Evaluating How Crime Rate Impacts Rental Housing Price

Brian Tung, Kevin Cahillane, Meng-Kang Kao, Nic Brathwaite

7/31/2022

Introduction

Home ownership has always been a core component in the American Dream as a milestone of prosperity, with many Americans viewing it with priority over aspects such as career, family, and college. In the effort to build wealth, some Americans seek investment opportunities in building a portfolio of rental properties. In a wealth-concentrated location like the Bay Area, second-home sales jumped 54% in 2021 from pre-pandemic levels[1].

However, today's competitive housing market and astronomical prices force buyers to make compromises in purchasing decisions, such as neighborhood safety. Experts have found that properties in high crime areas may be subject to frequent property damage, low-quality renters, and lower property values[2]. For home buyers with a tight budget, uncertainty in navigating real estate opportunities may likely lead to a poor investment decision.

The core objective of this study is to leverage rental listing data in the Bay Area along with city safety information in hopes to provide guidance for home buyers to make the best purchasing decision and maximize ROI. We use regression models to estimate the relationship between the rental price of a property and city crime rates with additional impacts of environmental and property features.

Data and Methodology

The data used in our study consists of datasets from three separate sources. The first dataset is part of "San Francisco Rentals" dataset on TidyTuesday, which was web-scraped from Bay Area Craigslist rental listings using the internet archive repository, Wayback Machine. This dataset contains detailed property information on city, county, rental price, number of bedrooms, number of bathrooms, and square footage. The second dataset is a violent crimes report from the FBI's Uniform Crime Reporting agency[3], which includes categorical crime rates and population data by city. The violent crimes dataset was collected via the Hierarchy Rule, which only counts the most serious offense in a multiple-offense criminal incident. The third dataset is the city area in square miles from Wikipedia[4] . We then derive two more fields of population density(in k), which is based on city population data from FBI crime dataset divided by square miles, and crime per 1000, which is the crime count divided by population times 1000. All of the datasets used in this study are observational.

[1] Hansen, Louis. "Record second-home purchases fueled by Bay Area wealth." Mercury News (2021) https://www.mercurynews.com/2021/11/28/record-second-home-purchases-fueled-by-bay-area-wealth/

[2] Reeder Asset Management, "Why Rental Property Owners Must Evaluate Crime Rates for ROI" (2021) https://www.reederproperties.com/blog/why-rental-property-owners-must-evaluate-crime-rates-for-roi

[3] FBI, "Crime in the United States 2012" (2012), https://ucr.fbi.gov/crime-in-the-u.s/2012/crime-in-the-u.s.-2012/violent-crime

[4] Wikipedia, "List of cities and towns in the San Francisco Bay Area", https://en.wikipedia.org/wiki/List_of_cities_and_towns_in_the_San_Francisco_Bay_Area

Figure 1 shows the plot of crime per 1000 over rental price. To avoid potential time series bias, we decide to limit our study to rental posts from 2012 and use 2011 rental posts as our test dataset for EDA. We also drop Emeryville from the dataset due to its extreme crime ratio outlier. Emeryville is commonly ranked as one of the most crime-ridden cities in the US[5].

