The Effect of Airbnb's Market Entry on Hotels in Berlin: A Regression Discontinuity Approach

Final Paper for Introduction to Collaborative Social Science Data Analysis

Daniel J Murphy & Paulo H Kalkhake

December 9, 2016

```r
# Dynamical Link to Data/Packages R script file (takes some time!)
source("code/finalpaper_analysis.R")
```

```
## Warning in read.table(file = file, header = header, sep = sep, quote =
## quote, : seek in einer gzfile Verbindung gab internen Fehler zurück

## Warning in read.table(file = file, header = header, sep = sep, quote =
## quote, : seek in einer gzfile Verbindung gab internen Fehler zurück
```

```r
##create a theme for all graphs (courtesy of Max Wolf)
fte_theme <- function() {

# Generate the colors for the chart procedurally with RColorBrewer
palette <- brewer.pal("Greys", n=9)
color.background = palette[2]
color.grid.major = palette[3]
color.axis.text = palette[6]
color.axis.title = palette[7]
color.title = palette[9]

# Begin construction of chart
theme_bw(base_size=9) +

# Set the entire chart region to a light gray color
theme(panel.background=element_rect(fill=color.background, color=color.background)) +
theme(plot.background=element_rect(fill=color.background, color=color.background)) +
theme(panel.border=element_rect(color=color.background)) +

# Format the grid
theme(panel.grid.major=element_line(color=color.grid.major,size=.25)) +
theme(panel.grid.minor=element_blank()) +
theme(axis.ticks=element_blank()) +

# Format the legend, but hide by default
theme(legend.position="none") +
theme(legend.background = element_rect(fill=color.background)) +
theme(legend.text = element_text(size=7,color=color.axis.title)) +

# Set title and axis labels, and format these and tick marks
theme(plot.title=element_text(color=color.title, size=10, vjust=1.25)) +
theme(axis.text.x=element_text(size=7,color=color.axis.text)) +
```

```r
theme(axis.text.y=element_text(size=7,color=color.axis.text)) +
theme(axis.title.x=element_text(size=8,color=color.axis.title, vjust=0)) +
theme(axis.title.y=element_text(size=8,color=color.axis.title, vjust=1.25)) +

# Plot margins
theme(plot.margin = unit(c(0.35, 0.2, 0.3, 0.35), "cm"))
}
```

Abstract

Introduction

Airbnb is a young company that has revolutionized the accomodation industry in much the same way that Uber has revolutionized transportation. It is an online platform that allows "hosts" to share their home or a rental home with travelers. For hosts, using Airbnb can be a valuable source of extra income. For guests, Airbnb provides accomodations that allows travelers to both save money and have a more authentic experience during their stay. Airbnb guests often remark that they interacted with their hosts and felt more like a local than a tourist. Indeed, Airbnb's recent "Don't Go There, Live There" advertising campaign reflects their philosophy in reshaping the accomodation industry.

In Berlin more than 20,000 hosts have been active over the past year, hosting more than half a million guests. According to Airbnb, their presence in Berlin brings more economic activity to the city, particularly to neighborhoods outside of the main hotel districts. Yet, Airbnb's entry in Berlin has been highly controversial. In 2014, the city passed a law that forbade renting out entire apartments to tourists on Airbnb. This came in response to an unprecedented 56 percent rise in rents in Berlin between 2009 and 2014, for which Airbnb and other online hosting platforms received some of the blame (France-Presse, 2016). The city's law was widely welcomed by Berliners hoping to slow the forces of gentrification, but it was also welcomed by another Airbnb victim in Berlin; the hotel industry.

In Berlin and around the world, the hotel industry has been quietly fighting a battle of their own against Airbnb. Airbnb represents an unprecedented challenge for the long-time incumbents, whose average price of 80 euros per night in Berlin cannot compete with the average Airbnb price of 55 euros per night (Skowronnek, Vogel, & Parnow, 2015). Moreover, Airbnb enjoys some real advantages over the hotel industry due to its new, still largely unregulated business model. For example, its informal business model has thus far allowed it to avoid the taxes and regulations that hotels and other traditional accomodation providers have to deal with (Guttentag, 2015). For their part, Airbnb claims that their service is complementary to hotels. They point out that most of their listings (77 percent in Berlin) are outside of cities' main hotel districts (Airbnb, 2015). Yet, many believe there must be some "Airbnb effect" on the hotel industry. A paper by Zervas, Proserpio, & Byers (2016) found that the rise of Airbnb had a negative effect on hotel revenue in the state of Texas. This shows that, while Airbnb may be complementary, they also compete with the incumbent hotel industry.

