ISM 6137: SDM Final Project Report

# Housing Affordability crisis

Suryateja Chalapati

# Table of Contents

# 1. Executive Summary

While there is no universal definition, housing affordability broadly refers to the financial ability to purchase a home. Homeownership rates among young adults today are even lower than in 1988, and the share of cost-burdened renters is significantly higher. The housing affordability crisis is a complex, multidimensional issue driven largely by the gap between housing prices and household income and is influenced by the balance between housing supply and demand, the labor market, and mortgage rates by way of Federal monetary policy.

The analysis in this paper was performed on all 50 states in the United States, in addition to the District of Columbia over the period 2010 to 2019. Statistical methods such as random effects models with log transformation on population and median income were used for the analysis. We performed quality checks and used plots to evaluate the relationships between variables.

Our models indicate that the percentage of subsided houses occupied, minimum wage, and unemployment rate are all factors in housing affordability. To help alleviate the lack of affordable housing for residents, we recommend that policy makers allocate greater resources towards subsidizing housing, explore increasing the minimum wage and create skills-based training programs targeted towards the unemployed, particularly in densely populated states.

# 2. Problem Significance

A large and growing share of American households cannot find housing they can afford to buy. Homeownership rates among young adults today are the lowest they have been in the past 20 years. From 1990-2016, the national median home price rose 41% faster than overall inflation. Experts generally agree that 30% of household income spent on housing for mortgage or rent is the threshold beyond which housing costs are deemed unaffordable; these households are considered "cost-burdened." In 2019, more than 37 million households were cost-burdened, representing more than a third of all households nationwide.

The target clients for this analysis are regional policy makers in the United States. This is because cost-burdened renter households move more frequently than those that are not, either to keep their housing cost-to-income ratio within manageable levels or to lower it. Housing affordability also has wider implications for economic security, with the cost of housing determining what is left over in household budgets for food, education, healthcare and other costs that affect lifestyle and quality of life. Persistent unaffordable housing has been linked to increased odds of poor self-rated health, as well as increased odds of experiencing a number of related health outcomes, such as hypertension. In order to maintain a productive workforce and continue to attract young workers, it is pivotal for policy makers to formulate policies geared at increasing the affordability of housing.

# 3. Prior Literature

| Study Title | Predictors | Findings | Citation |
| --- | --- | --- | --- |
| Is Housing Unaffordable? Why Isn't It More Affordable? | • Homeownership rate<br>• Income Quintile<br>• Poverty Rate<br>• Race<br>• Ethnicity<br>• Percentage of units renting for less than 30 percent of the median renter's income<br>• Nominal Price of New House | For the owner-occupied sector, lack of affordability is a problem for younger households. Here, modest changes in institutional arrangements could greatly affect the affordability of homeownership, especially for young households whose incomes will increase over the life cycle. For low-income renters, more aggressive policy is needed. | Quigley, John M., and Steven Raphael. "Is housing unaffordable? Why isn't it more affordable?." Journal of Economic Perspectives 18.1 (2004): 191-214. |
| The Rise of Housing Supply Regulation in the U.S.: Local Causes and Aggregate Implications | • Population<br>• Median Hourly Wage<br>• Median House Price<br>• Percentage of College Graduates<br>• Regulation<br>• Metropolitan Statistical Area (MSA) | Endogenous regulation choices account for 23% of the rise in wage inequality and 85% of the increase in house price dispersion between 1980 and 2007.Moreover, the rise of regulation reduces productivity due to misallocation of labor, and results in an output loss of 2.1%. | Parkhomenko, Andrii. "The rise of housing supply regulation in the US: Local causes and aggregate implications." University of Southern California (2018). |
| What are the most important factors that influence the changes in London Real Estate Prices? How to quantify them? | • Real Estate Development Investment<br>• Permanent Dwellings Started<br>• Permanent Dwellings Completed<br>• Interest Rate<br>• Gross Value Added (GVA)<br>• Gross Disposable Household Income<br>• Consumer Price Inflation | Population density, income, and GVA are the most significant factors that affect house price fluctuations in London which indicates that population and income are important point cuts to constrain housing increase. | Gu, Yiyang. "What are the most important factors that influence the changes in London Real Estate Prices? How to quantify them?." arXiv preprint arXiv:1802.08238 (2018). |
| The State of the Nation's Housing 2019 | • Household Income<br>• Housing Construction<br>• Household Growth<br>• Household Type<br>• Age<br>• Market Type<br>• Population<br>• Mortgage Interest Rate | The shortfall in new homes is keeping the pressure on house prices and rents, eroding affordability—particularly for modest-income households in high-cost markets. | Joint Center for Housing Studies. "The state of the nation's housing 2019." (2019). |
| Economic Prosperity and Housing Affordability in the United States: Lessons from the Booming 1990s | • Metropolitan Statistical Area (MSA)<br>• Percentage of cost-burdened households<br>• Race<br>• Poverty Rate<br>• Population<br>• Net Rate of Supply of Housing Units | Economic prosperity by itself cannot increase the availability of affordably priced housing. New, innovative federal, state, and local policies are required to make a significant impact on the nation's chronic and acute affordable housing crisis. | Anthony, Jerry. "Economic prosperity and housing affordability in the United States: lessons from the booming 1990s." Housing Policy Debate 28.3 (2018): 325-341. |
| A new measure of housing affordability: Estimates and analytical results | • Population<br>• Household Income<br>• Housing-Induced Poverty<br>• Cost Burden<br>• Region<br>• Race<br>• Ethnicity<br>• Housing Assistance Status | Official poverty rate underestimates the prevalence of a poverty standard of living. Furthermore, the geographic distribution of official poverty is different from that of housing-induced poverty, which is more concentrated in metropolitan areas. | Kutty, Nandinee K. "A new measure of housing affordability: Estimates and analytical results." Housing policy debate 16.1 (2005): 113-142. |
| Indicators of local housing affordability: Comparative and spatial approaches | • Median Household Income<br>• Median Gross Rent<br>• Shelter Poverty<br>•<br>• Rental Housing Affordability Mismatch Ratio<br>• Metropolitan Statistical Area (MSA) | Local housing assistance plans would benefit from the addition of spatially disaggregated measures of housing affordability. The combination of carefully selected indicators, coupled with maps and simple spatial tools, can be highly revealing of underlying | Bogdon, Amy S., and Ayse Can. "Indicators of local housing affordability: Comparative and spatial approaches." Real Estate Economics 25.1 (1997): 43-80. |