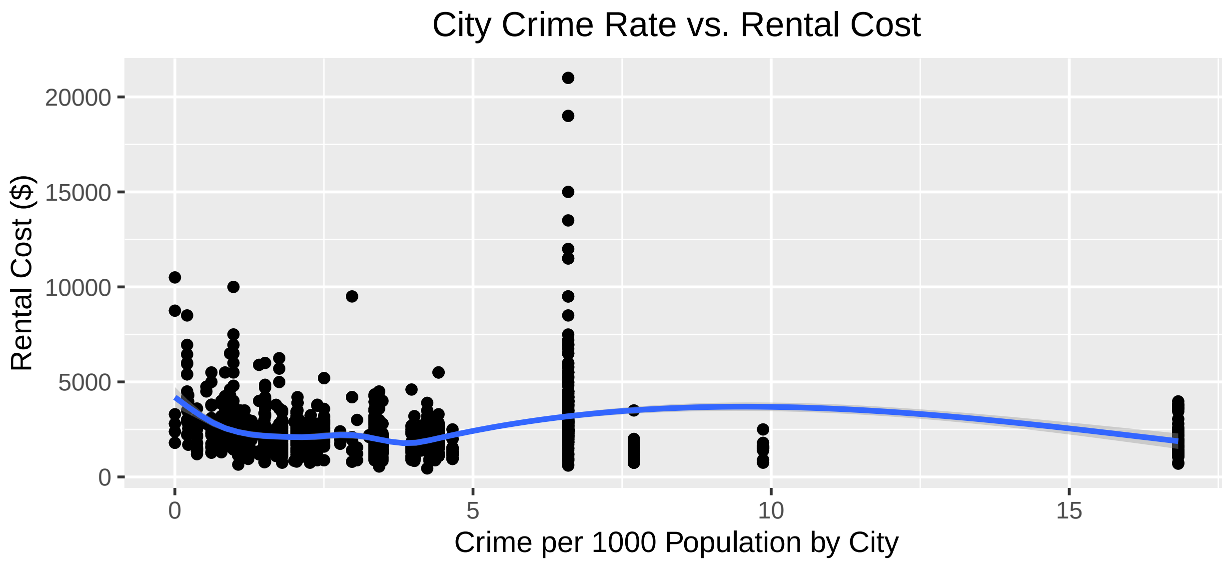

Figure 1: Emeryville has unproportionally high crime rate

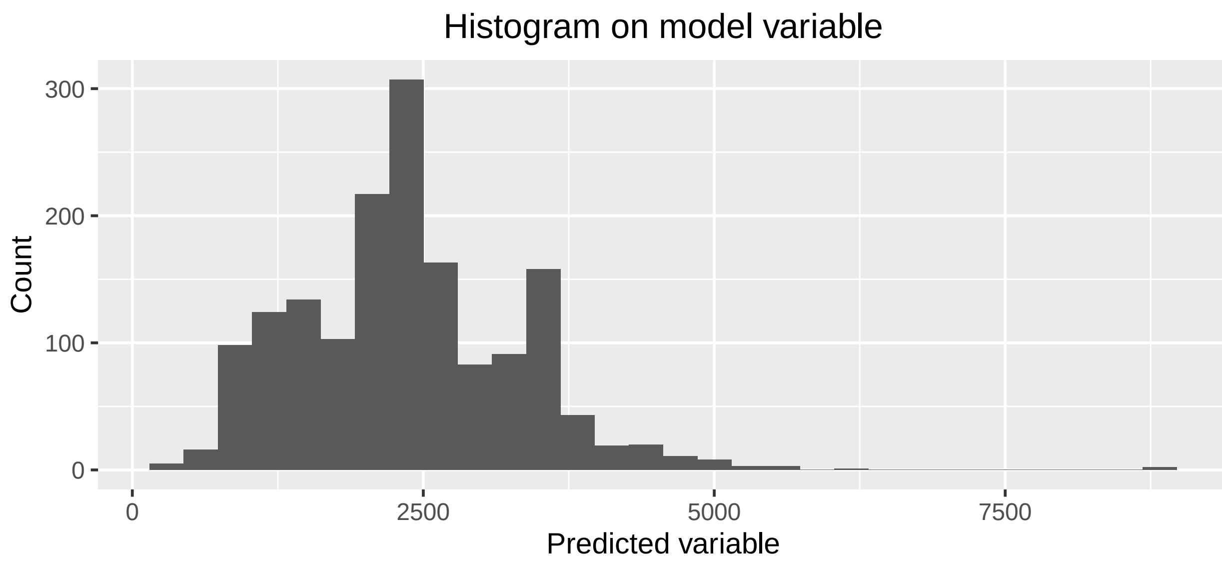

Figure 2: No severe skew to cause no unique BLP

Figure 2 shows the predicted value histogram to visually scan for heavy tails. The plot is approximately normal despite a slight right skew, which is not indicative of infinite variance. Additionally, we prove no perfect collinearity through R's functionality of dropping variables. Since no variables were dropped in R, there is no perfect collinearity.

