Scalable Analysis of Hedonic Prices and Local Competition in Hotel Room Prices

On a Large Set of Local Markets

by

Endre Márk Borza

Submitted to

Central European University

Department of Economics

In partial fulfilment of the requirements for the degree of

Master of Arts in Economics

Supervisor: Prof Gábor Békés

Budapest, Hungary

2019

I, the undersigned [Endre Márk Borza], candidate for the degree of Master of Arts in Economics at the Central European University Department of Economics, declare herewith that the present thesis is exclusively my own work, based on my research and only such external information as properly credited in notes and bibliography. I declare that no unidentified and illegitimate use was made of works of others, and no part the thesis infringes on any person's or intstitution's copyright. I also declare that no part the thesis has been submitted in this form to any other institution of higher education for an academic degree.

Budapest, 3 June 2019

Signature

Abstract

In this thesis, a large dataset of hotel room offers is compiled, from over 300 local markets, with at least 200 hotels for each one. Linear regression analysis of these markets shows remarkable heterogeneity, and the robust, price inflating effect of market power. It is suggested, that a lot of care is needed, when the predominant hedonic models in the literature are used to draw conclusions that go beyond a single local market. Further research can use the compiled data to explore the sources of this heterogeneity.

Table of Contents

1 Introduction

In 1997 a paper authored by Sala-i Martin (1997) was published by the National Bureau of Economic Research titled, ”I Just Ran Four Million Regressions”. Despite the somewhat cynical sounding title, it comes to rather optimistic conclusions, regarding economic growth literature. It claims that by looking at distributions of estimators for coefficients, useful insight can be gained that goes beyond testing for ”robustness”. This ended up being a very successful paper, cited over 3000 times, by papers in machine learning literature, or even the Journal of Clinical Epidemiology (Riaz et al., 2016).

Now, I only ran a couple of thousand regressions, neither of which have much to do with economic growth, but the gist is something similar. My goal is to look at hedonic pricing models of the same product but different markets, and find patterns in the behavior of the models, based on some attributes of the market. So I will also look at the distributions of model results, however I am not specifically looking for effects of variables on pricing, but the systematic differences in market behavior.

The product I will be looking at is a 1 night stay for 2 people in an online advertised accommodation - mostly hotels. The different markets are different cities, where one can purchase these. The choice is mostly based on the availability of data, and the ease of market specifications. Of course, it is possible for hotels from entirely different cities to compete with each other, and there might be hotels in the same city which have never been considered by the same potential guest. However, considering the need for a large sample of different markets, and the availability of vast booking agencies with precise data on pricing, community reviews and a lot of - hopefully reliable - information about hotels and rooms, this seemed like easily my best bet.

Running and comparing such a high number of regressions requires a great deal of automation and lots of data, in a uniform way. Evaluating or discarding models, handling outliers and missing values can not be done with human supervision for all markets, so a reliable and reproducible code-base is needed.

In the following, I go over the related literature, present the data, outline the models and draw some conclusions.

2 Local Market of Hotel Rooms

2.1 Hedonic Price Functions

The use of hedonic price functions based on the work of Lancaster (1966) expanded on by Lucas (1975) and Rosen (1974), is quite prevalent in hospitality pricing literature. It is assumed that a consumer's utility is determined by a vector of characteristics ($\mathbf{z} \in \mathbb{R}^n, \mathbf{z} \geq 0$), obtained by consuming certain products: $U(\mathbf{z})$. This utility is maximized in a non-linear intractable program, where goods available to the consumer ($\mathbf{x} \in \mathbb{R}^m, \mathbf{x} \geq 0$) are results of a linear transformation of an activity vector ($\mathbf{y} \in \mathbb{R}^k, \mathbf{y} \geq 0$ using an arbitrary matrix ($A \in \mathbb{R}^{k\times m}$). The characteristics are related to the consumed product indirectly, as they are also the result of a linear transformation ($B \in \mathbb{R}^{k\times n}$) of the activity vector. So with a linear budget constraint, the problem is:

$$\max U(\mathbf{z}) \text{ w.r.t.}$$
$$\mathbf{x} = \mathbf{y}A$$
$$\mathbf{z} = \mathbf{y}B$$
$$\mathbf{px} \leq I$$
$$\mathbf{x}, \mathbf{y}, \mathbf{z} \geq 0$$

Clearly, if exists, then in equilibrium, through $\mathbf{y}$, a linear connection can be drawn between $\mathbf{x}$ and $\mathbf{z}$, thus the effect of a good on a consumer's welfare can be deduced to factors of $\mathbf{z}$. With relatively simple assumptions and pure competition, this problem can be expanded to one, where market clearing $\mathbf{p}$ prices can be expressed as functions of $\mathbf{z}$.

Hedonic price functions suggest a very convenient empirical method for assigning value for certain characteristics of a family of products. In hospitality literature, one of the earliest examples of this is Carvell and Herrin (1990) who put a regression on cross-section data of San Francisco hotel room prices with a sample size of 567 (from 20 different hotels) for 9 attribute variables like AAA rating or a presence of a gym and achieved a 60% R^2. The only location specific variable there was a measure of distance from the city centre, as of course all hotels were from the same city. Not much later, Bull (1994) uses the same tools to focus more clearly on location, and concludes on a sample of 15 Australian motels that, while the exact effect is unclear, the way to go to measure the effect of location for hotel room pricing is the hedonic regression. Later, "chain association", meaning whether a hotel belongs to some brand-name or not, became one of the characteristics, as Wu (1999) and Israeli (2002) use the same method, again with mainly inconclusive results, but with larger data. Israeli used a dataset of 215 hotels from 9 locations with over 30000 rooms in Israel. For quite a while, these means are used in papers such as Monty and Skidmore (2003) or Thrane (2007), to generally find hedonic factors influencing prices in either bed and breakfasts in Wisconsin or hotels in Oslo, respectively. Zhang, Ye and Law (2011) and Abrate and Viglia (2016) utilize online resources of community hotel reviews and aggregating sites like tripadvisor and booking.com. They both note the trend of shifting from official hotel ranking and rating measures, to community based, online reviews in relevant characteristics. Corgel, Liu and White (2015) uses hedonic models to estimate determinants of hotel property prices on a sample of hotel transactions from 2005-2010.

