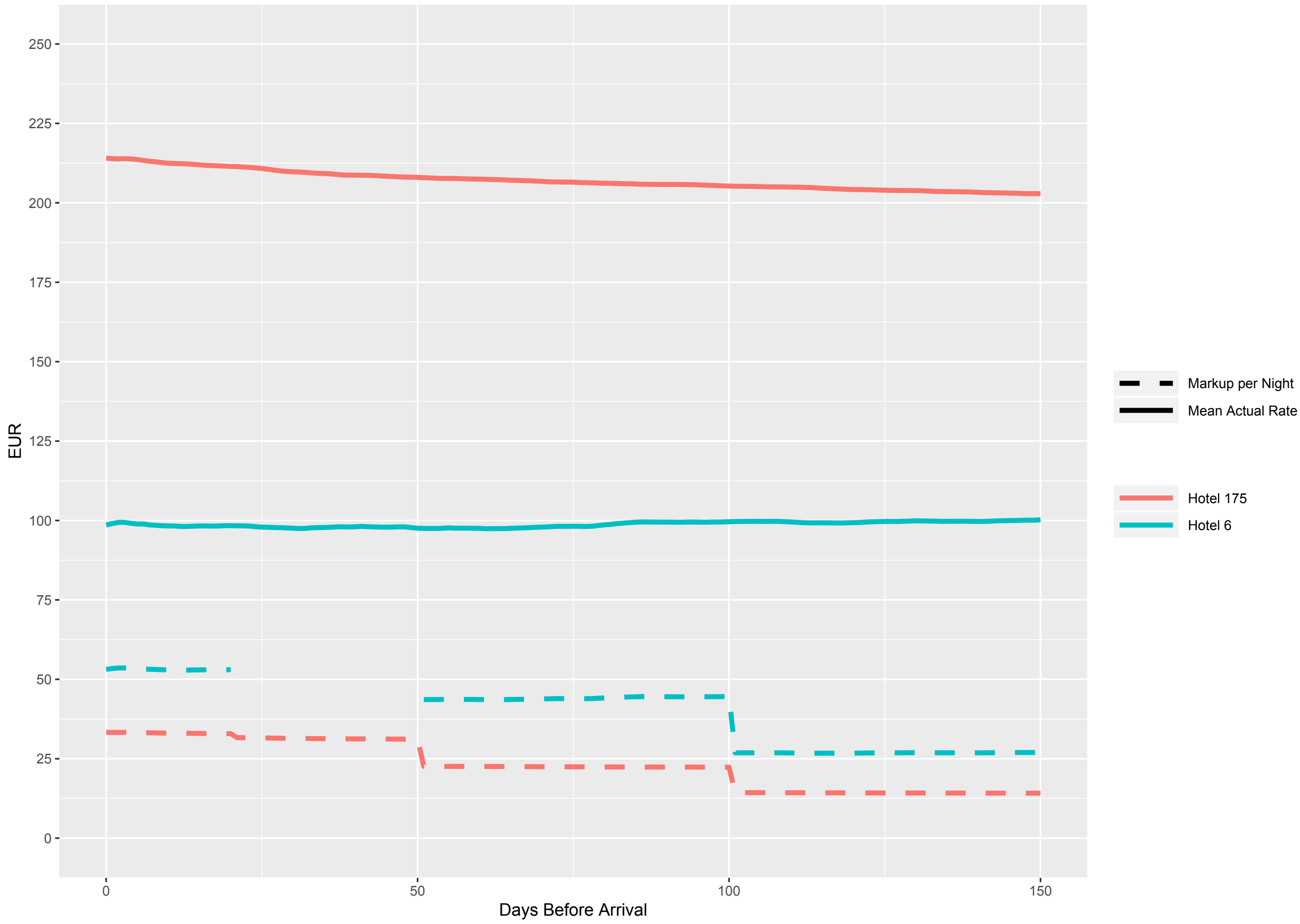


Figure 9: Markups and Acutal Rates by Days Before Arrival

Note: Bookings 21-50 days before arrival for Hotel 6 are excluded due to imprecise estimates (Table 10 column 4).