| | | | |
|---|---|---|---|
| | • Population<br>• Average Household Size | housing problems and the geographic distribution of need. | |
| The Impact of Building Restrictions on Housing Affordability | • Housing Supply<br>• Construction Costs<br>• Region<br>• City<br>• Median Family Income<br>• Percentage Population Growth | America is not facing a nationwide affordable housing crisis. In most of the country, home prices appear to be fairly close to the physical costs of construction. The bulk of the evidence marshaled in this paper suggests that zoning, and other land-use controls, are more responsible for high prices. | Glaeser, Edward L., and Gyourko Joseph. "The Impact of Building Restrictions on Housing Affordability." Economic Policy Review, Vol. 9, No. 2, June 2003 |

# 4. Data Source/Preparation

## 4.1. Data Source

Modelling housing affordability is very complex and there are many predictors to look at. The data was sourced from multiple databases like census.gov, laborlawcenter, data.bls.gov, consumerfinance,gov, huduser.gov etc.

| Variables | Source |
|---|---|
| Home values | zillow.com |
| Income | census.gov |
| Min wage | laborlawcenter.com |
| Unemployment | data.bls.gov |
| Demographics | census.gov |
| Mortgage delinquency | consumerfinance.gov |
| Housing assistance | huduser.gov |
| GDP | bea.gov |

*Table 1. Data Sources*

## 4.2. Data Cleaning

The data contained many features grouped state wise from years 2010 to 2019. The attributes were then combined them together using state and year as keys. The dataset also contained many missing values and these were all dropped for the sake of analysis. State and year were converted to factor variables as the data is expected to be of multi-level nature.

# 5. Descriptive Analysis

After the data is loaded as wrangled, the next part of the process is to look at the descriptive statistics. It is important to understand the data and determine the appropriate method to model the problem and identify the relevant predictors. We know that avg home prices distribution is continuous in nature and looking at the distribution of the response variable (avg home price), fig. 1, we can see that it is right skewed. If we take a log transformation of the variable, we get the distribution in fig.2, which is somewhat normal. It is not quite perfect, but it is sufficient to make analysis. Since the response variable is continuous type, a linear model is sufficient.

