Reference price effects in vacation rental markets

Shrabastee Banerjee
Boston University
Anita Rao
University of Chicago
Giorgos Zervas
Boston University

Cite as:
Banerjee Shrabastee, Rao Anita, Zervas Giorgos (2020), Reference price effects in vacation rental markets. *Proceedings of the European Marketing Academy*, 49th, (59390)

Paper from the 49th Annual EMAC Conference, Budapest, May 26-29, 2020.

Reference price effects in vacation rental markets

Consumers have been shown to respond to prices that are advertised by firms, but irrelevant to the actual purchase price. However, in most applications, the advertised price is higher than the focal price, and is framed as a discount. Our focus is on travel websites, who often advertise a “Starting from” price for each property. Using a combination of data sets and methods, we investigate whether consumers respond to these “From” prices. Our experimental results (obtained from Holidu.com) indicate that, while higher From price might lead to lower clickouts, it can attract more valuable customers conditional on purchase. To complement this experiment, our observational results (obtained from Airbnb.com) find that From prices are positively correlated with booking probability and occupancy rate. Taken together, our findings lend some evidence to the fact that consumers are guided by fairness perceptions, and are deterred if offered a very low initial From price, but a very high eventual price.

Keywords: Reference price, Field experiment, Causal machine learning
Track: Digital Marketing and Social Media

1. Introduction

A reference price is defined as the standard against which the purchase price of a product is judged (Monroe (1973)). In other words, the utility from any purchase depends not just on the price paid and product characteristics, but also on how much the price 'deviates' from the customer's reference point (e.g Winer (1986)). A sizeable literature in marketing, economics and psychology has established the various channels through which reference prices impact purchase likelihood, brand evaluations and other outcomes of interest.

Reference prices that are formulated by consumers themselves (typically "internal" or "external" reference prices) usually rely on past purchase experiences, or prices of other comparable products. Thus, retailers cannot exert direct control over them. On the other hand, advertised reference prices (ARPs) are supplied by sellers themselves, and can thus serve as a direct tool of comparative price advertising with which consumer price perceptions may be altered (see Jindal (2018) for a recent application).

Typically, ARPs are presented in a "was-now" framing, or as a striked through price to emphasize apparent savings off a given list price. However, online marketplaces offer far greater flexibility in showcasing these non-focal prices. With the vast majority of online prices being consumer-specific and customisable based on exact specifications, retailers have to make decisions with regard to defaults, i.e, prices that are displayed before any specifics are entered. This decision is particularly relevant in the context of marketplaces or aggregators, who are often themselves not price setters, but have complete control over the way price information is summarised or presented on the product page.

In the case of vacation rentals, which is our setting, the key problem faced by the aggregator is in deciding what price to show consumers who search for accommodation without entering any dates. The most common form that these price estimates take in this setting is the "Starting from" price, which computes the minimum based on certain criteria, e.g the minimum nightly price over the next calendar month.

Here, we aim to measure how the presence of such "Starting from" prices affects customers' booking behavior in the context of both a vacation rental marketplace (Airbnb.com) and a travel aggregator (Holidu.com).

2. Data And Empirical Set-Up

Our empirical analyses proceeds in two steps:

2.1. Experiment on Holidu.com

First, we conduct a large scale field experiment over 48 days (March and April 2018) on Holidu.com, a Europe-based travel aggregator. There are 1,802,867 users in total in the experiment, searching for accommodation across 35,950 cities in 89 countries. These users originate from 37 unique domains, which serve as a rough proxy for origin country locations.