Supplementary Materials

for

"Demand Estimation Using Managerial Responses to Automated Price Recommendations"

by D. Garcia, J. Tolvanen, and A.K. Wagner

A Dynamic Demand Model

A strategic consumer who considers hotel h at date t has three alternatives. First, she may decide to book a room and obtain $v_t - p_t$. Second, she may execute her outside option of w_t. Finally, she may decide to wait obtaining a (perceived) present value of $S_t(v_t, p_t)$, independent of w_t. We assume that $S_t(v_t, p_t)$ is weakly increasing in v_t and decreasing in p_t and that the function $v_t - p_t - S_t(v_t, p_t)$ is strictly increasing in v_t, strictly decreasing in p_t and there exists a unique value $v^*(p)$ such that $v^*(p) - p - S(v^*(p) - p) = 0$. Let $F_w(w)$ be the distribution of the value of the outside option and let $F_v^t(v)$ be the distribution of valuations for hotel h among consumers alive in period t (which may be endogenous). We define

$$\Delta_1 = \int_{v^*(p)} f_w(v-p) dF_v^t(v) + f_v^t(v^*(p)). \tag{15}$$

The first term captures those consumers who would otherwise buy now but a price increase leads to execute the outside option. The second term contains those consumers who are discouraged from buying now but are willing to wait. Similarly,

$$\Delta_2 = \int^{v^*(p)} f_w(S(v,p)) S_p(v,p) dF_v^t(v) \tag{16}$$

captures the fraction of consumers who are now discouraged from waiting further. Notice that in the basic framework presented in the main text, $\Delta_2 = 0$. In general, however, $\Delta_2 \geq 0$ as some consumers become pessimistic about their future prospects following a higher price realization. The main hypothesis can then be written as

$$\eta_t f_v^t(v^*(p_t)) < \int^{v^*(p_t)} f_w(S(v,p_t)) S_p(v,p_t) dF_v^t(v) \tag{17}$$

for all t. This will always hold if there is sufficient variation in the perceived distribution of prices over time so that $S_p(v, p_t)$ is low enough. This does not require, however, that prices are independent over time. For instance, if consumers expect that $p_t = p_{t-1}$ with probability q and p_t is distributed according to $H(p)$ with probability $1-q$, then $S_p(v, p_t) = 0$ even if the serial correlation in prices is arbitrarily large.

B Simulations

In this section we describe the simulation models which set up to estimate the price semi-elasticity of demand for the different econometric specifications presented in the main text. For the baseline simulation model, we first generate the price and quantity paths of 1000 rooms (or products) over a booking horizon of 100 days. For each room, we draw the 'initial utility' of a consumer, u_{iT}, using the linear model

$$u_{iT} = \sum_{k=1}^{250} v_{ki}\delta_k, \tag{18}$$

where $\delta_k = 0$ if variable k is irrelevant for demand and otherwise δ_k is distributed according to a log-normal with mean $2e^{1/2}$. Variable v_{ki} is the realization of variable k in hotel i, drawn from a standard uniform distribution. We keep the model sparse by setting parameter $\delta_k = 0$ with probability 0.8. We assume further that the time-specific mean utility a consumer received from product i in period $t = T, T-1, ..., 1$ follows a random walk, with

$$u_{it} = u_{it+1} + \epsilon_{it},$$

where ϵ_{it} is normally distributed (with standard deviation equal to 1/50 of the expected initial utility). We generate a similar model on the supply side with the initial marginal cost of production, c_{iT}, drawn according to

$$c_{iT} = \sum_{k=1}^{250} v_{ki}\theta_k, \tag{19}$$

and subsequent costs are given by

$$c_{it} = c_{it+1} + \varepsilon_{it},$$

with ε_{it} normally distributed (with standard deviation equal to 1/50 of the expected initial cost). Similarly to the demand component, $\theta_k = 0$ if variable k is irrelevant for supply and is distributed according to a log-normal with mean $2e^{1/2}$ otherwise. Cost shocks are correlated with prices but not with demand, introducing exogenous variation in prices.