Contemporary sharing economy pricing models also tend to pick the hedonic regression as a method. Gibbs et al. (2018) shows that such regressions on Airbnb accommodations behave similarly to hotels and bed and breakfasts, on a sample of 15,716 listings from Calgary, Montreal, Ottawa, Toronto and Vancouver. However, the authors concede that incorporating competition in the models is not possible, which is a major limitation. Chen and Xie (2017) uses 5,779 listings from Austin, TX to conclude that consumers tend to value intrinsic attributes of Airbnb listings, room-types and utility-bearing attributes, while less so social factors, with location playing some effect, but not highlighted. Wang and Nicolau (2017) uses a large sample of 180,533 listings from 33 cities in Europe North America and Australia. They identify 25 variables in the categories of host attributes, site and property attributes, amenities and services, rental rules, and number of online reviews and ratings, and find that all of them play an important role in determining the price

As mentioned in a great number of these papers from Rosen (1974) to Gibbs et al. (2018), the effect of competition is essentially neglected in these models. Also, most of them deal with one local market alone, so the results should not be distorted too much because of this, however generalizations are very hard to draw. Even though this approach has been linked to location as early as Rosen (1974) citing Tiebout (1956) and comparing these price-determining characteristics to public goods available in certain neighborhoods, and a number of papers treat location as an important characteristic, this location is always in relation to some unrelated utility-bearing resource. So the effect of co-location is neglected in this line of research.

2.2 Local competition

Simultaneously involving location and competition has a less robust history in the related literature. The theoretical fundamentals laid down by Hotelling (1929) and Porter (1998) don't show a clear path to empirical methods working well with simple regressions extending hedonic models, that dominate the field. The Hotelling model seems compelling as it incorporates location, competition and pricing in a fairly mathematical fashion, however in the hospitality business, application is not straightforward. In the Hotelling model, location is analogous with consumer taste and basically represents additional cost of moving from one place to another. This transforms well enough to a one dimensional distance measure from some arbitrary popular resource, as a characteristic involved in a hedonic regression, but it's a little more complicated to deduce features for concentration of hotels in a two dimensional space. The similarity of physical space and other attribute space can be utilized nevertheless, as differentiation can be quantified in a number of ways.

Porter (1998) introduces clusters, to address the paradox of technology and globalization decreasing the importance of locations and the continued prevalence of local concentrations of companies. On the one hand, some elements of the theory can apply to hotels, as geographical and cultural resources are highly localized and can lead to clusters of hotels forming. On the other hand consumption of goods of hotels - nights in rooms - is also localized, in contrast with manufacturing or a large chunk of service companies, which discourages co-location. No clear empirical or mathematical method is suggested though.

Some, though attempt to bridge the gap between local competition research and hedonic hotel room pricing models. Becerra, Santaló and Silva (2013) examine product differentiation strategies in the case of hotels. They find that both vertical and horizontal product differentiation is effective in the Spanish hotel market for reducing the need of large off-season discounts or allowing higher prices in general. Urtasun and Gutiérrez (2006) analyze this using an annual tourist guide book of Madrid to find that Hotels opening next to each other tend to be differentiated in other dimensions. Lee and Jang (2012) study seasonality and find it to be not uniform. They conclude that in Chicago, hotel prices have higher peaks and lower lows in the season, due to intense competition in the central business district Lee (2015) examined with a 2SLS method the extent of price competition among hotels in a georgaphical, and quality-differentiation sense, and found that, hotels that are quality-undifferentiated compete with a wider set of hotels on a geographical scale. Also suggested possible cooperation among hotels similar both in quality and location. Enz, Canina and Liu (2008) found that in the US, based on a sample of 15 000 hotels, hotel quality creates an externality that shows up in the pricing of hotels that don't necessarily fit a location profile. Meaning that low quality hotels can charge a premium for being located near high quality hotels. European hotels seem to generally overestimate the price flexibility of consumers in competitive, locations. As a sample of 3042 hotel observations over the two-year period, (1409 in 2006 and 1,633 in 2007) shows that maintaining higher rates and resisting the pressure to undercut competitors results in higher revenues (Enz and Canina, 2010)

2.3 Hedonic indices

Griliches (1961) claims, that up until that point, the main issue with price indices, namely that they don't account for quality change, has not been addressed in the literature. Then goes on tho lay down the foundations for a hedonic price index in the automotive industry. Later Goodman (1978) extends this model to housing, to analyze price differences of metropolitan and suburban markets. Over the years the method has been extended to a number of other markets, like spreadsheet handling software (Gandal, 1994) or PCs (Pakes, 2003).

I will not attempt to create a hedonic index for the hospitality market, but I will rely on similar methods to account for qualitative differences in the markets.

3 Data

3.1 Sources

Data is gathered using web crawlers from hotels.com. Hotels.com is a booking site that allows users to search for accommodation in a great number of regions, with detailed information about the hotels and rooms. More importantly, it also provides prices and booking opportunities.

The method of the collection was in compliance with the robots exclusion standard of hotels.com[1] found at `http://hotels.com/robots.txt`. At first, all hotel metadata was downloaded from links provided by the sitemap shown in robots.txt. These are over 600,000 links for information about hotels. At the time of downloading (2019-04-20), 641,082 pages were available, all of them were scraped for hotel level information. The detail of this data will be discussed in the next chapter, however one important attribute of hotels relevant here is the destination-id, on which the next phase of data collection is based.

In the second phase of data collection, room prices are collected, for groups of hotels expected to be competing in the same local market. Destination-id is an attribute of hotels, given by hotels.com, which is used here to determined local markets. This assumption is based on the fact that searching on hotels.com is mainly based on destination-id, so the hotels sharing the same destination-id are displayed on the same search results page, effectively providing options for the same intended purchase for customers. Hotels are filtered for destinations with at least 200, but not more than 500 hotels. 306 destinations fit this, with

[1] Protocol downloaded at 2019-04-20 does not prohibit `/ho*` or the Sitemap access for any User-Agent, and only these were accessed automatically as seen from the code.

a total of 90878 hotels. For these selected hotels, 2 dates in July, August and September are chosen, and room prices are queried for all of them. This results in $90878 \times 6 = 545268$ queries. [2]

Most hotels offer more than one room for each date, however a relevant portion of hotels do not list rooms at all, for any of the dates. This might be due to the fact that at one point these hotels registered to hotels.com, but for some reason decided to withdraw or restrict the bookings they make through the site. How these hotels are selected, or more generally, how the number of available rooms for a date is determined, is based on hotel attributes.