We randomly assign treatment users to see a From price which is between 5 and 10% higher than baseline (example in Figure 1). We further have access to a set of rich pre-treatment covariates (such as host IP, country and city being searched for, device used, number of searches by that user till date, etc). We find that overall, higher From prices lead to lower clickouts. However, there is a directional positive effect on the total booking value. Given the sparsity in bookings, we are under-powered and fail to detect a statistically significant effect for either the number of bookings or booking value for the full sample. To obtain additional precision, we condition on users who made a booking, and use flexible functions of our pre-treatment covariates as controls. We implement a rigorous Lasso (Rlasso) regression (Belloni et al. (2014)), a tool for variable selection that prevents over-fitting when sample sizes are limited compared to covariates and their interactions. We now find that treatment users spend significantly more on bookings relative to baseline. It is important to mention however that the conditional analysis might be subject to self-selection, i.e, we might not be comparing identical groups of consumers. Our parametric controls partially account for selection on observables - further, we employ a coarsened exact matching (CEM) algorithm that non-parametrically stratifies the covariate space and creates the best matches between treatment and control groups, computes a treatment effect for each strata, and weighs them according to the total number of observations in the strata (Iacus et al. (2012)). Finally, we use causal forest, a decision-tree based method that allows us to estimate treatment effects for every unit based on its best matched counterfactual unit (Wager and Athey (2018)). We obtain consistent results with all three methods. We are currently running a larger experiment to help us gain precision on the full sample.

2.2. Observational data from Airbnb.com

In the second step, we look at 'From...' prices on Airbnb.com. This price is calculated based on the minimum price that appears on the calendar of a listing over the next 30 days. We then examine how this minimum price affects bookings. We collect listing-level calendar data on Airbnb (scraped weekly for 2 months in 2015), with each scrape being associated with a

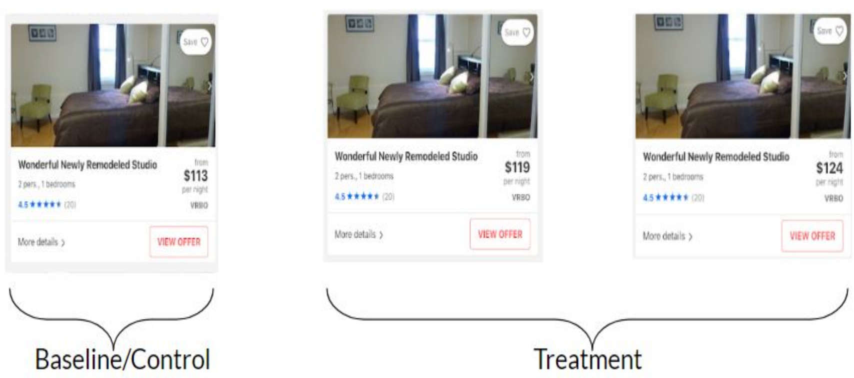

Figure 1: Example of experimental manipulation on Holidu.com.

minimum price. We have detailed information on roughly 23,000 listings in the greater Boston and New York area, and static as well as dynamic characteristics (location, size and number of rooms being offered, accommodation type, prices, and ratings). In total, we have 91,268 unique listing-scrape combinations. The identifying variation comes from scrape-to-scrape changes in From price for a given listing. Although we cannot observe consumer-level variables for this analysis, this setting allows us to look at the effect of From prices on the occupancy rate of hotels, since we have access to the entire calendar. This setting also allows us to model the difference between focal and From prices (which cannot be experimentally manipulated), and examine how this gap affects bookings. Our observational results indicate that, a larger From price, as well as a smaller relative difference between the displayed From price and the actual price paid, leads to both greater booking probability and higher occupancy rates, while controlling for focal prices and various time-invariant fixed effects, as well as different functional forms. To further alleviate concerns of endogeneity, we estimate an additional specification that compares listings who encounter a positive change in From price over the 2 month observation period with those who don't. The listings that see a change of greater than $5 are labeled as "treatment" listings, and matched with control listings based on a set of listing level characteristics (again, we use causal forest, and also robust Lasso to control for selection on observables). We find that treatment listings have a higher occupancy rate relative to control.

Taken together, these results lend evidence for the fact that consumers are guided by fairness perceptions, and are deterred if offered a very low initial From price, but a very high eventual price.

3. Results

3.1. Holidu

We estimate the effect of "Starting from..." prices on three main outcomes of interest. These are (1) outbound click (2) number of bookings and (3) total booking value. Given random assignment, our goal is to simply estimate the average treatment effect (ATE), i.e, differences in outcomes O for the treated group relative to the control group:

$$\Delta(O) = E[O_i|T_i = 1] - E[O_i|T_i = 0] \tag{1}$$

where i is a user and T_i indicates the treatment status of the user.

