Proposal for the Final Research Project in Introduction to Collaborative Social Science Data Analysis (MPP-E1180)

Paulo Kalkhake & Dan Murphy

27 October 2016

The effect of Airbnb's German market entry on Hotels in Berlin: A Regression Discontinuity Approach

Motivation & Research Questions

In recent years, the Berlin housing market has been the subject of a great deal of scrutiny. "Real Berliners" raised the alarm as a hoard of gentrifying students and entrepreneurs descended upon the city in recent years. However, they are not the only Berliners threatened by a new competitor. The rise of apartments on Airbnb as the place to stay has coincided with the rise of Berlin as the place to be and, like Real Berliners, hotels have noticed the new competition.

While Berliners fight tourists for apartments, hotels have been fighting apartments for tourists. This is no easy fight 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. BERLINVSAIRBNBPLACEHOLDER

We hypothesize that the higher the Airbnb supply in a given district in Berlin, the lower the hotel occupancy rate will be in that same district. Further, we believe that this effect 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. We also believe that, 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 AIRBNBPLACEHOLDER

justifying why the topic of study is interesting

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 (**???**).

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 (**???**).

providing a basic literature review (properly cited using BibTeX)

Data & Methodology

Our investigation is based on data from two different sources:

First, we use monthly tourism survey data from the Statistical Information System Berlin Brandenburg (PLACEHOLDER CITATION: StatIS-BBB) and the Regional Database Germany (PLACEHOLDER). The data provides reliable information about the current situation and especially the short-term development of tourism in Germany. The results are organized regionally according to districts, municipalities and also by region. This allows us to have specific data for each of the ten districts in Berlin and thus have a closer look the the geographically specific relationship between hotel occupancy rate and Airbnb supply. The surveys are carried out at the beginning of each month and refer to the reporting period of the previous month.

Third, our paper uses data downloaded from *InsideAirbnb.com* Cox (2016). The data behind *Inside Airbnb* is extracted from publicly available information from the Airbnb site between 18 July 2015 and 6 January 2016. It includes listings from more than 15 cities in 16 countries, among them Berlin. Using data directly extracted from Airbnb's website or API for our research would of course be preferable, but that data is not disclosed. The extracted data is an appropriate alternative. We have detailed information on all Listings for Berlin in the mentioned time period, such as Calendar Data, Review data, and Listing ID, which allows us to conduct time based analytics.

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

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 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. LogAbb 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

Cox, Murray. 2016. “Inside Airbnb.” http://insideairbnb.com/get-the-data.html.