For all dates and all hotels, prices are requested for one night with two adult guests. All different offers are collected, with attributes of rooms, as well as attributes of the offer, like the cancellation policy, inclusion of free breakfast and so on.

The data generation process is dependent on the connection between the accommodation provider and the website, as well as how hotels.com collects related data from location and review services. It is not clear, how the data provided by the accommodations is validated, but as both internal (hotels.com) and external (tripadvisor) review statistics are present, some community based validation can be expected.

3.2 Transformations

In its raw format, most of the data is text based, and relates to attributes of a hotel, or a room. Some of these can be converted to dummy variables, like whether there is a minibar, or an outdoor pool. However, in many cases, the data relates to some quantifiable measure like number of floor or rooms, or size of conference room. These attributes are parsed automatically, as if there are some where the only difference between two attributes is a

[2] if a query took one second, which for a single core in a single computer, with regular internet connection, it almost does, it would take about a week to run all this, so I needed to set up a cluster on aws to do this, which might not be appreciated by hotels.com, but according to their policy, it does not seem to be forbidden either

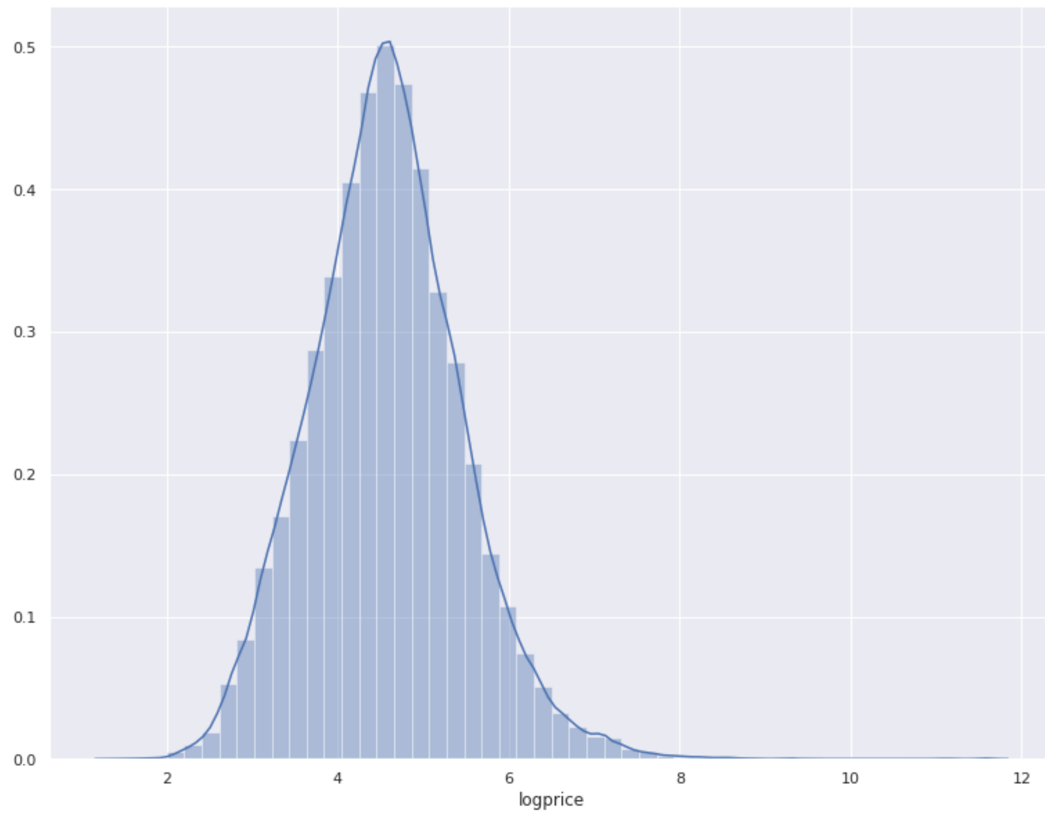

Figure 3.1: Distribution of log-price

number in the text, the text is replaced with a placeholder, and the number is taken as the variable.

Location information of hotels in the original data is restricted to coordinates and the destination-id. The coordinates are transformed to hotel-level attributes in the form of average distance from other hotels with the same destination-id, and to destination-level attributes in the form of median distance of two hotels from each other in the destination.

A number of variables, describing characteristics of the markets were constructed. To measure quality concentration, for each destination, the correlation between physical distance, price, star-rating, and tripadvisor review score difference is measured. For market share concentration, the rate of hotels offering at least 50% of rooms in the market is measured, as well as the difference between the prices for the top 5% and the rest.

Figure 3.1 shows the distribution of the target variable, the natural logarithm of average

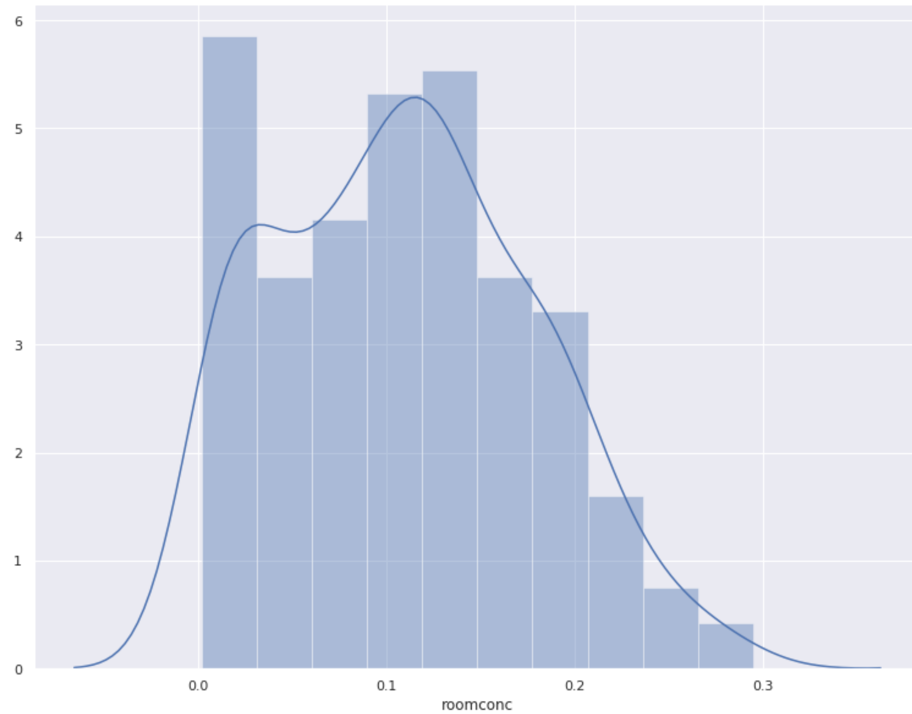

Figure 3.2: Distribution of room offer concentration

price, which has a few outliers, but other than that looks symmetric and without any noticable abnormalities.