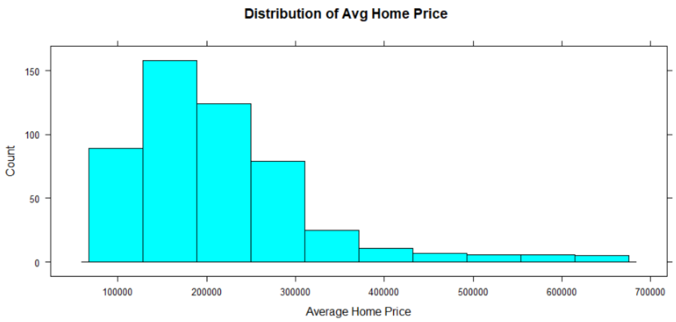



Figure 1. Distribution of Avg Home Price

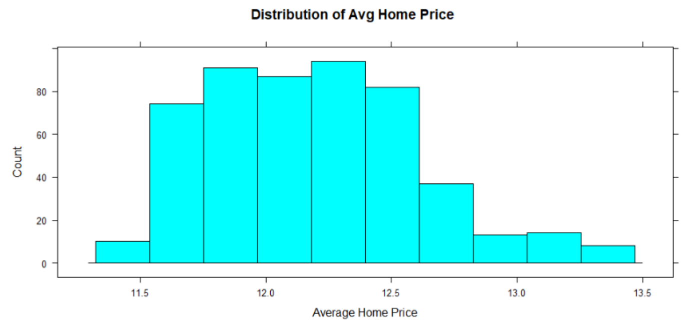



Figure 2. Log transformed distribution of Avg Home Price

The next step is to compute the correlation matrix as shown in fig 3 or fig 4. Looking at it, the first thing note is that gpd is highly correlated with population, subsidized units available and number of houses subsidized. So, we can drop gdp from the data. Population is highly correlated with subsidized units available and num of houses subsidized, but since population is a very import factor of housing price, we can keep it and drop subsidized units available. We can also drop number of houses subsidized from our analysis as its highly correlated with population. We will keep unemployment rate for analysis, as it is a strong predictor for home price. As the unemployment rate goes low, people have more disposable income and this will see a demand for houses. As the demand goes up, the prices are also driven up in the market. Since 30-to-90-day delinquent rate is more common, we will keep it for the analysis.