Given utility u_{it}, the probability that a consumer purchases in period t is $\lambda_t(u_{it} - bp_{it})$; parameter λ_t measures the arrival rate at period t and p_{it} is the realized price. We set the initial price equal to the optimal price, assuming unlimited capacity, and formulate the following adjustment model. The revenue manager submits a new recommendation in every period with a certain probability. The recommended price is interpreted as the revenue-maximizing price in that period. We assume the hotel manager can perfectly infer the demand shock from this recommendation. Whenever the hotel manager decides to adjust the price, she will choose the optimal price given the shock. Consistent with evidence from our data set, we assume that price adjustments are less frequent than updates in recommended rates.

We outline different simulation models as a proof of concept of our estimation approach. As shown in Table 11, the simulated models are then estimated for different econometric specifications, including the double-selected framework (DS linear), and compared against the simulated model's true price semi-elasticity. In the 'baseline' simulated model (column 1), parameters are chosen in order to obtain an average booking rate of 0.05, and a price semi-elasticity of approximately -1.5 p.p. In addition to the baseline simulation, we exlude supply shocks in the 'no supply shock' model (column 2) by setting the variance of ε equal to 0. In the 'low inertia' model (column 3), we change the frequency of price adjustments by increasing the probability of adjustment in prices from 0.2 to 0.4. Finally, we reduce the number of nuisance parameters from 250 to 100 in the 'low dimension' model (column 4). Note that we compute the true model-specific demand semi-elasticity using numerical derivatives because direct comparison across models is not possible.

Table 11 provides the estimated price semi-elasticity for each of the four different estimation specifications (in rows) and the four different simulation models (in columns). The first row shows the estimates of a linear model (LM) in which price recommendations are omitted. Similarly to the empirical results reported in the main text, the linear model is severely biased and yields *positive* elasticities for any of the simulated models. We also run a simple difference-in-differences model for which we use only variation in prices in periods in which the recommendation is constant. Since the sample size becomes relatively small, the coefficient for each of the simulation models is quite noisy as shown in row 2. Results of a simple linear model including the price recommendation and time fixed-effects are shown in row 3. Including recommended rates, we identify a negative price coefficient but the regres-

Table 11: Estimated Price Semi-Elasticity in p.p. for Simulated Models

	Simulation model			
	Baseline	No supply shocks	Low inertia	Low dimension
LM without recommended rate	0.56^{***}	0.57^{***}	0.79^{***}	0.65^{***}
	(0.01)	(0.01)	(0.01)	(0.01)
Diff-in-Diff	$\mathbf{-1.10}^{**}$	$\mathbf{-1.11}^{**}$	$\mathbf{-1.17}^{**}$	-1.14^{**}
	(0.39)	(0.39)	(0.37)	(0.36)
LM with recommended rate	-0.90^{***}	-0.74^{***}	$\mathbf{-1.50}^{***}$	-1.42^{***}
	(0.12)	(0.12)	(0.14)	(0.10)
DS linear	$\mathbf{-1.62}^{***}$	$\mathbf{-1.49}^{***}$	$\mathbf{-1.53}^{***}$	-2.10^{***}
	(0.16)	(0.17)	(0.12)	(0.14)
True coefficient	−1.39	−1.50	−1.48	−1.77

Notes: The dependent variable is the probability of a booking. The specifications of the simulated models are described in the text of this section. The econometric specifications used to estimate price semi-elasticities for each simulated model are the ones used in the main text. Bootstrapped standard errors for a simulated sample with 1000 rooms and 100 booking days in parentheses. Boldface estimates if we cannot reject the true value at $p = 0.05$. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

sion model underestimates its magnitude. We find that difference-in-differences estimates are close to the true coefficient only if there is substantially more variation in prices ('low inertia' model in column 4) or fewer confounders ('low dimension' model in column 5). As can be seen from Table 11, our preferred double-selected linear model obtains estimates much closer to the true price effect than any other regression specification for all but the simulated model in column 4.