In Figures 3.2, 3.2 and 3.4, the distributions for the market describing variables can be seen. We can see that the price difference doesn't seem to systematically correlate with geographical distance, and the other two distributions have noticeable peaks. In the case of room concentration, in a remarkably high number of cases, less than 3% of hotels are actually offering half of the available rooms. While the difference between the log of top 5% and the rest seem to be mostly between 1 and 2.5, so about 3-12 times.

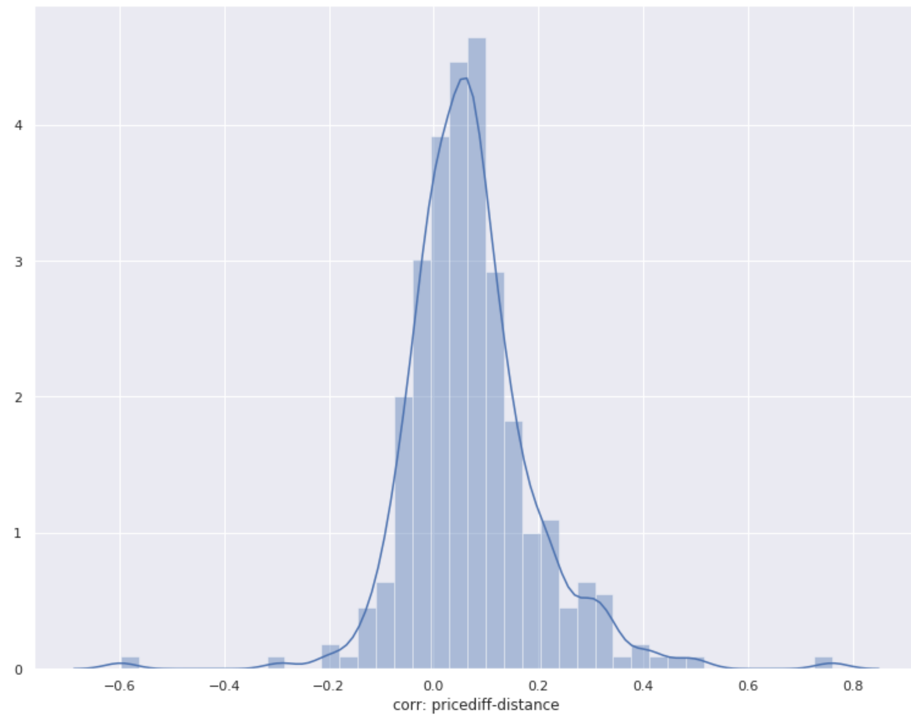

Figure 3.3: Distribution of price difference-physical distance correlations among destinations

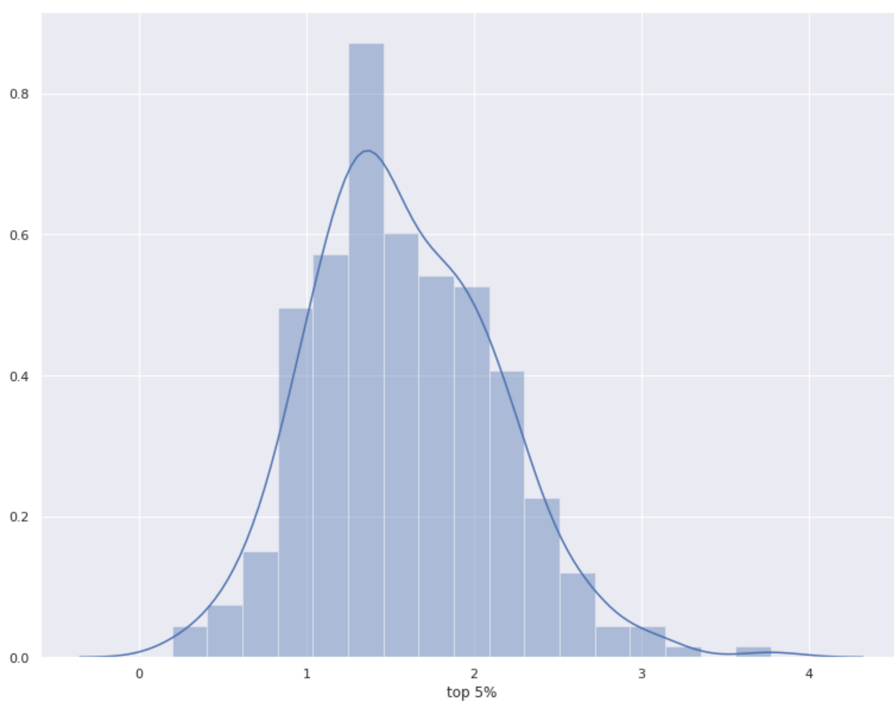

Figure 3.4: Distribution of price differences between the top 5% most expensive, and the rest (log difference)

4 Model

4.1 Model Definition

A number of regressions were ran for all destinations. The statsmodels (Seabold and Perktold, 2010) python package is used for running all regressions, with White (1980) standard errors corrected to $\frac{n}{n-p}(\mathbb{X}'\mathbb{X})^{-1}(\mathbb{X}' \operatorname{diag}(\widehat{u}_1^2, \ldots, \widehat{u}_n^2)\mathbb{X})(\mathbb{X}'\mathbb{X})^{-1}$ in MacKinnon and White (1985).