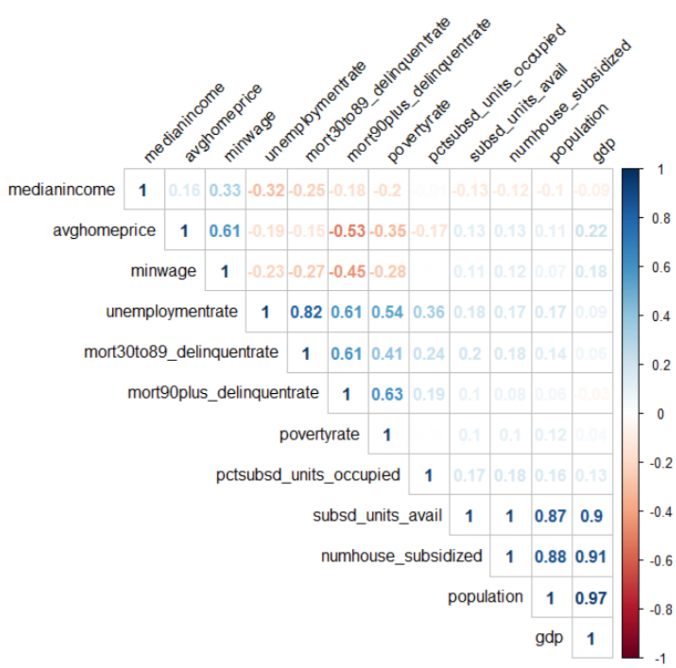



Figure 3. Correlation Matrix

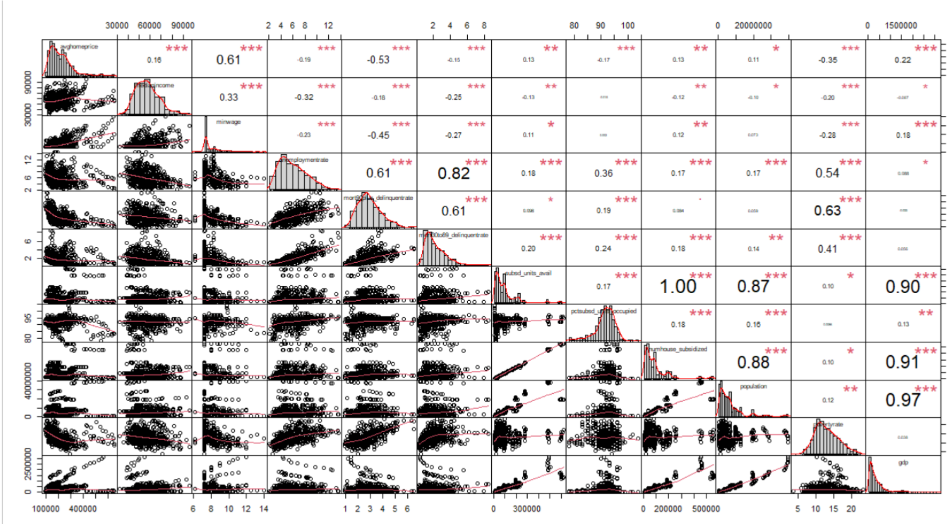


*Figure 4. Correlation Matrix with distribution of predictors*

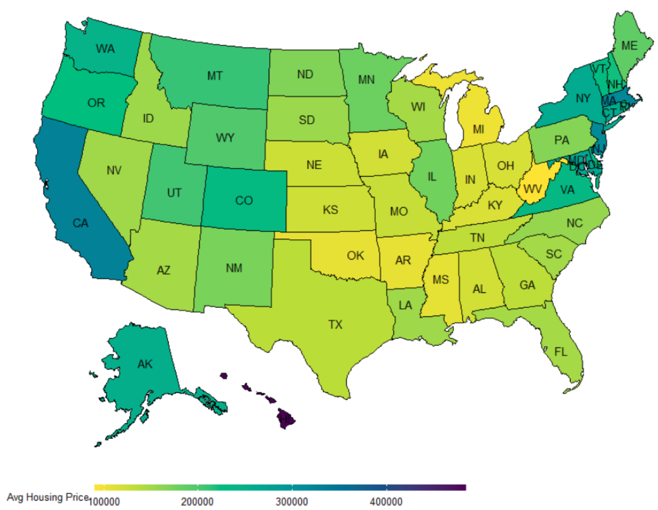
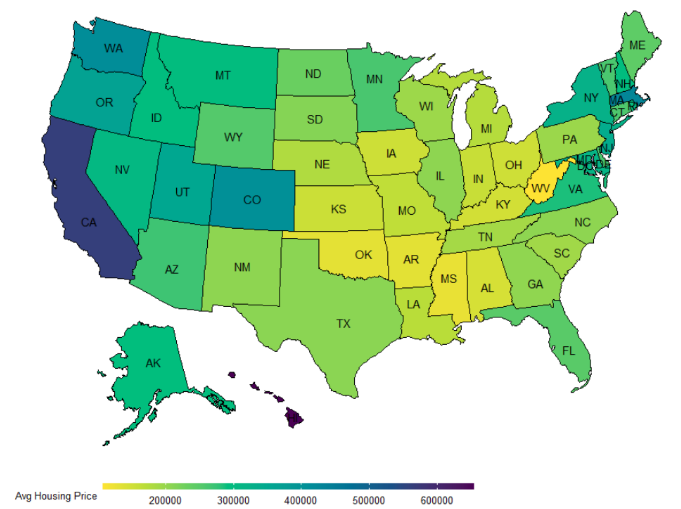
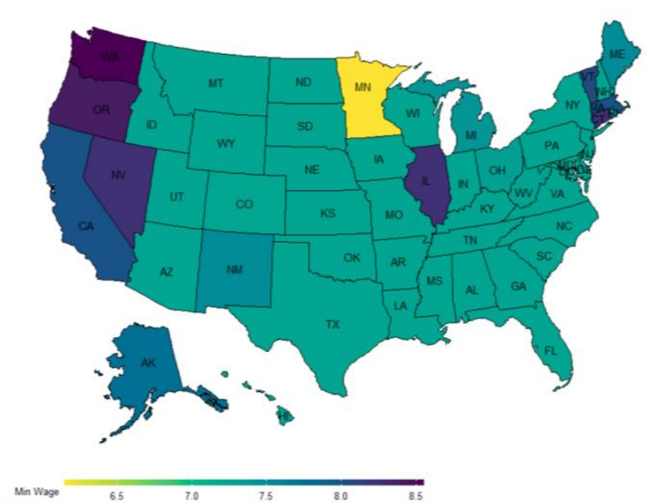
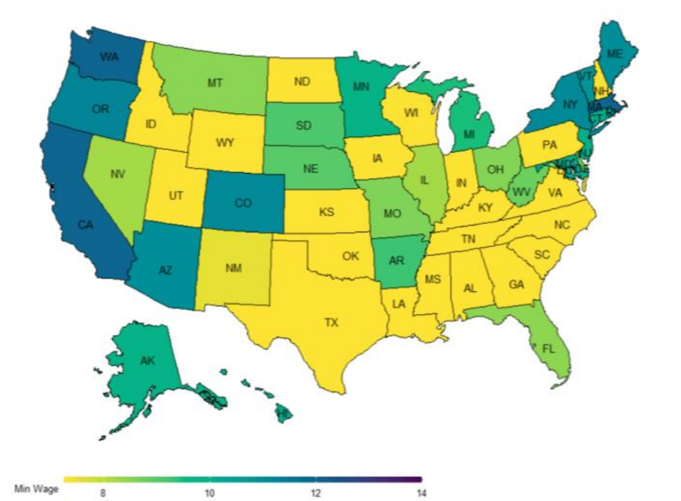
| Attribute | Year: 2009 | Year: 2019 |
|---|---|---|
| **Avg Housing Price** |   |   |
| **Min Wage** |   |   |

