Santa Clara University

Seasonal Airbnb Price Differences in Differences Analysis

Peter Sushko and Wesley Denison

Economics 173: Applied Econometrics

Professor Helmers

March 2019

Causal Research Question:

Here in sunny California many of us are lucky to rarely if ever experience a "real winter" with snow and frigid temperatures, however for much of the country the effects of harsh winter weather are an annual reality. Curious about quantifying the economic effects of harsh winters, we landed on the cause research question of "How do harsh, snowy winters, affect rental prices in US cities?" Our initial hypothesis is that we suspected the presence of a harsh winter would have a significant negative effect on rental prices; representing in general people's lack of desire to stay and/or visit cold, snow-covered cities.

Empirical Identification Strategy:

Because there is no way to randomly assign harsh winters as a treatment to cities and observe their effect on rental prices, and there is no discontinuity as cities that have harsh winters have always had harsh winters, we settled upon a Differences in Differences method to answer our research question. We will be comparing the difference in summer and winter rental prices in cities with harsh winters to the difference in prices in cities without.

The main assumption needed to use the Differences in Differences method is that treatment and control outcomes move in parallel in the absence of treatment. For us this would be the assumption that the rental prices of the two cities being compared (ex: Seattle and LA) would move in parallel if Seattle did not experience harsh winter weather. This is a very significant assumption. Because of the many variables individually influencing cities' rental prices there is undoubtedly some difference in price trends even after removing the harsh winter treatment . To combat this we attempted to compare similar cities (this is outlined later in the paper) and also compared 3 different pairs to ensure our results were consistent. Even so, there is undoubtedly Omitted Variable Bias we were unable to account for due to the constraints of our data and tools. (such as…)

Data:

Inside Airbnb is an independent project with the mission of making publicly available tools and data on Airbnb listings in numerous cities around the world. Monthly scraped datum containing all listings with detailed parameters regarding reviews, host, property, and price information are made readily available on their website insideairbnb.com. Among the variety of datasets published for each city is calendar.csv which contains every posted Airbnb listing along with the listed price for every day in a given city spanning a full year ahead of the date the data was scraped.

Using this dataset, we planned to plot and analyze the listed price of available rentals over the course of a year, then using a differences in differences analysis between northern and southern cities. To ensure that the hosts have adjusted their prices according to differences in

demand between season and we were analyzing dates set comparably in the future from the scraping date, we used data scraped roughly a month before the dates we were analyzing.

The sheer size of the calendar data sets, while beneficial in many regards, became a significant challenge as with millions of rows, the file was unable to be fully opened by excel, and attempting to run regressions in R took over 10 minutes and quickly exceeded R's maximum allocated memory. As a result, we were forced to manually truncate the data by removing unavailable listings and limiting our analysis to two months: January and July (representing Winter and Summer respectively).

R Analysis:

We broke the cities up into three pairs: Los Angeles & Seattle, Austin & Minneapolis, and Chicago & New Orleans. The reason we paired them this way is to account for factors that depend on the cities' longitude. Such factors include the cultural difference that is suspected to play a significant role in price development. We also paired cities with similar population size, once again to ensure the cities were as similar as possible to facilitate comparison and limit omitted variable bias.

The datasets we used contained a lot of data about prices, so we decided to narrow it down to single data points by using Median and Average values as parameters to represent the prices during summer and winter seasons. In all cases, the average price was above the median. This is due to luxury properties significantly affecting the average because of their extremely high costs (some properties were valued over $5,000 a night) and/or test and placeholder listings with arbitrarily high prices. However, these extremely expensive homes don't appear in the data set very often, so they have a minuscule effect on the median.

Our data set was split into two groups, one containing rental prices in July (summer prices) and the other containing January prices (winter prices). In order to use the differences in differences method, its necessary to compare the seasonal price change in both group of cities: those who receive the treatment and those who don't. The first step was to create a boolean dummy variable for the treatment itself (1 = winter = January data, 0 = summer = July data) named Season. Then, a second dummy variable named Treated was created to identify the cities exposed to the treatment. In our case these cities are Chicago, Seattle, and Minneapolis as being northern cities, they are exposed to our harsh winter treatment while Chicago, LA, and Austin are not.

