Airbnb Optimal Pricing: Combining Structural Modeling and Machine Leanrning

Zheyu Ni

May 2023

1 Goal: Optimal Pricing for Airbnb: Smart Pricing?

The general machine learning approach will be training dataset with target Y which is the price in our case and its corresponding features X extracted for each listing, e.g. number of bedrooms, whether it is a part of the property or whole property, available amenity and etc. With the Airbnb dataset, we could extract thousands of property features and use the trained model to predict the price for a new listing with specific features.

The potential drawbacks of this approach:

- The model doesn't fully capture the demand/supply factors of the market. A high market demand could drive a high price, e.g. people's unobserved interest in the city gets higher, and festivals held in the city.

- A newly added listing could squeeze out demand for the existing listings. The machine learning model predicts based on the training set instead of the dynamic between listings.

But we do know how the general pattern of how people make the purchase: choose the most satisfied, and of how sellers decide the price: choose the price to maximize profit. At least, homeowners put their properties on Airbnb not to make losses and consumers won't choose a property that doesn't fit their needs. Structural models provide more insight into the data and enable us to specify specific relationships between the target and features. We could combine machine learning and structural modeling to make a more reasonable and effective prediction.

2 Model: Consumer Demand

Here is the general model for consumer demand. We observe c=1,..., C hot spots (markets) over t= 1,..., T time periods. There are j = 1,..., J_{ct} There are

$i = 1,..., N_{ct}$ consumers in each city-period. The indirect utility from booking a property $j \in J_t$ at time t in city c is

$$u_{jct} = x_j\beta + \alpha^{CO}p_{jct} + \sigma_j^D + \tau_t^D + \nu_c^D + \xi_{jct}$$

where x_j is a vector of observable listing features, p_{jrt} is the listing price, σ_j^D is the valuation of unobserved listing characteristics, τ_t^D is the time-specific valuation of unobserved product characteristics, ν_c^D is the market-specific valuation of unobserved product characteristics, ξ_{jct} is the unobserved quality valuation. ϵ_{jct} is the type 1 extreme value shock. α^{CO} and β parameters are the effect of price and features on consumer utility.

Denote $\delta_{jct} = x_j\beta + \alpha^{CO}p_{jct} + \sigma_j^D + \tau_t^D + \nu_c^D$ which is the mean utility of booking property j at time t in market c, the market share of this property will be

$$P_{jtc} = \frac{\exp \delta_{jct}}{1 + \exp \delta_{jct}} \tag{1}$$

which will be matched to the observed market share to recover model parameters α^{CO} and β.

Alternative to linear model: Lasso, or Random forests.

Given the recovered parameters, the price elasticities of the listing

$$\frac{\partial s}{\partial p} \cdot \frac{s}{p}$$

3 Seller's pricing

Given the demand estimates, each home-owner set prices for their properties to maximize the profit

$$\underset{p_j}{\text{Max}} \left\{ \sum_{j \in \mathcal{J}_t} (p_j - c_j) M \cdot s_j(\mathbf{p}) \right\}$$

where c_j is the cost of maintaining and cleaning the property j, $s_j(p)$ is the market share of property j given price p. M is the market size.

By maximizing the profit, the vectored first-order condition of the optimization problem for each market m will be

$$p_m = \underbrace{c_m}_{cost} \underbrace{- \left[\Omega_m^d \circ \left(\frac{\partial s_m(p_m; \theta_d)}{\partial p_m} \right) \right]^{-1} s_m(p_m; \theta_d)}_{markup} \tag{2}$$

Ω_m^d is the ownership matrix accounting for the fact that the owner may have multiple properties listed. $\left(\frac{\partial s_m(p_m; \theta_d)}{\partial p_m} \right)$ captures demand substitution patterns given demand estimates θ_d. It's showing that the price is the cost of maintaining + markup.

We are able to recover markup given demand estimation results and therefore recover cost c_m.

4   Optimal Pricing

Recompute the equilibrium price p^* after adding a new listing and forming a new listing profile J from:

$$p^* = mc - \left[\Omega_m^d \circ \left(\frac{\partial s(p^*; \hat{\theta}_d)}{\partial p^*}\right)\right]^{-1} s(p^*; \hat{\theta}_d)$$

which accounts for the substitution pattern between listings, property ownership, and costs of the listing. For the effect of ownership, consider the case where all properties are owned by one large institution/individual owner, the listing price will be significantly higher due to lack of competition.

We can then calculate the gain from combining a structural model vs the case using machine learning only.

5   Note

The first step will be clustering using Machine Learning: find the hotspot (markets). Once we have identified hotspot areas, we can compute the elasticity/profitability based on different hotspot areas. Another interpretation of this hotspot area will be the consumer's consideration set.