*Table 2. Average home price and minimum wage for year 2010 and 2019*

Looking at the table above, avg home price in the years 2010 and 2019, the states with the highest average home prices are Hawaii, District of Columbia, California, and Massachusetts. Hawaii is a tourist/vacation spot, so the prices will be high, and the other states are metro cities located on the west coast and east coast. The avg home prices across the states are in range of 400,000 to 650,000. Similarly, the states with the lowest housing prices are West Virginia, Mississippi, Oklahoma, and Arkansas. These are states without many technical/software hubs and the median income and min wages in these cities are low when compared to metro cities. The avg home prices across the states are in range 100,000 to 150,000. Looking at the min wage in 2010, we see a similar story, the states with highest housing prices are the same states with highest minimum wages, an average of $8.25. The states like West Virginia, Mississippi, Oklahoma, and Arkansas have and avg minimum wage of $7.25. These averages increase to $12 for metro cities in 2019 and other states remain same. The only anomaly is that the state of Hawaii has a one of the lowest minimum wage of $7.25 but also having the highest housing prices in 2019. This is most likely because of the state being very tourist centric. This balances out in 2019 when the min wage increases to $10 for Hawaii.

# 6. Model

Before we start building our models, we need to identify the relevant predictors and form a hypothesis as shown in table 3. When we looked at the min wage and housing prices, we can see that the wage only went up an avg of $2 for some states but the housing prices have gone up almost $200,000. So, there are many other factors to look at while identifying the reasons for surging housing prices.

| Predictors | Effect | Rationale |
|---|---|---|
| Response variable: `AvgHomePrice` | | |
| `medianincome` | + | As the median income in a state increase, we expect to see an increase in the average increase in the housing prices. Since increasing median income means more disposable income. |
| `unemploymentrate` | - | Usually there is an inverse relation between unemployment rate and housing affordability. As unemployment rate goes up, people cannot afford houses and the demand goes down, the housing prices decrease |
| `mort90plus_delinquentrate` | - | The 90-day delinquency rate is a measure of serious delinquencies. It captures borrowers that have missed three or more payments. As the delinquent rate increase, the housing prices are expected to go down as more people cannot afford houses. |
| `mort30to89_delinquentrate` | - | The 30-89 mortgage delinquency rate is a measure of early-stage delinquencies and can be an early indicator of the mortgage market's overall health. It captures borrowers that have missed one or two payments. As the delinquent rate increase, the housing prices are expected to go down as more people cannot afford houses. |
| `pctsubsd_units_occupied` | ? | Occupied units as the percent of units available. The exact effect is unknown, but we expect to see a decrease in housing prices. |

| | | |
|---|---|---|
| minwage | + | As min wage increases, the housing prices are expected to increase. This is mainly since increasing min wage means more income. The housing listing price will then go up. |
| year | ? | The avg housing prices increase as year goes by but it will also depend on the market. So, we must keep it to identify the price trend over the years. Also, to account for multilevel effect, we have to model with year as a predictor. |
| population | + | As population increases, the demand for houses increases, resulting in increased avg housing prices. |
| state | ? | The avg housing prices across the states will vary so used in the model to account for multilevel effect. |
| povertyrate | - | As poverty rate increases, people cannot afford houses, so expect to see decrease in avg housing price |

*Table 3. Predictor Hypothesis Table*

## 6.1. Model

Based on the hypotheses and insights from descriptive analysis, we built the following models. Our target variable average house price is continuous that's why we have constructed a Linear mixed effect model using lme4 and the data is grouped into levels by states and year as average house price may differ across states for different years. Since the distribution of home price is not normal, its log transformed. To account for multilevel nature of data have taken random effect into consideration for years and state. Here level 2 is state and level 1 is home prices from year to year. Like the response variable, population variable and median income variable is also right skewed because of states like California, Texas, Florida, and few others. Therefore, we log transform both variables.

**Model 1:**

```
m1 <- lm(log(avghomeprice) ~ log(medianincome) + unemploymentrate+ mort90plus_delinquentrate+
              mort30to89_delinquentrate + pctsubsd_units_occupied + minwage*year + povertyrate
              + log(population), data = d)
```

The first model is a linear model which is used as baseline. Median income and population were log transformed to obtain a distribution as normal as possible. To analyze the effect of Minimum wage over the year we have added an interaction between year and minimum wage.

**Model 2:**

```
m2 <- lmer(log(avghomeprice) ~ log(medianincome) + minwage*year +unemploymentrate+
              mort90plus_delinquentrate+  mort30to89_delinquentrate + pctsubsd_units_occupied+
              povertyrate + log(population) + (1 | year) + (1 | state), data = d, REML=FALSE)
```

This model is a mixed effect model. As there is expect changes in the housing prices across states and year to year, a lmer model is used. Minimum wage can't be interpreted as a unique effect on Average house price. There is a possibility that wages were low in year 2010 and because of the economic growth over the years minimum wage has increased, that's the reason interaction between Minimum wage and year is added in model2 to account the change in Minimum wage with respect to time/year. Year and state re used as independent randomized effect.