4.1.1 Determinants of a Regression

Filter Combination

A filter combination filters the rows of the final data-table. It is a combination, as the filters are based on more than one criteria. In this concrete case the criteria are:

- destination-id: data is filtered for each of the 306 destinations selected
- star rating: 3 possible filters are: 4+ star hotels, non 4+ star hotels, and all hotels
- location: 3 possible filters are the bottom 30 percentile in average distance from other hotels[1] within the city, the "central" hotels, the remaining 70 percentiles, "periphery" hotels, and all hotels

This gives $306 \times 3 \times 3 = 2754$ filter combinations, the specified further filtering of the location sets are mainly to check for robustness of variable coefficients in subgroups.

[1]distances are calculate

Independent Variable Set

For all filter combinations, a number of *fixed* variables are given, which are included in all regressions. In addition a selection from another pool of *additional* variable-sets are added.

- *Fixed* variable-sets
  - dummies for the dates of the queried room rates
  - star-rating: the star rating of the hotel
  - parsed reviews gathered by hotels.com for the accomodation (hotels.com and tripadvisor reviews, the hotels.com reviews can be split to a number of categories)
  - room attributes: these are general attributes for the specific room the hotel is offering
- *Additional* variable-sets
  - inhotel attributes, describing services of the hotel, like free wifi, or valet parking
  - base hotel attributes like the age of the hotel or the number of floors
  - location attribute, meaning the normalized distance from other hotels in the same destination for each hotel

4.1.2 Runtime Checks for Regressions

For each filter, a number of things are checked, and therefore the proposed regression is either processed, processed with some modifications, or discarded.

On the variable level, each independent variable is checked, which is proposed to be included in the regression. A variable is discarded if it either has 0 standard deviation, has over 95% correlation with another variable already selected to be in the regression, or has the same value over 92.5% of observations. Of course if the first condition holds, this last one doesn't need to be checked. Also, if a variable from a group of dummies is dropped, another one needs to be randomly selected to drop as well to avoid loss of rank in the matrix.

On the whole regression level, if degrees of freedom falls below 0.7 times the number of observations, or an entire predefined variable-set is dropped in the previous check, the regression is canceled.

4.1.3 Regression Statistics

The output of the above, fits into two tables. One, where each record is a regression, that passed the checks and ended up being processed, thus some statistics about it could be collected. The other is a table of logs, where each regression that was discarded, or the variables that were dismissed, get recorded, so that they are not processed again, and later model specifications can be aided.

The regression statistics output, contains regression and variable level information. Parameter values, heteroscedasticity robust standard errors and calculated p-values are included on the variable level. On regression level, adjusted R^2, Bayesian and Akaike information criterion, and five-fold cross-validated root mean squared error, plus degrees of freedom and number of observations.

4.2 Model Results

Figure 4.1 shows a scatter of all selected regressions. For all markets (with no additional filter), a *best* regression was selected, based on 5-fold cross-validated rmse, which seems to be nicely correlated with in-sample adjusted R^2 measures.

In these regressions, and the further filtered ones, parameters of the variables were evaluated, to determine hedonic factors for possible indices. However, only a handful of variables produced significant positive parameters in more than half of the cases, all of them describing some rating of the hotel - either reviews or star-rating, or the size of the hotel, like the number of rooms in it. While only one variable was negative significant in over 40% of cases, the room attribute for a room only having a shower.

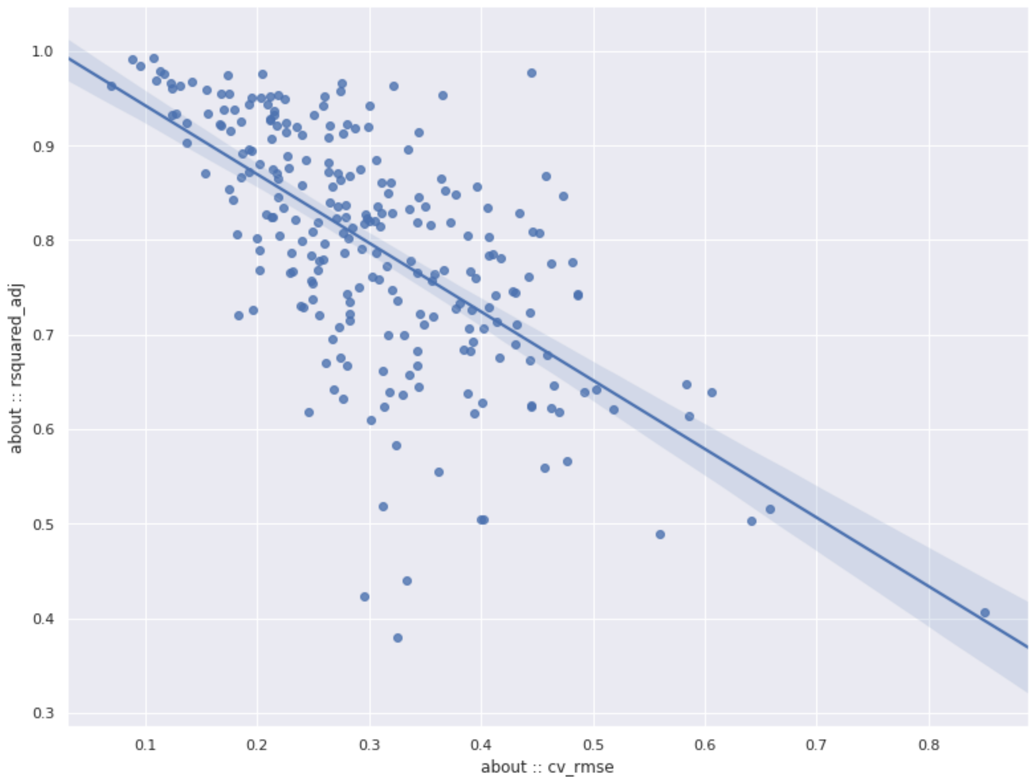

Figure 4.1: Scatter of adjusted r squared and 5-fold cross-validated rmse

This clearly limits the possibility of creating a useful cross-market hedonic index, but how the parameters of the reliably positive variables behave can still be inspected.