In this paper we will seek to illustrate the magnitude of the "Airbnb effect" on hotels in Berlin. To that end, this paper will proceed as follows. First, we will review the relevant literature in the field and outline our hypotheses going forward. Second, we will discuss the data sources and variables that we will use to explore this relationship. Third, we will outline the methology we use in this study and describe our model in detail. Fourth, we will analyze the data both descriptively and inferentially to evaluate our hypotheses. Finally, we will discuss our results, the policy implications of our findings, and opporunities for future research.

Literature Review

While the literature on the effect of Airbnb on hotels is quite limited, there is a diverse body of literature available on different aspects of Airbnb's business. In Zervas, Proserpio, & Byers (2015), the authors explore user ratings on Airbnb and TripAdvisor. They find that the average user rating of Airbnb is significantly higher than that of the average hotel on TripAdvisor, 4.5 out of 5 stars compared to 3.8 out of 5 stars, respectively. They theorize that Airbnb hosts, being more entrepreneurial than their hotelier competitors, go above and beyond to ensure that they recieve good reviews that bolster their individual "brand". Along similar lines, Gutt & Herrmann (2015) investigated how Airbnb hosts change their prices in response to their first review. They found that increased "rating visibility" causes hosts to raise their prices by an average of 2.69 euros.

In Berlin, much literature and discussion has focused on the impact of Airbnb on rents in the city. Schäfer, Braun, Reed, & Johnston (2016), for example, found that 5,555 residential apartments are currently being misused for Airbnb in Berlin. However, as the *Zweckentfremdungsverbot* (Airbnb misuse law) came into effect this year, the main focus of this issue is likely to be enforcement going forward.

Like Uber, Airbnb has also been widely praised for providing a more personalized experience for consumers in an industry that they might not have expected to find it before. In both Yannopoulou, Moufahim, & Bian (2013) and Guttentag (2015), the authors find that the desire for a personalized user experience has played a role in the rise of Airbnb and has changed the very nature of the accomodation and tourism industry. A working paper by Stors & Kagermeier (2015) shows that this trend extends to Berlin, and indicates that the motives for a tourist to use Airbnb rather than a hotel are both monetary and otherwise. In fact, the Stors & Kagermeier found that guests' expectations of having a more authentic, personal experience during their stay were just as important as their monetary concerns. This suggests that the hotel industry in Berlin is not only being undercut in pricing, but also outclassed in the experience they provide.

In Guttentag (2015), the author argues that while many points discounting Airbnb's ability to compete with the traditional accomodation industry are valid, there is still no reason to completely discount the young disruptor's impact. Airbnb will probably never completely replace the hotel industry, but it is already large enough to have an impact nevertheless. Indeed, Guttentag estimates that Airbnb sold about 15 million room nights in 2012, which would have made its footprint on the industry that year similar to that of Fairfield Inn & Suites or InterContinental, two major hotel brands. Guttentag concludes his paper with some recommendations for future research on Airbnb, and highlights the need to investigate the impact of Airbnb on hotel occupancy rates and room prices.

Along those lines, there is one prominent paper that helped inspire our investigation. In their paper, Zervas et al. (2016) used a difference in differences method (DID) to measure the impact that Airbnb's presence had on Texas hotel revenue, and how this impact differed by region and over time. They found that a 10 percent size increase of the Airbnb market in Texas resulted in a .39 percent decrease in hotel revenue. This effect varied widely by region, however. Austin, for example showed a 8 to 10 percent impact on hotel revenue. Zervas et al. also went on to investigate the impact of Airbnb's presence on different types of hotels and how that presence may effect hotel pricing models. They found that hotels with lower prices are more affected than higher priced hotels, and that Airbnb has significantly hindered hotels' ability to raise prices during high demand periods.