The results from this estimation are provided in Table 1. We see that there are significantly fewer outbound clicks in the treatment group. We fail to recover a significant effect on bookings, or booking value. However, booking value appears to be positive directionally, and has a low p-value.

To further investigate this effect, we look at booking value conditional on users who actually made a booking. The estimated equation is:

Table 1: The impact of a high From price on outbound clicks, bookings and booking value.

	Dependent variable:		
	Outbound	Bookings Per User	Booking value
	(1)	(2)	(3)
Pooled treatment (5% to 10%)	−0.023	−0.00000	0.158
	(0.004)***	(0.0001)	(0.140)
	p = 0.000	p = 0.977	p = 0.261
Constant	1.071	0.003	2.242
	(0.003)***	(0.0001)***	(0.114)***
	p = 0.000	p = 0.000	p = 0.000
Pre-treatment covariates	No	No	No
Observations	1,802,867	1,802,867	1,802,867
Adjusted R^2	0.00002	−0.00000	0.00000
Note:			*p<0.1; **p<0.05; ***p<0.01

$$\text{TotalBookingAmount}_i = \alpha + \beta T_i + X_i + \epsilon_i \tag{2}$$

where X_i = Region Country × Browser Family × Host + Host × Number Of Search In Journey × Device Type.

Given the high dimensional nature of the parameters, directly adding them to the equation could be subject to over-fitting. Hence, we make a correction as suggested by Chernozhukov et al. (2016). This allows us to recover the treatment effect of interest while controlling for "nuisance parameters" that tend to add noise to estimation. Their approach uses Lasso selection in a high dimensional setting to first identify a subset of relevant control variables, and then use OLS for estimation using the selected variables and the treatment indicator (for more details please refer to Chernozhukov et al. (2016)). Results are reported in Table 2. Column (1) shows results without any controls, Column (2) introduces direct controls and Column (3) implements R-Lasso. We find that, upon covariate adjustment, treatment users are significantly more valuable.

The main concern with the above specification is whether users are comparable, conditional on booking. In addition to controlling for confounders above, we check to ensure that there is no difference in the basic searching behavior of the users (as measured by the prices of apartments they view when searching with dates, and also the total number of apartments viewed - results not reported). Finally, we also implement the two matching strategies described in Section 2.1: CEM and causal forest. Results remain consistent (Table 3, Table 4 and Figure 2).

3.2. Airbnb

Our goal here is to estimate the impact of the "Starting from" price, over and above the focal price and other listing and time level characteristics. To do so, we make use of three key

Table 2: Total booking amount across conditions: (1) no controls, (2) controls and (3) RLasso

	Dependent variable:		
	totalBookingAmount		
	(1)	(2)	(3)
Pooled Treatment (5% to 10%)	31.340	63.336	51.59
	(27.811)	(24.892)**	(23.5)**
	p = 0.260	p = 0.011	p = 0.028
Pre-treatment covariates	No	Yes	Yes - rlasso
Observations	6,299	6,299	6,299
Note:	*p<0.1; **p<0.05; ***p<0.01		

Table 3: CEM estimates for the effect of From prices on total booking value, within users who booked.

	Dependent variable:
	TotalBookingAmount
Pooled Treatment (5% to 10%)	68.511**
	(27.202)
Constant	751.743***
	(21.932)
Note:	*p<0.1; **p<0.05; ***p<0.01