Figure 4.2 shows that the correlation between the average price in a market and the parameters of the variables supposedly providing the value in the hotels is remarkably small. This points to the conclusion that either in each market, some specific aspect of accommodation is valued, that can't be traced in other markets, and/or the pricing of hotels in different markets is clearly not determined by the different pricing of these hedonic attributes. Even if we separate the parameter for the most reliable predictor of price, the parameter of star-rating, in Figure 4.3, there is virtually no connection. These results show that a simple mean of log-price, and the parameters of generally utility bearing attributes of products in a hedonic model, show different things.

As seen from 4.2, the sum of the parameters is not even positive for these variables, in nearly half of the markets. However, as 4.1 shows, the models are accurate and not over-

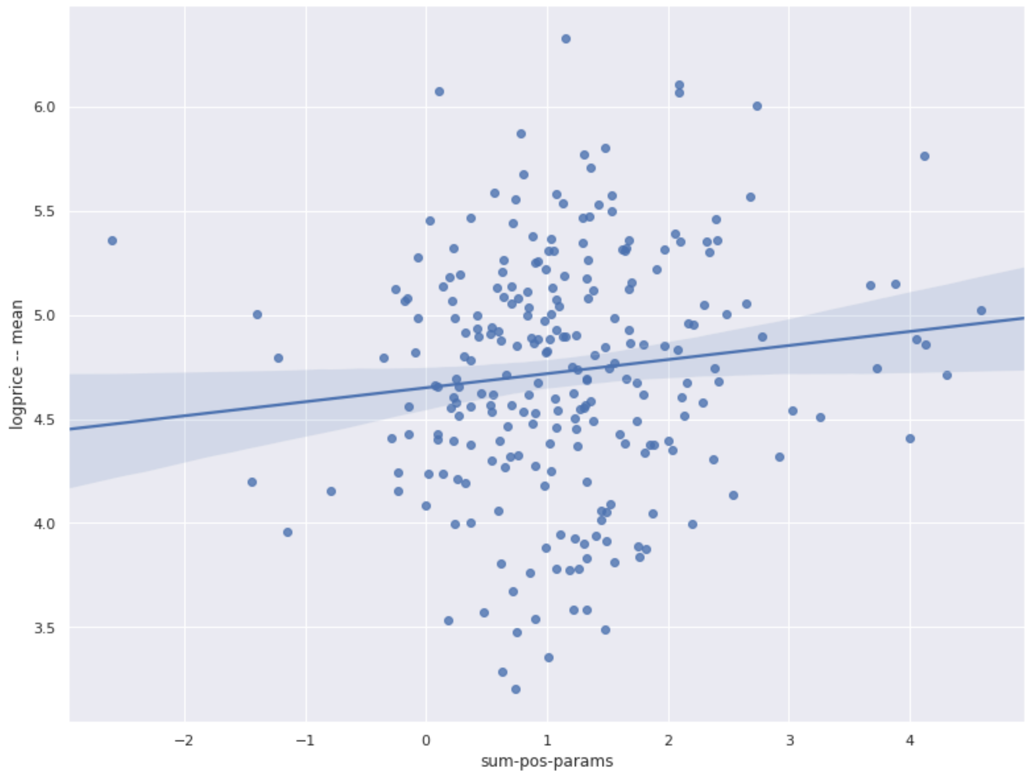

Figure 4.2: Scatter of sum of parameter values of reliably positive variables and mean log-price for each market

fitted. There are over 200 variables, and a few regression level statistics that could form any number of patterns, but as far as the scope of this paper goes, the heterogeneity of the local markets creates an extreme amount of noise, so that very little can be extracted.

Of the market features defined above, neither distance-correlations, nor top 5% pricing difference measurements seem to correlate with other attributes of the markets measured in the regressions. However, room concentration has a strong negative correlation with the mean price, while having seeming no connection with the parameters. This is shown in Figures 4.4 and 4.5. This is a rather interesting result, as the logic of market-power suggests that in the case of high concentration - as a lower portion of hotels in a market are offering the same portion of rooms - prices can be inflated. However, due to the serious heterogeneity of the markets, it is not clear what exactly costs more. The room concentration measure having a negative relationship with certain parameters would more precisely show where the

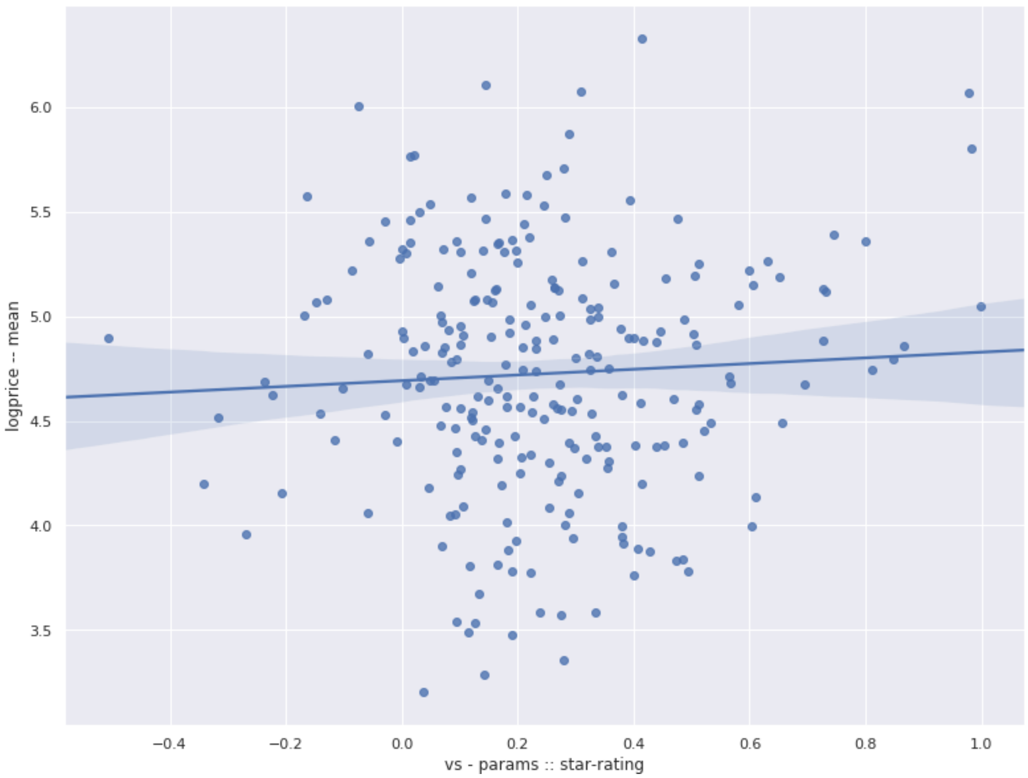

Figure 4.3: Scatter of parameter for star-rating and in sample mean log-price for each market

rent for market power goes, but this way, it simply seems to be independent of the identified variables.