**Model 3:**

```
m3 <- lmer(log(avghomeprice) ~ log(medianincome) + minwage + unemploymentrate+
           mort90plus_delinquentrate+ mort30to89_delinquentrate + pctsubsd_units_occupied+
           povertyrate + log(population) + (1 | year) + (1 | state), data = d, REML=FALSE)
```

Model 3 is same as the model 2, and it's used for checking multicollinearity

| Predictor | Model 1 | Model 2 | Model 3 |
|---|---|---|---|
| log(medianincome) | -0.015 (0.082) | 0.045 (0.041) | 0.063 (0.039) |
| unemploymentrate | 0.015 (0.014) | -0.018*** (0.003) | -0.019*** (0.003) |
| mort90plus_delinquentrate | -0.214*** (0.022) | -0.003 (0.011) | -0.001 (0.010) |
| mort30to89_delinquentrate | 0.111*** (0.019) | -0.017*** (0.004) | -0.019*** (0.004) |
| pctsubsd_units_occupied | -0.008** (0.004) | -0.003*** (0.001) | -0.004*** (0.001) |
| minwage | 0.151 (0.097) | -0.001 (0.016) | 0.005 (0.003) |
| year2011 | 0.227 (0.976) | 0.063 (0.142) | |
| year2012 | 0.540 (0.929) | 0.003 (0.136) | |
| year2013 | 0.080 (0.909) | -0.153 (0.133) | |
| year2014 | -0.514 (0.859) | -0.163 (0.128) | |
| year2015 | -0.344 (0.810) | -0.144 (0.124) | |
| year2016 | -0.067 (0.774) | -0.106 (0.120) | |
| year2017 | 0.016 (0.754) | -0.078 (0.118) | |
| year2018 | 0.191 (0.745) | -0.046 (0.118) | |
| year2019 | 0.413 (0.738) | 0.007 (0.118) | |
| povertyrate | -0.020***(0.005) | -0.001 (0.002) | -0.001 (0.002) |
| log(population) | -0.018 (0.013) | 2.012*** (0.110) | 2.004*** (0.111) |
| minwage:year2011 | -0.039 (0.131) | -0.018 (0.019) | |
| minwage:year2012 | -0.076 (0.124) | -0.011 (0.018) | |
| minwage:year2013 | -0.009 (0.122) | 0.012 (0.018) | |
| minwage:year2014 | 0.065 (0.114) | 0.012 (0.017) | |
| minwage:year2015 | 0.045 (0.107) | 0.010 (0.016) | |
| minwage:year2016 | 0.012 (0.103) | 0.008 (0.016) | |
| minwage:year2017 | 0.004 (0.101) | 0.008 (0.016) | |
| minwage:year2018 | -0.015 (0.100) | 0.007 (0.016) | |
| minwage:year2019 | -0.040 (0.099) | 0.003 (0.016) | |
| Constant | 12.703***(1.164) | -18.265*** (1.773) | -18.378*** (1.777) |
| R2 | 0.57 | | |
| Adjusted R2 | 0.547 | | |
| Log Likelihood | | 650.93 | 626.239 |
| Akaike Inf. Crit. | | -1,243.86 | -1,228.48 |
| Bayesian Inf. Crit. | | -1,121.06 | -1,177.67 |
| Residual Std. Error | 0.279 (df = 483) | | |
| F Statistic | 24.655 | | |

*Table 4. Stargazer Output*

## 6.2. Model Evaluation

We know that the model is multilevel and so we will not be selecting model 1. To select between model 2 and we can use the AIC and BIC. Akaike's Information Criterion (AIC) is an estimator of the relative quality of statistical models for given set of data. AIC provides a mean for model selection. Lower the AIC value better the model. Bayesian information criteria (BIC) is a variant of AIC with a stronger penalty for including additional variables to the model. Looking at the AIC metric, models 2 through 3 are all within a couple points of each other. Model 2 has lowest AIC and BIC score and therefore we can go with model 2.

## 6.3. Assumptions

We can perform various assumption test to check for the quality of the model.

- **Multicollinearity**

  We ran Variance Inflation Factor (VIF) test on Model 3 which is used to measure how much the variance of a regression coefficient is inflated due to multicollinearity in the model. VIF value for all the predictor are less than 5 indication that there is no correlation among the predictors.

```
         log(medianincome)                    minwage           unemploymentrate mort90plus_delinquentrate
                  1.042529                   1.092325                   1.846450                  1.550547
mort30to89_delinquentrate  pctsubsd_units_occupied                povertyrate           log(population)
                  1.813582                   1.108704                   1.173789                  1.181631
```

*Figure 5. VIF test*

- **Autocorrelation/ Independence**

  The Durbin Watson (DW) statistic is a test for autocorrelation in the residuals from a statistical regression analysis. DW = 1.9236, p-value = 0.3872. Durbin-Watson statistic have value close to 2.0 meaning that there is no autocorrelation detected in the sample.

