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

Paulo Kalkhake & Dan Murphy

27 October 2016

Motivation & Research Questions

The effects of Airbnb on the Berlin housing market has recently been the subject of a great deal of scrutiny. 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 platforms recieved some of the blame. The city's law was widely welcomed by Berliners hoping to slow the forcs of gentrification, but it was also welcomed by another Airbnb victim in Berlin, the hotel industry.

While Berliners fought tourists for apartments, hotels were quietly fighting Airbnb for tourists. That's a tall order for the long-time incumbents, whose average price of 80 euros per night cannot compete with the average Airbnb price of 55 euros per night. To make matters worse, the rise of Berlin as the place to be happened to coincide with the rise of Airbnb as the place to stay. This prevented the hotels of Berlin from fully realizing profits that were, in their minds, rightfully theirs. In this paper we will seek to illustrate the magnitude of this "Airbnb effect" on hotels in Berlin. To that end, three hypotheses will guide our thinking. BERLINVSAIRBNBPLACEHOLDER

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: The effect of H1 will be more pronounced in districts with lower hotel density, since 77 percent of Berlin's Airbnb listings are outside of the main hotel districts. AIRBNBPLACEHOLDER

H3: Given the relative elasticity of travelers in choosing which district they stay in, hotel occupancy rates may decrease in a given district before Airbnb has even penetrated that district, if Airbnb has sufficiently penetrated a neighboring district.

Literature Review

Our work is closely related to that of Zervas, Byers, & Prosperio, whose 2016 paper explored the impact of Airbnb on the hotel industry in Texas. In their paper, Zervas et. al (2016) used a difference in differences method to measure the impact that Airbnb's presence had on Texas's hotel industry, and how this impact differed by region and by 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, howeve. Austin, for example showed a 8 to 10 percent impact on hotel revenue zervas2016rise.

Zervas et al. (2016) go on to investigate the impact of Airbnb's presence on different types of hotels and how the presence of Airbnb may effect hotels' pricing models. They find 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. Fascinating though this is, data and time limitations will prevent us from investigating similar trends in Berlin. This we leave to future researchers whom the data gods have smiled upon more generously zervas2016rise.

Data

Our investigation utilizes data from two different sources:

First, we will use monthly tourism survey data from the Statistical Information System Berlin Brandenburg (PLACEHOLDER CITATION: StatIS-BBB) and the Regional Database Germany (PLACEHOLDER). These data sets provide reliable information about the current situation and especially the short-term development of tourism in Germany. 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 have specific data for each of the ten districts in Berlin. We will use this geographic specificity to take a close look at the relationship between hotel occupancy rate and Airbnb supply in each of Berlin's districts.

Second, our paper uses data from *InsideAirbnb.com* Airbnb:2016aa. This data was extracted from the Airbnb site between 18 July 2015 and 6 January 2016. It includes listings from more than 15 cities in 16 countries, one of them being Berlin. While using data directly extracted from Airbnb's website or API for our research would be preferable, it is not disclosed. However, the data extracted from *InsideAirbnb.com* is an appropriate alternative. We have detailed information on all Airbnb listings for Berlin between 2010 and 2014, including calendar data, review data, and listing ID numbers. This allows us to conduct comparable time based analytics in concert with our hotel data.

Our most significant data challenge is the absence of precise listing availability data during our period of interest. However, in keeping with the Zervas et al. methodolgy, we will use review data as a proxy with which we can establish whether or not a listing was available in in a given month. Zervas et al. acknowledge that while this is not ideal, Airbnb itself is unable to produce exact availability data due to the prevalence of "stale vacancies" on the site. That is, listings whose owners have not accurately updated their availability. zervas2016rise

Methodology

Having described our data sources and the basic properties of our data, we propose to approach our research questions using a regression discontinuity model. Contrary to the Texas Model, our analysis cannot take advantage of a treatment and control region accounting for the different Airbnb market entry patterns. Mainly, because

Instead, we propose the use of a Regression Discontinuity Model (RDM). By introducing a binary variable which is one for all observation after Airbnb's market entry and completely interacted 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.

Taking the arguments above into account, our main specification for our proposal would look like this:

$$\log OccupRate_{it} = /beta_i * \log Abb_i t + \tau_t + X + \varepsilon_i t$$

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

Bibliography