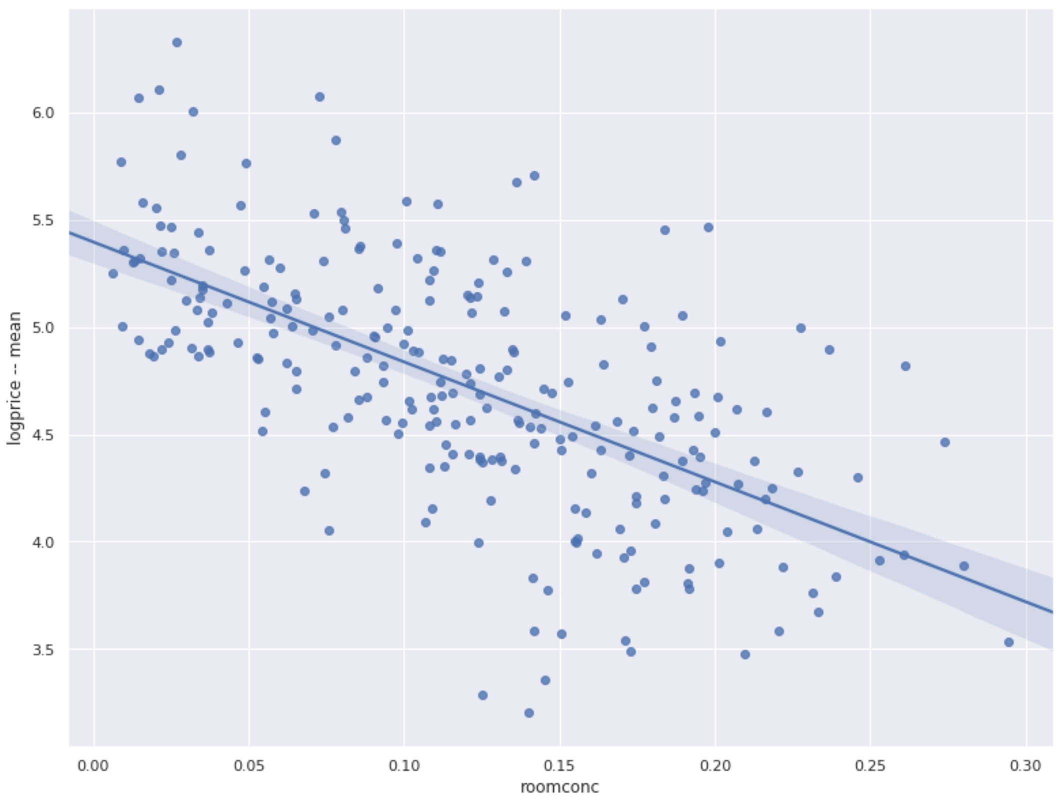

Figure 4.4: Scatter of ratio of hotels offering half of available rooms, and in sample mean log-price for each market

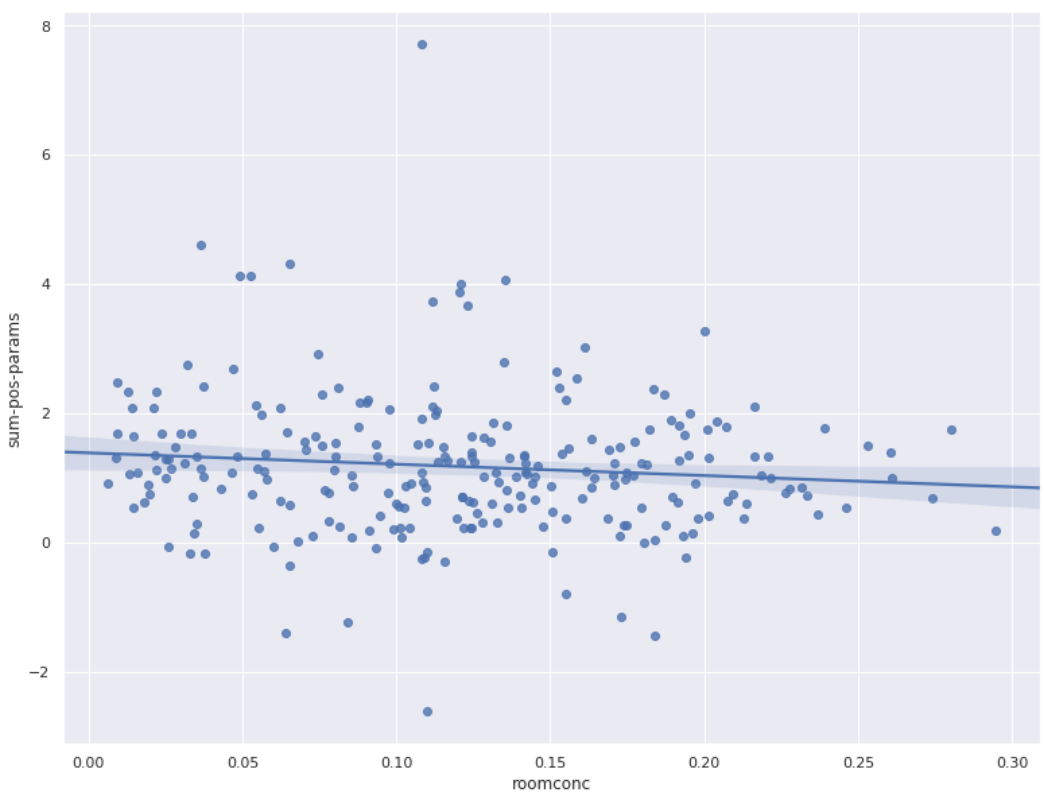

Figure 4.5: Scatter of ratio of hotels offering half of available rooms, and sum of parameter values of reliably positive variables

5 Conclusion

Running a few thousand regressions on a few million rows of hotel data with a couple of hundred features, there doesn't seem to be a straightforward method to create a cross market hedonic model, that dissects the global market for one night stays, to global markets of star-ratings and free breakfasts. Even the most reliable hedonic factors seem to lose their effects in a relevant portion of the markets, and parameters are nowhere near stable enough to determine prices for hedonic factors, they are barely stable enough to determine whether a variable is positively influencing price or not.