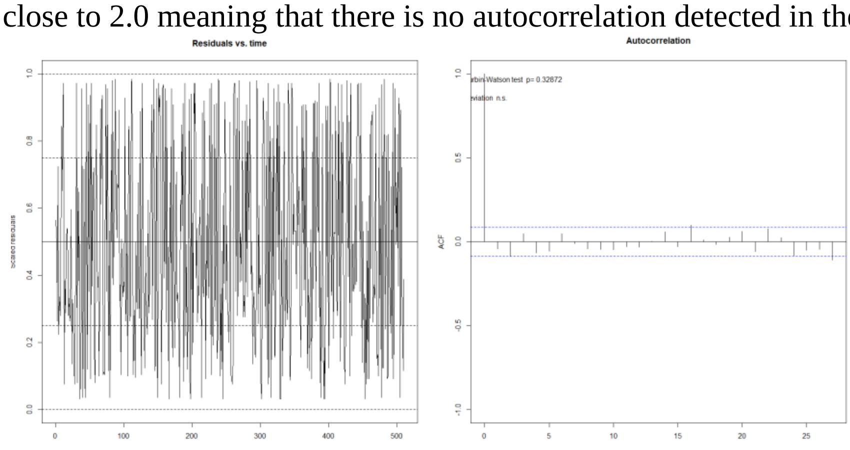



*Figure 6. Autocorrelation Test*

## 7. Results & Interpretations

| Predictors | Interpretations |
|---|---|
| `log(medianincome)` | 1 % increase in medium income will increase average house price by 0.045 % |
| `unemploymentrate` | 1 unit increase in unemployment rate will decrease average house price by 1.8 % |
| `mort90plus_delinquentrate` | 1 unit increase in Mortgage rate after 90 days will decrease average house price by 0.3% |
| `mort30to89_delinquentrate` | 1 unit increase in Mortgage rate for 30 to 90 days will decrease average house price by 1.7% |
| `pctsubsd_units_occupied` | 1 unit increase in subsidized unit occupied will decrease average house price by 0.03% |
| `minwage` | 1 unit increase in Minimum wage will decrease average house price by 0.01%, keeping all the variable constant.<br>Interaction with year:<br>Minimum wage will vary according to year. 2013 and 2014 have the highest effect on Average house price when base level is 2010. 2011 have the least effect on Average house price. |
| `year` | Year 2014 has highest average house price, exceeding 2010 by 16.3%. The effect is decreasing till 2019. |
| `povertyrate` | 1 unit increase in Poverty rate will decrease average house price by 0.1% |
| `log(population)` | 1 % increase in population will increase average house price by 2 % |

## 8. Conclusion & Recommendations

Housing Affordability is a very common problem that we've all been effect by throughout the last decades. It was the objective of this analysis to find reasonable solutions and insights and give actionable recommendations so we can tackle the issue. Based on the given data and analysis, increase in the percentage of subsided houses occupied resulted in decrease in the average home price by 0.4%. So, policy makers can formulate policies to increase the percentage of subsided houses in the state. This can lower the excessive demand for the houses and lower the average home price. Increasing the minimum wage is also a recommendation, as change in wage results in increase in housing price by 0.5%. By doubling the minimum wage in the current market means people can have more disposable income. This could also mean they can qualify for better mortgage, and have affordable housing cost.

## 9. References


[1] "Housing Data." *Zillow Research*, 25 Mar. 2021, www.zillow.com/research/data/.

[2] Bureau, US Census. "Historical Income Tables: Households." *The United States Census Bureau*, 8 Sept. 2020, www.census.gov/data/tables/time-series/demo/income-poverty/historical-income-households.html.

[3] *2021 Minimum Wage Rates by State*, www.laborlawcenter.com/state-minimum-wage-rates/.

[4] "Unemployment Statistics Map." *U.S. Bureau of Labor Statistics*, U.S. Bureau of Labor Statistics.

[5] Bureau, US Census. "State Population by Characteristics: 2010-2019." *The United States Census Bureau*, 20 Apr. 2021, www.census.gov/data/tables/time-series/demo/popest/2010s-state-detail.html.

[6] "Download the Data." *Consumer Financial Protection Bureau*, www.consumerfinance.gov/data-research/mortgage-performance-trends/download-the-data/.

[7] "Family Data on Public and Indian Housing." *Assisted Housing: National and Local | HUD USER*, www.huduser.gov/portal/datasets/assthsg.html#2009-2020_data.