In order to add to this body of research, and at the behest of Guttentag (2015), our investigation will focus on the effect of Airbnb on hotel occupancy rates in Berlin. Specifically, two hypotheses will guide our thinking.

> *H1: The higher the Airbnb supply in a given district in Berlin, the lower the hotel occupancy rate will be in that same district.*

> *H2: Since Airbnb's listings are concentrated in districts with low hotel density, the effect of substitution will be more pronounced in those districts.*

Data & Variables

Our data comes from three different sources, the Statistical Information System Berlin/Brandenburg (SBB) (StatIS-BBB, 2016), the Federal Statistical Office and the statistical offices of the Länder (FSO)[1] (Germany, 2015), and *InsideAirbnb.com* (Cox, 2016). In order to conduct our analysis we needed to clean, merge and manipulate these data sets.

From the Statistical Information System Berlin/Brandenburg, we collected monthly data on the number of overnight stays and the number of guests to arrive at their accommodations in the reporting period in Berlin (StatIS-BBB, 2016). The surveys are carried out at the beginning of each month and refer to the reporting period of the previous month. The results are organized regionally according to districts and municipalities, allowing us to use specific data for each of the twelve districts in Berlin. We also gathered data for yearly household income groups and the number of employed and unemployed people per district. Based on the data, we calculated a yearly average household income and unemployment rate per district.

[1] Both databases use JAVA-based website, which did not allow direct web scraping. The data was manually downloaded.

OVERNIGHT VARIABLE

From *InsideAirbnb.com*, we scraped data on 15,368 listings, i.e. apartments or rooms, for Berlin from August, 2008 until October, 2015. This data was extracted from the Airbnb site between 18 July 2015 and 6 January 2016. Amongst 92 variables for each listing covering topics ranging from room price to information on the host, the data includes (1) the neighbourhood of each listing, (2) the date that an Airbnb host signed up, and (3) the date of the first review of each listing. We further include a proxy for Airbnb's market entry. It is based on date of of acquisition of the German peer-to-peer market place Accoleo in June 2011, which is considered Airbnb's first move to expand their business model to the European market (TheNextWeb, 2011).

Our most significant methodological challenge is the absence of precise listing availability during our period of interest, as it is not directily available in the data. However, in keeping with the Zervas et al. (2016) methodology, we used a date 6 months prior to the first review as a proxy for market entry. We then construct a variable for cumulative supply based on this information, i.e. the total number of listings in a district which have had their first reviews six month before to that month. This is not ideal, as this proxy does not take into account whether or not a listing was available in a given month. However, Airbnb itself is also unable to produce exact supply data. This is because owners do not accurately update their listings' availability.

The final data set covers 720 monthly obversations across tewlve districts in Berlin and covers the time period between 2010 and 2014.

Methodology & Analysis

Descriptive Analysis

Airbnb‘s popularity amongst users and hosts has increased substantially since it was founded in August 2008. As Berlin's popularity has increased over that same period, many Berliners began listing apartments on Airbnb. The growth in Airbnb listings has not been equal across all neighbourhoods in Berlin, but it has been positive in every year from 2010 to 2014. That trend has accelerated each year, with more and more Berliners listing apartments on the site.

Upon plotting Airbnb supply against hotel room occupancy, we were surprised to see a clear positive correlation. However, we realized the importance of accounting for general demand in an area as it becomes more attractive to tourists. Zervas et al. (2016) accomplished this by controlling for passengers listing the local airport as their final destination. We will incorporate a similar statistic in our final analysis, as tourism in Berlin has increased dramatically in recent years. Given confirmation from Zervas et al. that Airbnb does indeed compete with hotels, we believe that this increase in Berlin's popularity is largely responsible for the positive correlation we observe here.

Inferential Analysis

Following Zervas, Proserpio, & Byers (2013), our analysis takes advantage of the fact that Airbnb offically entered the German market in June 2011 with the acquisition of Accoleo, and also grew differently depending on the neighbourhood. By introducing a binary variable which is equal to one for all observations after Airbnb's market entry in Berlin and interacting that variable with all regressors in our model, we account for all the changes in hotel occupancy rates before and after Airbnb's market entry in Berlin, i. e. the dicontinuity at market entry.