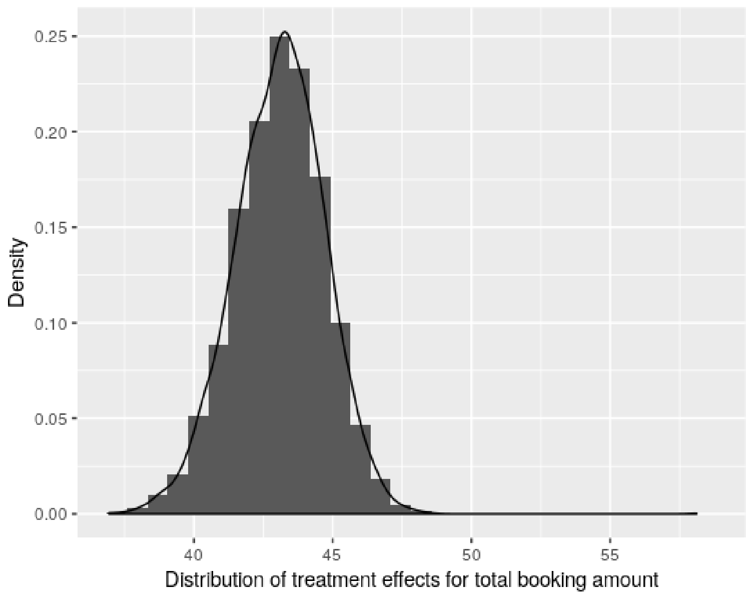

Figure 2: Plot of causal forest estimates of the individual treatment effects (total booking amount) for Holidu, within listings that are booked.

Table 4: Causal forest estimates for the effect of From prices on booking amount, within users who booked. Estimates somewhat noisy (Significant at ~ 10% level), but all individual level estimates are positive (refer to Figure 2).

	TotalBookingAmount
Estimate	48.23
Std. Error	25.78

independent variables (denoted by From_{it} in the equation):

1. the "Starting from...' price associated with a given listing i and scrape t (Price_{From})
2. the absolute difference (Diff) between the focal and the "Starting from..." price for every listing i, scrape t and calendar date j and,
3. the relative difference (Frac.Diff = $\frac{Diff}{Price}$) between the focal and the "Starting from..." price for every listing i, scrape t and calendar date j.

Now, we compute the occupancy rate (defined as the fraction of booked nights out of all available nights). The unit of analysis here is aggregated at the listing-scrape level. Hence, for every listing i and every scrape instance t, we take averages of the variables of interest over all calendar dates j. We include fixed effects for listing, month of scrape and month of calendar, to account for time invariant characteristics and transient time shocks respectively. Standard errors are clustered at the listing level.

The equations estimated are of the form:

$$\text{OccupancyRate}_{it} = \beta_1 \text{From}_{it} + \overline{\text{Price}_{it}} + \alpha_i + \delta_t + \epsilon_{it} \tag{3}$$

where From_{it} is either (1) the average Diff for a given scrape, (2) the average Frac.Diff for a given scrape, or (3)Price_{From} for a given scrape.

The results are reported in Table 5. We see that a larger From price, as well as a smaller relative difference between the displayed From price and the actual price paid, leads to higher occupancy rate. Column (1) uses $\overline{\text{Price}}$ and Price_{From} as independent variables; column (2) looks at their mean absolute difference (by scrape), and column (3) looks at their mean difference relative to Price (by scrape). We also conduct the above analysis for booking probability at the calendar date level, and find similar results (not reported).

Note that the above analysis only uses within listing variation in From prices, which is quite small in a two month period (SD = \$5). This also partially accounts for the small effect size. Hence, we now report results from an across-listing identification strategy. We pick listings that experience a large change in From prices (\$5 and more) as the "treated" listings. To ensure comparability between treated and control listings, we add covariates for (1) neighborhood, (2) number of beds, (3) bed type, (4) city, (5) whether instant bookable, (6) price for an extra person, (6) number of reviews, (7) number of pictures and (8) person capacity. We could not use star

Table 5: A larger From price, as well as a smaller relative difference between the displayed From price and the actual price paid, leads to higher occupancy rate (calculated as the fraction of total available calendar days that are booked).