We used the differences in differences method by comparing both Medians and Means in each of three pairs. The regression table accounted for the difference in seasons and the difference in treatment, which in this case is the cold climate in some cities. A new dummy DiD was created to represent the differences in differences. DiD is the product of Treated and Season because cold winters happen only in cities that are in the northern part of the United States, and as the name suggests, only during winter.

Lastly, we used a linear regression model to estimate the effect of cold weather on rental prices and found a consistent result. We compared to pairs of cities between each other, and also look at the aggregate result.

Results:

Here is the regression table for pairs:

Regression

	Dependent variable:					
	Avg_Price	Median_Price	Avg_Price2	Median_Price2	Avg_Price3	Median_Price3
	Chicago vs New Orleans	Chicago vs New Orleans	LA vs Seattle	LA vs Seattle	Minn. vs Austin	Minn. vs Austin
	(1)	(2)	(3)	(4)	(5)	(6)
treated	-37.29	-30.00	3.67	94.00	436.11	115.00
season	-7.38	-12.00	-22.10	26.00	-13.27	-13.00
DiD	-22.63	-10.00	-62.70	-106.00	-90.15	-63.00
Constant	198.57	142.00	209.38	85.00	100.84	80.00
Observations	4	4	4	4	4	4
R2	1.00	1.00	1.00	1.00	1.00	1.00
Note:						*p<0.1; **p<0.05; ***p<0.01

Columns 1 and 2 compare the differences in Mean and Median between the cities of Chicago and New Orleans. Columns 3 and 4 compare the Mean and Median between Los Angeles and Seattle. Columns 5 and 6 look at the differences between Minneapolis and Austin. Almost all data showed negative seasonal coefficient, which shows that a decrease in Airbnb rentals during winter. The only outlier, in this case, was Los Angeles, because in that median price regression the season coefficient is positive. This is due to the fact that median rental prices are higher in winter than summer in LA. This outcome is slightly unexpected. Our explanation is that it just a random error because the average rental price is higher in summer in LA. It could also be the case that because summer is a tourist season, a lot of smaller properties start renting their houses on Airbnb driving the median down.

Additionally, the Treated variable shows the differences in prices associated with the cities' geography. The variable takes a negative value in the first comparison which means that in the Chicago and New Orleans pair, the warmer city has higher average and median rental prices. In other cases, northern cities were more expensive, which is somewhat expected in the LA and Seattle pair and less in the Minneapolis and Austin comparison. A big difference arises when the mean values of Minneapolis and Austin are compared: average Airbnb rental in Minneapolis is $436 more expensive than an average rental in Austin. This is due to the fact that the dataset contains a lot of very expensive rentals out in small Minneapolis which hugely spikes up the price. The issue gets better when the median is used instead of the mean, and in that case, we see only a 115 dollar difference between the cities.

DiD, the most important variable in the regression, is negative throughout the comparisons. The values it takes for the Chicago and New Orleans pair are $-22 and $-10. These are reasonably close estimators of the effect of cold winters on Airbnb's prices in the city of Chicago. For the LA and Seattle pair, the estimation goes down and is measured at $-106 and $-62. From this information, we can conclude that Seattle's demand for short term rental

homes decreases significantly during winters. Lastly, the pair of Minneapolis and Austin produces a DiD value of $-90 and $-63 for both parameters. This is more consistent with the LA to Seattle pair.