> The key identification assumption we have to make to support a causal interpretation of this DD estimate is that there are no unobserved, time-varying, city-specific factors that are correlated with both Airbnb entry and hotel room revenue.Stated differently, we assume that unobserved factors that could potentially jointly affect both Airbnb adoption and hotel room revenue do not systematically vary both between different cities and over time.

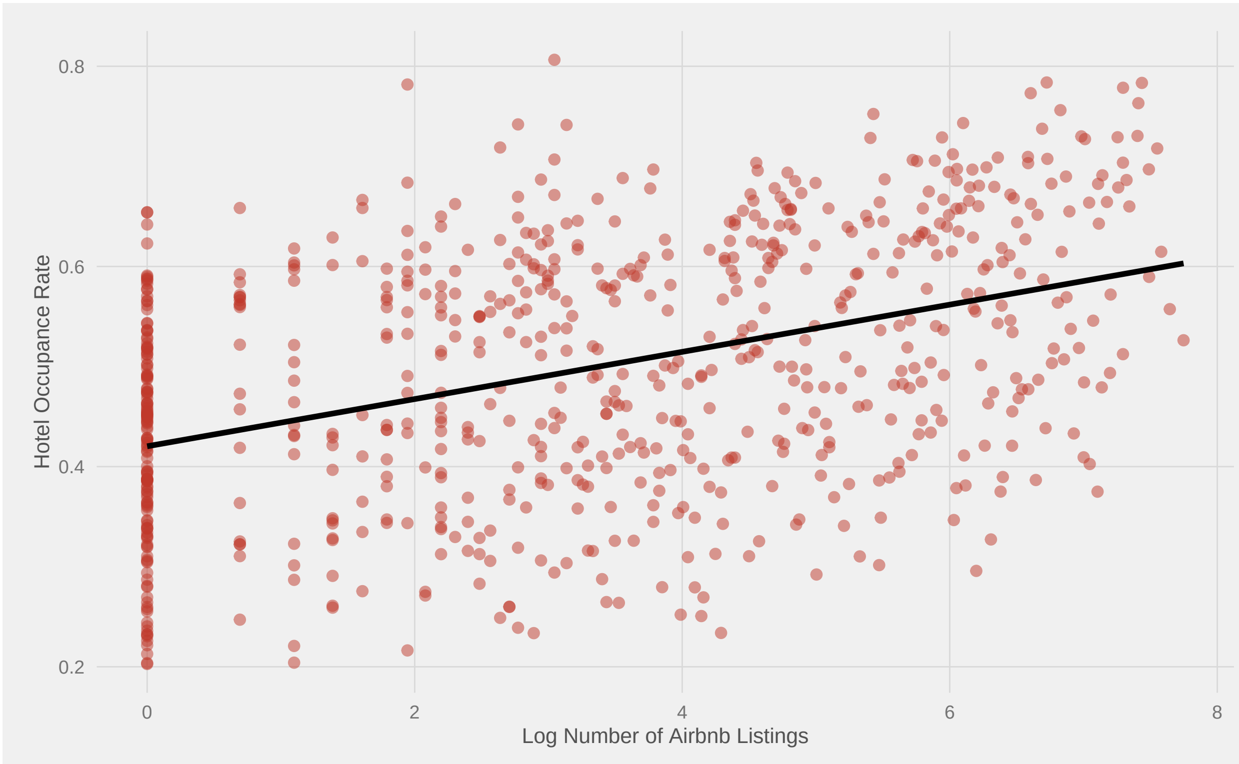

Figure 1: Effect of Change in Airbnb Listings on Hotel Occupancy Rates in Berlin (2010 - 2014)

Unlike Zervas et al. (2016), we cannot use hotel revenue as our main unit of analysis due to data limitations. Instead, our analysis will focus on hotel occupancy rates. Also unlike the model used by the authors, our analysis cannot take advantage of a treatment and control region accounting for the different Airbnb market entry patterns. Instead, we propose the use of a Regression Discontinuity Model (RDM) (Jacob, Zhu, Somers, & Bloom, 2012).