[8] "GDP: U.S. Bureau of Economic Analysis (BEA)." *GDP | U.S. Bureau of Economic Analysis (BEA)*, www.bea.gov/data/gdp.

# 10. Appendix

```
#Setting the Working Directory
setwd(dirname(rstudioapi::getActiveDocumentContext()$path))

#clear environment & plot
rm(list=ls())
dev.off()
options(scipen=999)

#load relavent libraries
library(pacman)
p_load(rio, stargazer,Hmisc, PerformanceAnalytics, tidyverse)

#import main data
df <- import("housingData.xlsx")
names(df) <- stringr::str_to_lower(names(df))
str(df)

#import gdp data
gdp <- read.csv("current_dollar_gdp.csv", header = TRUE, sep = ",")
colnames(gdp)[3:13] <- c("2009":"2019")
gdp$`2009` <- NULL
gdp %>% pivot_longer(col = starts_with("2"), names_to = "Year", values_to = "GDP" ) -> gdp
str(gdp)

names(gdp) <- stringr::str_to_lower(names(gdp))
names <- c("state", "year", "region")
gdp[,names] <- lapply(gdp[,names],factor)

# merge temp & stores dataframes by store
d <- merge(df, gdp, by=c("year", "state"))

#categorize features,
names <- c("state", "year")
d[,names] <- lapply(d[,names],factor)
colnames(d)[7:11] <- c("mort90plus_delinquentrate","mort30to89_delinquentrate", "subsd_units_avail",
                       "pctsubsd_units_occupied", "numhouse_subsidized")

#check for missing
which(!complete.cases(d))
colSums(is.na(d))


#Data Visualizations
hist(d$avghomeprice)
hist(log(d$avghomeprice))

library(lattice)
histogram(~avghomeprice,data=d,
          type="count",
          xlab="Average Home Price",
          main="Distribution of Avg Home Price")

hist.data.frame(d)

library(purrr)
d %>% keep(is.numeric) %>% pairs(. , panel= panel.smooth)
d %>% keep(is.numeric) %>% chart.Correlation(., histogram=TRUE, pch=19)


#https://liuyanguu.github.io/post/2019/04/17/ggplot-heatmap-us-50-states-map-and-china-province-map/
par(mfrow = c(1,2))

start = levels(d$year)[1]
end = levels(d$year)[10]
```

```r
library(usmap)
for (i in c(start,end)){
  df = subset(d, year == i)
  map <- plot_usmap(data = df, values = "minwage", labels = T) +
          labs(fill = 'Min Wage') +
          scale_fill_gradientn(colours=hcl.colors(10, rev = TRUE),na.value="grey",
                               guide = guide_colourbar(barwidth = 25, barheight = 0.4)) +
          theme(legend.position = "bottom",
                legend.title=element_text(size=10),
                legend.text=element_text(size=10))

  print(map)

}

d %>% subset(d$year == 2019) %>% group_by(region) %>% summarise(mean = mean(avghomeprice)) %>%
arrange(mean)
d %>% subset(d$year == 2019) %>% group_by(state) %>% summarise(homeprice = avghomeprice) %>%
arrange(homeprice) %>% top_n(-5)



library(corrplot)
d %>% keep(is.numeric) %>% cor() %>% corrplot(., method = "number", type = "upper", order =
"hclust", tl.col = "black", tl.srt = 45)

d$numhouse_subsidized <- NULL
d$gdp <- NULL
d$subsd_units_avail <- NULL

with(d, interaction.plot(year,minwage,avghomeprice )) #interaction check


#Model
library(lme4)
m1 <- lm(log(avghomeprice) ~ log(medianincome) + unemploymentrate+ mort90plus_delinquentrate+
             mort30to89_delinquentrate + pctsubsd_units_occupied + minwage*year + povertyrate +
             log(population), data = d)
m2 <- lmer(log(avghomeprice) ~ log(medianincome) + minwage*year +unemploymentrate+
mort90plus_delinquentrate+
             mort30to89_delinquentrate + pctsubsd_units_occupied+ povertyrate +
             log(population) + (1 | state) +  (1 | state), data = d, REML=FALSE)
m3 <- lmer(log(avghomeprice) ~ log(medianincome) + minwage + unemploymentrate+
mort90plus_delinquentrate+
             mort30to89_delinquentrate + pctsubsd_units_occupied+ povertyrate +
             log(population) + (1 | year) + (1 | state), data = d, REML=FALSE)
summary(m1)
summary(m2)
stargazer(m1, m2, m3, type='text', single.row = TRUE)


#model assumptions
library(DHARMa) # for autocorrelation test

lmer_assumptions <- function(model){
  #' Tests for assumptions

  print(vif(model))                                    # Variance inflation factor

  print(testTemporalAutocorrelation(model, time = NULL)) # Autocorrelation

}


lmer_assumptions(m2)
```