To better quantify the effect of cold winters on Airbnb rental prices, we decided to add the southern cities together and compare them to the combined prices of the northern cities. The regression produced the following table:

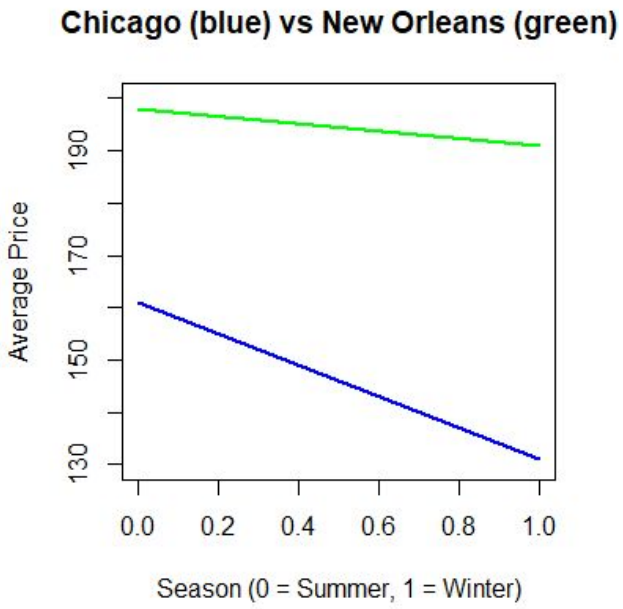
Combined Pairs

	Dependent variable:	
	Avg_Price_Total Total Average (1)	Median_Price_Total Total Median (2)
treated	134.16	59.67
season	-14.25	0.33
DiD	-58.49	-59.67
Constant	169.59	102.33
Observations	4	4
R2	1.00	1.00
Note:	*p<0.1; **p<0.05; ***p<0.01	

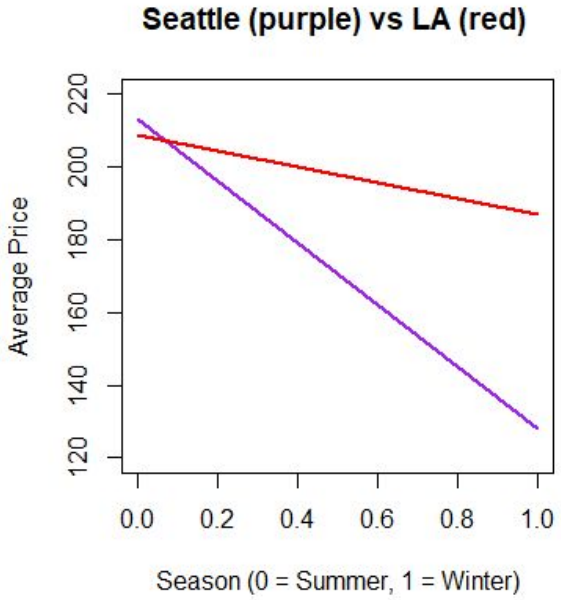
Here, we still get that northern cities have more expensive average and median prices. This is most likely do to the fact that out of the cities we used, northern cities have more developed economies and hence, higher real estate prices. Additionally, the season variable is negative when the averages are used, and zero when the medians are compared. This says that in our dataset, all cities decrease their Airbnb rental prices by a small amount in the winter. The Differences in Differences component, however, takes very similar values of $-60 and $-59. This means that on average, the effect of cold winter costs American cities a decrease of around $60 per night during winter.

We conclude that cold winters decrease demand and, therefore, Airbnb rental prices in northern American cities by around $60 during the winter season. This value is close to what we expected it to be at the beginning of the project, and it also makes a lot of sense, as we expect people to travel to warmer places during winters. We also created graphs which show differences in average prices between the three cities and allows for the visualization of the Difference in Differences comparison.

The first pair has higher prices for New Orleans and lower prices for Chicago. The latter also takes a bigger hit during winter, which we wanted to quantify.



The relationship between LA and Seattle is very similar to Chicago and New Orleans, with the only exception being that in the summer Seattle is slightly more expensive than Los Angeles.



Lastly, Minneapolis and Austin follow the same line trends. This graph also highlights the big differences in prices, but this difference is accounted for in the DiD method.