	Dependent variable:		
	Occupancy Rate		
	(1)	(2)	(3)
$\overline{\text{Price}}$	-0.0002	0.00000	-0.00000
	$(0.00004)^{***}$	(0.00002)	(0.00002)
	p = 0.00000	p = 0.887	p = 0.823
Price_{From}	0.0002		
	$(0.00003)^{***}$		
	p = 0.00000		
$\overline{\text{Price - Price}_{From}}$		-0.0002	
		$(0.00003)^{***}$	
		p = 0.00000	
$\overline{\frac{\text{Price - Price}_{From}}{\text{Price}}}$			-0.068
			$(0.011)^{***}$
			p = 0.000
Listing Fixed Effect	Yes	Yes	Yes
Scrape month Fixed Effect	Yes	Yes	Yes
Observations	89,491	89,491	89,491
Adjusted R^2	0.147	0.147	0.148
Note:			*p<0.1; **p<0.05; ***p<0.01

rating as a control because Airbnb only releases ratings once listings get more than 3 reviews (which most listings in our sample do not). However, conditional on the above covariates, the mean price across treatment and control listings are not significantly different. Thus, we can argue that we are capturing underlying listing quality with these variables. The results are reported in Table 6 - column (1) has no controls and column (2) implements RLasso. Again, we find that treated listings (which experience an increase in From price) have a higher occupancy rate relative to control. Causal forest estimates are also consistent (0.0027 with a standard error of 0.0014 - refer to Figure 3 for the distribution).

Table 6: Occupancy rate for "treated" listings with a positive change in min price ($\geq$ 5) vs those with smaller or no change.

	Dependent variable:	
	Occupancy Rate	
	(1)	(2)
Treated	0.005	0.003
	$(0.002)^{***}$	$(0.0015)^{*}$
	p = 0.003	p = 0.058
Listing level covariates	No	Yes -rlasso
Observations	20,396	20,396
Adjusted R^2	0.0004	0.009
Note:	*p<0.1; **p<0.05; ***p<0.01	

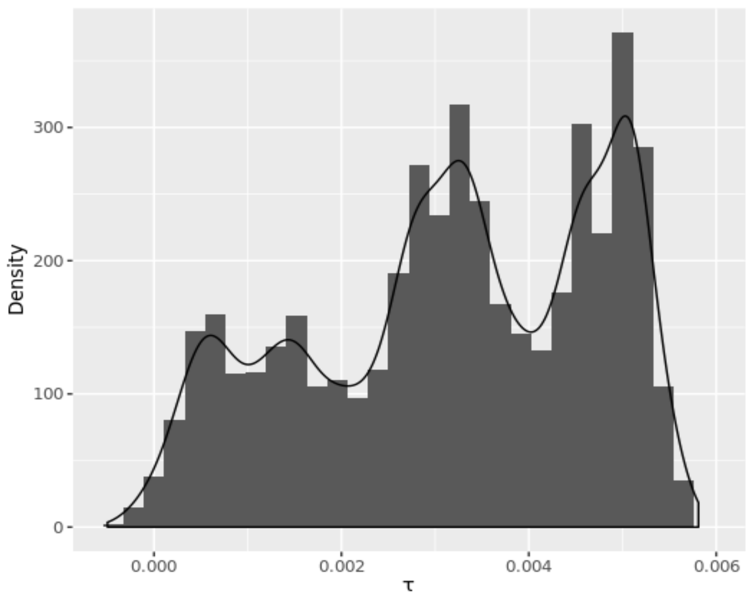

Figure 3: Plot of causal forest estimates of the individual treatment effects for occupancy rate.

4. Conclusion

We demonstrate that From prices have an impact on consumer booking behavior in an online vacation rental setting. Although high From prices might deter outbound clicks, they can positively affect booking value and occupancy rates. This points to the fact that consumers may be guided by fairness perceptions, and although they may click out, they would be dissuaded from purchasing if confronted with a low initial price, but a high eventual price. Hence, a "click-bait" approach to price advertising might not always lead to favourable outcomes.

References

Belloni, A., Chernozhukov, V., and Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies*, 81(2):608–650.

Chernozhukov, V., Hansen, C., and Spindler, M. (2016). High-dimensional metrics in r. *arXiv preprint arXiv:1603.01700*.

Iacus, S. M., King, G., and Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political analysis*, 20(1):1–24.

Jindal, P. (2018). Reference dependence and price negotiations–the role of advertised reference prices.

Monroe, K. B. (1973). Buyers' subjective perceptions of price. *Journal of marketing research*, pages 70–80.

Wager, S. and Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523):1228–1242.

Winer, R. S. (1986). A reference price model of brand choice for frequently purchased products. *Journal of consumer research*, 13(2):250–256.