Taking the arguments above into account, we use a region fixed effects to account for time-invariant differences in hotel occpuancies main specification for our proposal would look like this:

(1) $Overnights_{it} = \beta_i * \log Abb_{it} + \tau_i + \beta_j * X' + \varepsilon_{it}$

(2) $OccupRate_{it} = \beta_i * \log Abb_{it} + \tau_i + \beta_j * X' + \varepsilon_{it}$

Here, $OccupRate$ is the occupancy rate for all hotels in district i at time t. Log Abb is the log of the approximated total number of Airbnb listings available in district i at time t. Further, we control for economic conditions (unemployment rate and GDP per capita at federal level) and a district-specific linear time trend to account for unobserved heterogeneous variation across districts by using a month-year time dummy τ.

To decide whether to use a fixed or random effects model, we ran a Hausman test, since a regular OLS regression does not consider heterogeneity across groups or time (SOURCE: Green, 2008). It tests if the unique erros are or are not correlated with the regressors. Performing the test we get a p-value of 1 (check whether this is really true). Thus, we reject the null hypothesis that they are not. The Breusch-Pagan Lagrange multiplier test confirms our results, as there is a panel effect, i.e. a significant difference across neighbourhoods (p-value = 0.01).

Results & Discussion

Table 1 compares the results for the different models. Column 1 is a simple model where we only include region and month-year fixed effects to account for changes in time in variation across regions. We estimate the coefficient beta = 0.035, or equivalently, a 10% increase in Airbnb listings is associated with a statistically significant 0.35% ($p < 0.01$) decrease in monthly hotel room revenue.As stated earlier, we interpret a negative coefficient beta as indicative of some Airbnb stays substituting for hotel stays in cities with an established Airbnb presence. when looking at the effect of a percentage change of Airbnb Appartments on the change in hotel overnights in Berlin. The positive and highly significant effect of log Airbnb supply is a surprise. Hence, Column 2 includes economic control varibales to mitigate the above effect. Column 2 introduces observations for the unemployment rate or the log of average household income varying differently across neighbourhoods. We then want to control for unobservables that evolve following some trend that can differ from one neighbourhood to another. This is done by including a neighbourhood-specific time trend. The results are presented in Column 3.

By using the number of overnights per month and district as dependent variable we imply a constant number of beds over the course of the explored time period. This is certainly not the case, as capacity in Berlin hotels grew by xx% from 2010 to 2014. To account for the increased capacity we compute the theoretical hotel occupancy rate. From the Regional Database of Germany, we collected data on the supply of hotel beds in each district in a given year. Unfortunately this data is only recorded annually. Hence, our analysis must assume that the number of beds in each hotel stay constant throughout the year. For our main dependent variable, we used data on the number of beds generally available in each district per day in a given year and the monthly data on the number of overnight stays from guests to compute a variable as a proxy for hotel occupancy (as a percentage).

Whilst has the expected negative sign in all our specifications, it is only significant when we abstract from any additional control.

Table 1: Effect of Airbnb Supply on Hotels in Berlin (2010 - 2014)

	Dependent variable:			
	Overnight			Occupancy Rate
	(1)	(2)	(3)	(4)
log Airbnb listings	0.02^{***}	0.02^{***}	0.02^{**}	0.01^{***}
	(0.01)	(0.01)	(0.01)	(0.003)
Airbnb listings			0.001^{**}	0.0002^{**}
			(0.0002)	(0.0001)
Log*Level			-0.0001^{**}	-0.0000^{*}
			(0.0000)	(0.0000)
Average HH income (log)		0.59^{***}	-0.34	-0.49^{***}
		(0.20)	(0.25)	(0.10)
Unemployment rate		-1.72^{***}	-1.78^{***}	0.13
		(0.40)	(0.39)	(0.15)
Constant	13.01^{***}	8.90^{***}	15.79^{***}	4.03^{***}
	(0.03)	(1.54)	(1.91)	(0.73)
Neighbourhood-specific trend	Yes	Yes	Yes	Yes
Time trend	Yes	Yes	Yes	Yes
District FE?	No	No	No	No
Observations	720	720	720	720
R^2	0.99	0.99	0.99	0.91
Adjusted R^2	0.99	0.99	0.99	0.81
Note:				$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Our new main specification takes into account that the effect of Airbnb is likely best observed once it has reached a critical mass. However, even when introducing an interaction between log Airbnb supply and the absolute level of Airbnb listings we find that Airbnb supply has a highly significant positive effect on hotel occupancy, with a p-value on the .01 % level.

