MODELING HOUSING PRICES WITH COMPOUND POISSON PROCESS

HAMZAH SAMI
STAT 482
ID: 025305655

Introduction

According to Norada Real Estate, a website that analyzes real estate and property management in various markets, the LA real estate market is considered to be "one of the premier markets for both investors and homeowners. It is also touted as the nation's least affordable housing market." While cities such as Beverly Hills and Palos Verdes are certainly communities where property values are high, smaller cities such as Hawthorne are undergoing a transformation in property value due to an influx of tech companies establishing offices. Established corporations such as Northrup Grumman and SpaceX have made Hawthorne the place to be and have brought along tech companies SnapChat and Ring have built offices in Hawthorne. As a result, this has led to a migration of a younger demographic of people into what has traditionally been an older community.

According to researchers from the Financial website GOBankingRates, Hawthorne was considered the best place to buy real estate in California because of the abundance of engineering companies such as Northrup Grumman and SpaceX as well as easy access to the beach and Downtown LA,

The purpose of my project is to predict the total value and the total number of houses sold in Hawthorne in a calendar year. In doing so, the goal is to learn more about how many houses can be sold in a year and to predict a trajectory for housing sales given a two-month primer on properties sold in Hawthorne.

Data Set Description

The data used for this project was obtained from Redfin, a website and app that helps house buyers find homes, organize tours, and connects home buyers with real estate agents to help them through the process. For my project, I looked specifically at houses that were sold within a two-month window (September 2020 and October 2020) in Hawthorne, California. According to Redfin, there were 46 houses that were sold from September to the end of October with prices ranging from $500,000 to $1,000,000.

Methodology

In order to determine the total number and value of the houses sold in a year, I wanted to use the Compound Poisson method. I thought this method was best as it would allow me to separately model the houses sold as well as their selling price. From further investigation of the data, it was apparent that the houses sold modeled a Poisson distribution. To determine a value for lambda, I took the total number of houses sold in the timeframe being studied (46) and divided by the number of days (61).

To model the selling price, I first checked to see whether it was normally distributed or not. From creating a histogram of the prices and from running the Shapiro-Wilk test, I was able to conclude that the housing prices were normally distributed with mean $776906.50 and variance 28896811100. From here, I proceeded to program the code to simulate the trajectory in R.

Results

From running the Compound Poisson process, the total profit over 365 days of houses sold is $207,041,368 and 284 houses will be sold over this time period. The numbers were quite large and I thought of transforming the numbers by a factor or a log transform. At the end of the day, I decided to keep the numbers as is because a transformation at any step would affect the ability to make reasonable predictions.

Based on the answers calculated, these predictions appear to be quite reasonable. Over the course of a year, 284 homes sold amounts to about 24 homes sold every month. Seeing as how industrialized Hawthorne is becoming, this is a fair statement to make. However, there are obvious factors that could affect these results in real-time. For instance, the quarantine may end up slowing down growth in the city and slow down the market as more people are staying at home. In addition, factors such as inflation could cause the price of homes to swell up even more than the model predicted.

The plotted trajectory indicates very strong growth in value over the next year for houses. While external factors will certainly skew our answers, the plotted trajectory falls in line with several news headlines that have been calling Hawthorne the best place to buy property in California.

Conclusion

The goal of my project was to predict the potential number of houses that would be sold in a calendar year as well as the total market value of properties that were sold. In order to have a lambda value, I found data on Redfin about 46 properties sold in Hawthorne from September 1st to October 31st. In doing so, I determined lambda to be 46/61 or 0.754. From here, I programmed the trajectory and determined that approximately 284 homes were sold within the year with a total market value of $207,041,368. This was an interesting result, given that covid-19 and other external factors will affect the buying and selling of properties. Furthermore, the variance in the data was so extreme that I was concerned the lambda value would be heavily skewed or useless in helping to simulate a reasonable trajectory. This project did highlight to me how powerful these results can be; with just a small amount of data, I could simulate a potential buying pattern in the real-estate market.

With that said, the most impressive result obtained was regarding the plotted trajectory. The model projects a linear growth in total home value over the next year so even though the city may be struggling financially due to the quarantine, the housing market is showing strong signs of life. These calculations fall in line with news headlines as tech companies such as Ring and SnapChat are relocating to Hawthorne with their workers in tow. While it's good to have strong technological roots, it is somewhat concerning that these properties are reaching astronomical value which will make it difficult for local residents to live in the same neighborhoods. This could mean a loss of culture and identity; a similar crisis enveloped San Francisco and it is frightening to think Hawthorne is headed in the same direction. I hope that the changes that do result are meaningful and enable Hawthorne to remain the 'city of good neighbors.'