The hedonic models, that dominate the hospitality literature, can find very little ground in this data-set. This data can provide useful ground to examine how extrapolation might be possible from a subset of markets, it seems highly unlikely that it is possible form a single market.

Large hotels with many stars and good reviews can charge high prices, while markets with dominant hotels with high market shares, tend to be more expensive. These conclusions stand the heterogeneity of local hospitality markets, but not much else does.

References

Abrate, Graziano, and Giampaolo Viglia. 2016. "Strategic and tactical price decisions in hotel revenue management." *Tourism Management*, 55: 123–132.

Becerra, Manuel, Juan Santaló, and Rosario Silva. 2013. "Being better vs. being different: Differentiation, competition, and pricing strategies in the Spanish hotel industry." *Tourism Management*.

Bull, Adrian O. 1994. "Pricing a Motel ' s Location." *International Journal of Contemporary Hospitality Management*, 6(6): 10—-15.

Carvell, Steven A., and William E. Herrin. 1990. "Pricing in the Hospitality Industry: An Implicit Markets Approach." *Hospitality Review*, 8(2): 27–37.

Chen, Yong, and Karen L. Xie. 2017. "Consumer valuation of Airbnb listings: a hedonic pricing approach." *International Journal of Contemporary Hospitality Management*, 29(9): 2405–2424.

Corgel, Jack B., Crocker Liu, and Robert M. White. 2015. "Determinants of Hotel Property Prices." *Journal of Real Estate Finance and Economics*, 51(3): 415–439.

Enz, Cathy A., and Linda Canina. 2010. "Competitive pricing in European hotels." *Advances in Hospitality and Leisure*, 6: 3–25.

Enz, Cathy A., Linda Canina, and Zhaoping Liu. 2008. "Competitive dynamics and pricing behavior in US hotels: The role of co-location." *Scandinavian Journal of Hospitality and Tourism*.

Gandal, Neil. 1994. "Hedonic price indexes for spreadsheets and an empirical test for network externalities." *the RAND Journal of Economics*, 160—-170.

Gibbs, Chris, Daniel Guttentag, Ulrike Gretzel, Jym Morton, and Alasdair Goodwill. 2018. "Pricing in the sharing economy: a hedonic pricing model applied to Airbnb listings." *Journal of Travel and Tourism Marketing*, 35(1): 46–56.

Goodman, C. 1978. "Hedonic prices, price indices and housing markets." *Journal of Urban Economics*, 5: 471.

Griliches, Zvi. 1961. "Hedonic price indexes for automobiles: An econometric of quality change." In *The price statistics of the federal goverment.* 173–196. NBER.

Hotelling, Harold. 1929. "Stability in Competition." *The Economic Journal*, 39(153): 41–57.

Israeli, Aviad A. 2002. "Star rating and corporate affiliation: Their influence on room price and performance of hotels in Israel." *International Journal of Hospitality Management*, 21(4): 405–424.

Lancaster, Kelvin J. 1966. "A New Approach to Consumer Theory." *Journal of Political Economy*, 74(2): 132–157.

Lee, Seul Ki. 2015. "Quality differentiation and conditional spatial price competition among hotels." *Tourism Management.*

Lee, Seul Ki, and Soo Cheong Jang. 2012. "Premium or discount in hotel room rates? the dual effects of a central downtown location." *Cornell Hospitality Quarterly.*

Lucas, Robert. 1975. "Hedonic price functions." *Economic Inquiry*, June(2): 157—-178.

MacKinnon, James G., and Halbert Working White. 1985. "Some Heteroskedasticity Consistent Covariance Matrix Estimators with Improved Finite Sample Properties." *Journal of Econometrics*, 29(3): 305—-325.

Monty, Ben, and Mark Skidmore. 2003. "Hedonic pricing and willingness to pay for bed and breakfast amenities in Southeast Wisconsin." *Journal of Travel Research*, 42(2): 195–199.

Pakes, Ariel. 2003. "A reconsideration of hedonic price indexes with an application to PC's." *American Economic Review*, 93(5): 1578–1596.

Porter, Michael E. 1998. "Clusters and the New Economics of Competition." *Harvard Business Review*, , (November): 77–90.

Riaz, Nadeem, Suzanne L Wolden, Daphna Y Gelblum, and J Eric. 2016. "Assessment of vibration of effects due to model specification can demonstrate the instability of observational associations Chirag." *Journal of clinical epidemiology*, 118(24): 6072–6078.

Rosen, Sherwin. 1974. "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition." *Journal of Political Economy*, 82(1): 34–55.

Sala-i Martin, Xavier X. 1997. "I Just Ran Four Million Regressions." *National Bureau of Economic Research*.

Seabold, Skipper, and Josef Perktold. 2010. "Statsmodels: econometric and statistical modeling with Python." *9th Python in Science Conference*, , (Scipy): 57–61.

Thrane, Christer. 2007. "Examining the determinants of room rates for hotels in capital cities: The Oslo experience." *Journal of Revenue and Pricing Management*, 5(4): 315–323.

Tiebout, Charles M. 1956. "A Pure Theory of Local Expenditures." *Journal of Political Economy*, 64(5): 416–424.

Urtasun, Ainhoa, and Isabel Gutiérrez. 2006. "Hotel location in tourism cities: Madrid 1936-1998." *Annals of Tourism Research*.

Wang, Dan, and Juan L. Nicolau. 2017. "Price determinants of sharing economy based accommodation rental: A study of listings from 33 cities on Airbnb.com." *International Journal of Hospitality Management*.

White, Halbert. 1980. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica*, 48(4): 817—-838.

Wu, Lawrence. 1999. "The pricing of a brand name product: Franchising in the motel services industry." *Journal of Business Venturing*, 14(1): 87–102.

Zhang, Ziqiong, Qiang Ye, and Rob Law. 2011. "Determinants of hotel room price: An exploration of travelers' hierarchy of accommodation needs." *International Journal of Contemporary Hospitality Management*, 23(7): 972–981.