As a further robustness test of our empirical approach, we also study the performance of our empirical estimation framework by relaxing some of the following main assumptions. In the simulated model with 'endogenous adjustments' (column 1), we implement adjustment costs of hotel manager such that prices adjust only if the difference between the current price and the optimal price exceeds 4% of the mean price (conditional on an observation). In the model assuming 'biased recommendations' (column 2), recommendations are allowed to

Table 12: Estimated Price Semi-Elasticity in p.p. for Simulated Models

	Simulation model			
	End. adjustment	Biased recs	Mean-reverting	Strategic
LM without recommended rate	0.80^{***}	1.13^{***}	0.79^{***}	0.79^{***}
	(0.01)	(0.01)	(0.01)	(0.01)
Diff-in-Diff	$\mathbf{-1.18}^{**}$	-2.12^{***}	$\mathbf{-1.12}^{**}$	$\mathbf{-0.92}^{*}$
	(0.41)	(0.87)	(0.37)	(0.37)
LM with recommended rate	-0.87^{***}	-2.52^{***}	-0.71^{**}	-0.02
	(0.11)	(0.25)	(0.12)	(0.13)
DS linear	$\mathbf{-1.61}^{***}$	$\mathbf{-3.43}^{***}$	$\mathbf{-1.39}^{***}$	$\mathbf{-1.16}^{***}$
	(0.16)	(0.36)	(0.12)	(0.16)
True coefficient	-1.52	-3.92	-1.42	-0.86

Notes: The dependent variable is the probability of a booking. The specifications of the simulated models are described in the text. The econometric specifications used to estimate price semi-elasticities for each simulated model are the ones used in the main text. Bootstrapped standard errors for a simulated sample with 1000 rooms and 100 booking days in parentheses. Boldface estimates if we cannot reject the true value at $p = 0.05$. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

be biased so the manager chooses $p_{it} = \phi r_{it} + (1-\phi)p_0 + \nu_{it}$, with $\phi = 0.8$ and ν_{it} being white noise (equal to the variance of ϵ). In the 'mean-reverting' simulated model (column 3), we allow for the consumer's utility u_{it} to be mean-reverting by assuming an autocorrelation parameter of 0.9. Finally, we also introduce a simulation model with strategic consumers (column 4). We model this by simply assuming that if a consumer arrives but does not buy in a given period, she stays in the market in the future with probability $\delta = 0.3$.

Estimation results are reported in Table 12. We find that, in all of the simulated models, estimates of our DS linear regression model are consistent with the true price elasticities and are much more precise than the difference-in-differences estimates. Note also that the average price elasticity in the model with biased recommendations (column 2) is much higher because realized prices depart substantially from the static optimal, and the average selling probability is only half of the one in the other models. In addition, we also run the DS linear model specification with lagged prices (as the one in Table 9) using the simulated data with patient consumers. In this case, we identify a slightly higher contemporaneous price elasticity (-1.29) than in the specification without a lag (see column 4). The effect of the lag is positive (0.37) and around 1/3 of the main coefficient but insignificant due to lack of power; these results are in line with the ones found for hotel 6 in Table 9.

C Heterogeneous Price Elasticities

Our method can also be applied to estimate the extent of heterogeneity in demand elasticities that is critical for the development of more complex dynamic pricing algorithms. To demonstrate the flexibility of the method, we estimate the demand semi-elasticities first for different booking weekdays and then for different arrival seasons.

Heterogeneity in booking patterns creates an opportunity for more nuanced pricing if consumers, who differ in their weekday booking pattern, also exhibit differences in willingness to pay for a room. In the airline industry, for example, it is known that different types of consumers typically book their flights on different weekdays (e.g. Puller and Taylor, 2012). We investigate whether similar booking patterns are also present in our hotel data. The first row in Table 13 presents the lowest, highest and the median estimate of the price semi-elasticity over the seven possible booking weekdays. In hotel 175 the elasticities are highly constant across days of the week, except for weekends (especially Saturdays), when the demand elasticity is considerably lower than on other booking weekdays. In hotel 6 we see somewhat more variation across booking days than in hotel 175. The most elastic booking day for hotel 6 is Thursday while Tuesdays are the least elastic. The difference between the two hotels is likely to be due to differences in consumer types, because hotel 6 is a city hotel and hotel 175 a holiday resort.