References

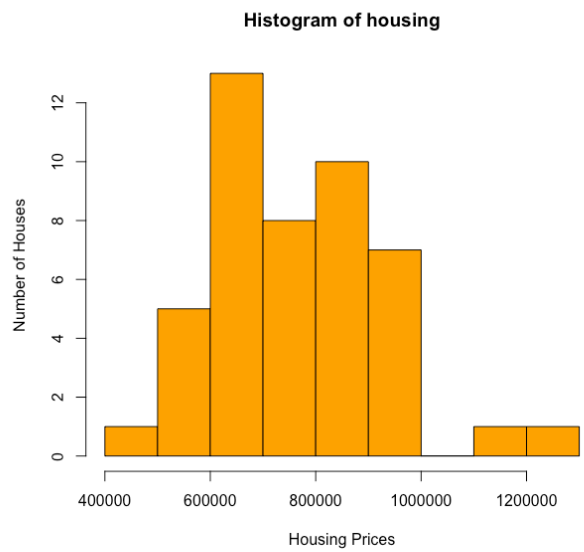
https://www.noradarealestate.com/blog/los-angeles-real
estatemarket/#Los_Angeles_Real_Estate_Investment

https://www.redfin.com/city/8306/CA/Hawthorne/housing-market

https://fisherrealestate.com/the-top-place-to-buy-a-home-in-california-is-hawthorne/

Appendix I

```
housing <- data <- read.csv(file = "~/Documents/csulb statistics/stat 482/housing.csv",
header = TRUE, sep = ",")

#histogram of housing
hist(housing,col= 'orange',xlab = "Housing Prices",
ylab = "Number of Houses")
```

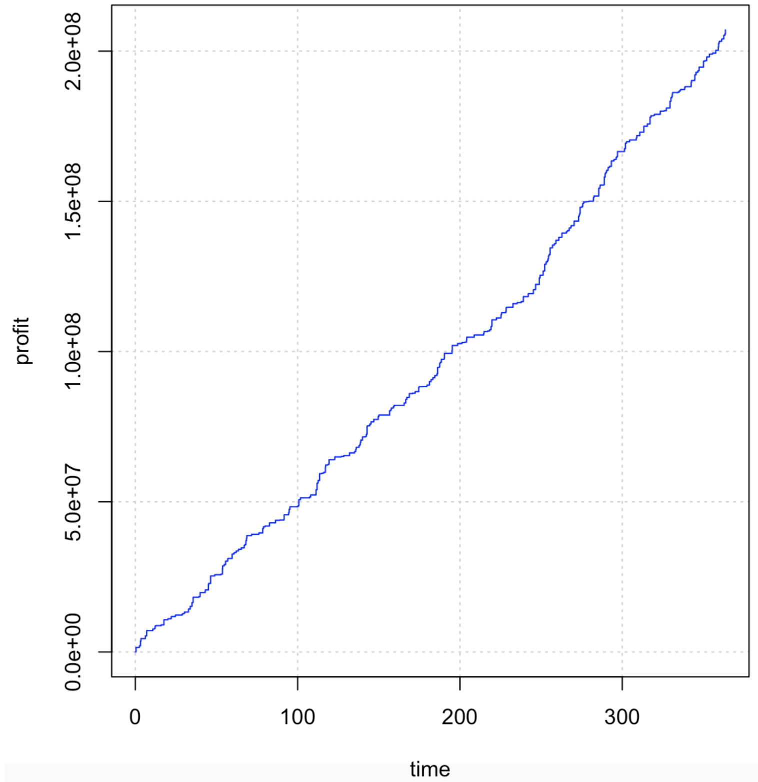



```
#verifying distribution of housing prices
shapiro.test(housing)
```

```
	Shapiro-Wilk normality test

data:  housing
W = 0.96353, p-value = 0.1571
```

```
house_avg = mean(housing)
house_var = var(housing)

#method 2. Order Statistics.

t<- 365
lambda<- 46/61

#generating N(t)
set.seed(12345)
Nt<- rpois(1,lambda*t)


#generating N(t) standard uniforms
```

```
unif<- 1:Nt
for(i in 1:Nt)
  u[i]<- runif(1)

#sorting standard uniforms
sort_unif<- sort(unif)

#computing N(t) event times
events<- t* sort_unif


#simulating trajectory for one year (365
days)
time<- 1:2*Nt
profit<- 1:2*Nt
time[1]<- 0
profit[1]<- 0
for (i in seq(2, 2*Nt, 2)) {
  time[i]<- events[i/2]-0.001
  time[i+1]<- events[i/2]
  profit[i]<- profit[i-1]
  profit[i+1]<- profit[i-1]+(abs(rnorm(1)))
}

#plotting simulated trajectory
plot(time, profit, type='l', col=4, panel.first=grid())

Nt

profit[length(profit)]
```




```
> Nt
[1] 284
```

```
> profit[length(profit)]
[1] 207041368
```

Appendix II

$$X(t) = \sum_{i=1}^{N(t)} Y_i$$

Expected Value

$$E[X(t)] = \lambda t\, E(Y_1)$$

$$\begin{aligned} E[X(365)] &= 365*(0.754)*(776906.50) \\ &= 213812437.5 \text{ dollars} \end{aligned}$$

Variance

$$Var(X(t)) = \lambda t Var(Y_1) + \lambda t(E(Y_1))^2$$

$$\begin{aligned} Var(X(365)) &= (0.754)(28896811100) + (0.754)(365)(776906.50)^2 \\ &= 24660287002.3 \text{ dollars-squared} \end{aligned}$$