Further methods of data exploration

> To do that, we would define supply as the number of airbnb's available in a given apartment in a given month. This would be calculated by looking at the date that a host signed up for Airbnb and the date of the individual reviews. Upon the sign up date, the listing would be assigned a "life period" of 6 months, which would then be renewed upon each subsequent review. If a listing exceeds its life period without a review, then we assume that it is no longer available. However, once it recieves another review it would then be considered available again for another X months.

> We then want to control for unobservables that evolve following some trend that can differ from a region to another. This is often done in DD papers by including a linear or a quadratic region-specific time trend. In Zervas et al. (2014), they only report estimates with the quadratic city-specific trend since their results were apparently not really affected by this choice.

how listing's characteristics matter?

Airbnb lists all sort of accommodations, from a house or an apartment to shared rooms. This has an influence on how closely they can replace a hotel room, and thus have a different impact on hotel occupancy. We thus excluded all listings categorized as "shared room".

In contrast to Zervas et al. (2014) we did not find that Airbnb had a significant negative impact on hotel revenue per available room when looking at Berlin. Several reasons migh

First of all, they had used individual data for all hotels while we only had a computed value on hotel occupancy at the district level. Thus, we measure the impact of Airbnb supply on a much smaller scale and might have some sort of omitted variable bias.

Policy Implications & Conclusion

The purpose of this paper was to measure the impact of Airbnb on the hotel industry Berlin.

Bibliography

Airbnb. (2015). Airbnb Economic Impact. Retrieved from http://blog.airbnb.com/economic-impact-airbnb/

Cox, M. (2016). Inside Airbnb. Retrieved from http://insideairbnb.com/get-the-data.html

France-Presse, A. (2016, May). Berlin's government legislates against Airbnb. *The Guardian*. electronically.

Germany, R. D. (2015). Retrieved from https://www.regionalstatistik.de

Gutt, D., & Herrmann, P. (2015). *Sharing Means Caring? Hosts' Price Reaction to Rating Visibility*. Working Paper.

Guttentag, D. (2015). Airbnb: disruptive innovation and the rise of an informal tourism accommodation sector. *Current Issues in Tourism*, *18*(12), 1192–1217.

Jacob, R. T., Zhu, P., Somers, M.-A., & Bloom, H. S. (2012). *A practical guide to regression discontinuity*. Citeseer.

Schäfer, P., Braun, N., Reed, R., & Johnston, N. (2016). Misuse through short-term rentals on the Berlin housing market. *International Journal of Housing Markets and Analysis*, *9*(2).

Skowronnek, A., Vogel, L., & Parnow, J. (2015, February). Airbnb vs Berlin. Retrieved from http://airbnbvsberlin.com/#introduction

StatIS-BBB. (2016). Retrieved from https://www.statistik-berlin-brandenburg.de/webapi/jsf/dataCatalogueExplorer.xhtml

Stors, N., & Kagermeier, A. (2015). Motives for using Airbnb in metropolitan tourism – why do people sleep in the bed of a stranger? *Regions Magazine*, *299*(1), 17–9.

TheNextWeb. (2011). Airbnb takes on Europe. Will it revolutionize the industry, again? Retrieved from http://thenextweb.com/insider/2011/07/06/airbnb-takes-on-europe-will-it-revolutionize-the-industry-again/

Yannopoulou, N., Moufahim, M., & Bian, X. (2013). User-generated brands and social media: Couchsurfing and AirBnb. *Contemporary Management Research*, *9*(1), 85.

Zervas, G., Proserpio, D., & Byers, J. (2013). The rise of the sharing economy: Estimating the impact of Airbnb on the hotel industry. *Boston U. School of Management Research Paper*, *16*, 2014.

Zervas, G., Proserpio, D., & Byers, J. (2015). A first look at online reputation on Airbnb, where every stay is above average. *Where Every Stay Is Above Average (January 28, 2015)*.

Zervas, G., Proserpio, D., & Byers, J. (2016). The rise of the sharing economy: Estimating the impact of Airbnb on the hotel industry. *Boston U. School of Management Research Paper*, (2013-16).