[5]CrimeGrade.org, "The Safest and Most Dangerous Places in Emeryville, CA: Crime Maps and Statistics", https://crimegrade.org/safest-places-in-emeryville-ca/

We operationalize rent cost at the household-level, using the posted price of Craigslist advertisements. These were pulled using the Wayback Machine which archived Craigslist on random days. As a result, if there were any changes in the listed price over the course of the year, our rent variable shows a randomly selected price from the life of the posting. To estimate the effect that crime rate has on rental prices in the Bay Area, we develop a linear regression model with the crime-to-population ratio (crime_ratio) as our primary causal variable and the rent price (price) as our dependent variable. After running a regression on this model, we decide to control for the potentially confounding variable of urbanness. We believe that omission of this variable causes the crime ratio coefficient to have a highly positive bias. We approximate urbanness using the population density of each city (pop_density).We also choose to control for other aspects of the household (house_dimensions) that commonly affect rent prices such as the number of bedrooms, bathrooms, and total square footage. Finally, we remove square footage from the house dimensions variables since it is highly correlated with the number of bedrooms.

Below is the third and final model. The results from all models can be found in the Results section.

$$\widehat{Price} = \beta_0 + \beta_1 Number\ of\ Bedrooms + \beta_2 Number\ of\ Bathrooms + \beta_3 Population\ density + \beta_4 Crime\ rate$$

Results

Table 1 below shows the three most appropriate models. All models show evidence that the crime rate per 1000 people has a statistically significant effect on rent. Model 3 in Table 1 below shows that crime rate has a negative coefficient of -128.98. In other words, an increase of 1 crime per 1000 people per year would lower the monthly rent costs by approximately $130. Given the crime rate in the Bay Area is between 0 and 6 per 1000 people, this coefficient shows practical significance.

Model 3 removes square footage from the linear model and produces statistically significant coefficients for number of bedrooms (245.37), number of bathrooms (689.61), and the population density (137.28). Meaning, for every bedroom in an apartment the price of rent increases by approximately $245, every additional bathroom increases rent by approximately $690, and for every 1000 additional people per square mile rent increases by an estimate of $137.

By comparing Model 2 and Model 3 with Anova, we see Model 3 has a highly statistically significant P-value. Thus, we conclude that Model 3 is a better representative of the true rental housing price by adding population density variable to the linear model.

Limitations

We do not believe that our dataset is IID, since Craigslist data is not able to capture the entire rental market. As highlighted by the author of the dataset, there is reason to expect that the dataset fails to capture the highest end of the market, where housing transactions are primarily managed by real estate agencies[6].

Regression models on large samples also require the existence and uniqueness of a best linear predictor. During our EDA visualization phase, we did not find evidence of a heavy-tail distribution. There is also no evidence of perfect collinearity, as no variables were automatically dropped.

Omitted variables that could introduce structural biases include average income, and Bay Area demographics due to their potential correlation with both rent price and crime rate. We expect income to have a negative effect on crime rate. Thus, we expect the omitted variable bias to push the negative coefficient of crime towards zero.

[6] Pennington Kate, "Craigslist Scrape Methodology" (2018), https://www.katepennington.org/clmethod

Table 1: Estimated Regressions

	Output Variable: Rental Cost		
	(1)	(2)	(3)
Number of bedrooms	−89.67	36.40	245.37***
	(71.73)	(67.05)	(26.97)
Number of bathrooms	511.53***	486.17***	689.61***
	(98.04)	(89.51)	(57.22)
Square feet	0.70**	0.61**	
	(0.23)	(0.21)	
Population(K) per Sq.Mi		130.22***	137.28***
		(4.70)	(4.65)
Crime per 1000 population	88.70***	−125.23***	−128.98***
	(11.22)	(8.39)	(9.08)
Constant	417.99***	34.81	−81.39
	(84.21)	(83.16)	(94.08)
Observations	3,374	3,374	3,374
Adjusted R^2	0.36	0.50	0.45
Residual Std. Error	979.43 (df = 3369)	866.84 (df = 3368)	912.68 (df = 3369)
Note:			*p<0.05; **p<0.01; ***p<0.001

Conclusion

This study estimates the impact of city level crime rates on rental prices in the Bay Area and supports our beliefs that crime rates have a significant role in influencing rental listing price, reducing rental prices with each crime event in 1000 people. During our model exploration phase, we found that omitting population density from the model resulted in a positive crime coefficient, and was corrected upon implementation. We also explore additional predictors including number of bedrooms, number of bathrooms, and city population density, which were all found to have a positive coefficient and a statistically significant impact on rental price.

We encourage future research to explore which crime has the most impact on rent price. Additionally, we did not include average income in our model due to lack of reliable data sources, but future studies should also consider exploring this direction to provide a different perspective. Our goal for this area of research is to provide potential investment property owners a better understanding of the relationship between rental price and crime rate.