In the regions where the two hotels operate, the months from December to March and June to September are considered high seasons, especially among holiday travelers, and remaining months are considered off-season. We estimate the hotels' demand semi-elasticities for these three groups of months. The 'arrival season' results are presented in Table 13. The expected demand pattern is reflected in the estimates for hotel 175: demand is highly inelastic during the winter months and fairly elastic during the off-season. For the city hotel 6, demand varies much less across seasons.

Table 13: Double-Selected Demand Semi-Elasticities in p.p.

	Hotel 175			Hotel 6		
	Min	Median	Max	Min	Median	Max
Booking weekday	-2.82^{***}	-2.51^{***}	-0.62^{*}	-4.22^{***}	-2.97^{***}	-1.36^{*}
	(0.21)	(0.23)	(0.33)	(0.60)	(0.70)	(0.60)
	Winter	Summer	Off-season	Winter	Summer	Off-season
Arrival season	-0.49	-2.42^{***}	-4.22^{***}	-3.18^{***}	-2.71^{***}	-2.17^{***}
	(0.29)	(0.18)	(0.23)	(0.75)	(0.82)	(0.43)

Notes: The dependent variable is the probability of a booking for a given product on a given date. Control variables include month and weekday of arrival, month of arrival, their interaction, number of days before arrival, room type, and a third-degree polynomial of the recommended rate, interacted with month of arrival, weekday of booking and the number of days before arrival. Standard errors in parentheses. Semi-elasticity estimate uses the average probability of a booking as the quantity sold. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Finally, note that the method presented in Semenova et al. (2017) can be combined with our method to select and estimate the most important dimensions of demand heterogeneity. However, such application is beyond the scope of this paper.

D Cross-Price Effects

Cross-price effects, if not included in the analysis, can lead to biased price elasticity estimates. The analysis in the main text abstracted from cross-price effects by assuming implicitly that a hotel sells a single room type. In this section, we comment on how our approach can be adapted to account for cross-price effects between products. We then present results of regression specifications including prices and price recommendations of other products (e.g. substitutes) and establish, for our data, robustness of the approach presented in the main text.

First, note that the estimation approach presented in the main text remains valid and produces consistent estimates as long as price changes, conditional on covariates and recommendations, are crucially not correlated across different rooms in the same hotel. Table 5 shows that price changes, conditional on the recommendation, are highly correlated across different dates in a given arrival week for the same room but not across different room categories in the same week or across weeks. Given that the average length of stay is less than three days and the fact that prices of a room are almost perfectly correlated within a week, our estimates would be robust to substitution of the same room at a different day of the week; consumers would not find it profitable to switch across nights given the strong co-movement of relevant prices.

Considering a possible substitution between different room types for the same arrival day, we next show that our original price estimates are robust to adding cross-price effects across room types to our model. To this end, we use the regression specification in equation 8 in the main text, restrict the data to the most popular room type (usually a standard double), and include prices and price recommendations of the most popular alternative room type (most often a single room or a suite) for a hotel. Moreover, we restrict the sample so that both types of rooms are offered at the same time. We then compare the estimated price coefficient (η) from regressions *including* and *excluding* prices and price recommendations of the most-popular alternative room type.

The estimated price coefficients for the two specifications for hotel 6 and 175 are as follows. For hotel 175, the coefficient is -6.42e-04*** (8.67e-05) if the alternative room type is excluded from the regression and -5.28e-04*** (8.65e-05) with the alternative room included. For hotel 6, the coefficients are -0.0011*** (1.40e-04) with the alternative room type excluded and -6.94e-04* (3.50e-04) in the case of inclusion. The above results show that the inclusion of the alternative room type does not change price coefficient estimates significantly as the null hypotheses that the coefficients are the same cannot be rejected at the 5% level for the two hotels. The same result holds for all other hotels in our data set which had significantly negative elasticity estimates in the original model presented in the main text.