Assignment 3

Daniel J Murphy & Paulo Kalkhake

November 9, 2016

Introduction

While the effects of Airbnb on Berlin apartment prices has garnered a great deal of focus, significantly less attention has been paid to Airbnb's effect on the hotel industry. Airbnb claims that they are largely complementary to hotels, since most Airbnb listings are found outside of the main hotel district in a given city. In Berlin for example, 77% of Airbnb listings are outside of the major hotel districts.

However, a paper by @zervas2016rise 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 complentary, 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, 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, the Regional Database of Germany, and *InsideAirbnb.com*.

From the Statistical Information System: Berlin/Brandenburg, we collected monthly data on the number of nights spent in hotels and the number of guests that arrived in hotels in Berlin [@Brandenburg:2016aa].

From the Regional Database of Germany, we collected data on the supply of hotel beds in each neighborhood in a given year. Unfortunately this data is only recorded annually, so our model will assume that the supply stays constant throughout the year [@Germany:2015aa].

From *InsideAirbnb.com*, we scraped data on listings and reviews on all Airbnb listings for Berlin between 2010 and 2014. The most important variables for our model are the following [@Airbnb:2016aa].

1. The neighbourhood of each listing
2. The date that an Airbnb host signed up
3. The date of the last review of each listing
4. The date of each inividual review of each listing

Data Cleaning & Merging

In the original detailed listings data set from *InsideAirbnb.com*, there were 92 variables covering topics ranging from room price to information on the host. Most of these variables were superfluous, however, leading us to drop all but six variables. Those were listing ID, host ID, listing Neighbourhood, room type, number of beds, date of host registration, date of the first review, and the date of the last review.

Methodology

We will use a Regression Discontinuity Model to evaluate whether the emergence of Airbnb in Berlin has affected hotel occupancy rates. By introducing a binary variable that 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 will 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 [@jacob2012practical].

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

(1) $\log OccupRate_{it} = \beta_i * \log Abb_{it} + \tau_i + \beta_j * X_j + \varepsilon_{it}$

Here, log *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 specific to each neighbourhood (unemployment rate and income) and a district-specific linear time trend to account for unobserved heterogeneous variation across districts by using a month-year time dummy τ.

Analysis

(both descriptive and inferential statistics)