EC 420 Final Project - Housing Prices and Election Years

David Selemba

2023-04-10

#Research Question/Motivation: Does the instability of election years impact the US housing market?

Due to the wild surges in the real estate market of the past few years, I thought it would be interesting to see if there is any correlation on large shifts relating to upcoming elections in the US. With politicians more focused on their campaigns and reelection or becoming lame-duck in office, they may have less focus towards stimulating the economy through policy. In addition, since US home buyers will have less certainty on what they can expect from their government and both local and national politicians, they will not be able to as accurately predict their feelings of the market and may be less inclined to make this large financial commitment.

#Method:

The econometric model I used to carry out this analysis was a multivariate regression model. The choice of model was due to the numerous factors that play into the property valuation and sale prices, with my analysis accounting for a number of different macroeconomic factors. These features, such as the unemployment rate, inflation, mortgage rate, foreclosure rate, and change in GDP by quarter, as well as cost of rental homes later in the analysis. With the use of a dummy variable, we will be able to model for whether or not it was an election year and see the correlation this has with housing prices. I also created a model where the rent CPI is added to the multivariate regression, as this is seemingly a ommitted variable when initially carrying out my models.

#Data:

The data I am using comes from a number of different sources, including Kaggle and the St. Louis Federal Reserve. Each column of the below combined data frame housing is as follows: ***HPI*** - Housing Price Index, index that measures the movement of a single family home over a period of time, with the first quarter as the base(Dependent variable in models) ***DATE*** - current quarterly date at time of observation ***UNEM_RATE*** - current unemployment rate during the quarter ***MORTGAGE*** - current mortgage rate during the quarter ***Real.GDP..Percent.Change.*** - real GDP adjusted for inflation during the quarter ***FORECLOSURES*** - foreclosures per 100 mortgages per quarter ***Inflation.Rate*** - current inflation rate per quarter ***Effective.Federal.Funds.Rate*** - current federal funds rate per quarter ***Year*** - current year ***Rent CPI*** - Consumer Price Index of the cost of renting a home

```
#Read in data for the project
housing_demand = read.csv('demand_data.csv')
inflation = read.csv('index.csv')
```

```
#Manipulate inflation/demand data set and add to housing set
inflation$DATE = as.Date(paste(inflation$Year, inflation$Month, inflation$Day, sep = "-"))
inflation = inflation%>%
  filter(Year >= 2000, Month == 1 | Month == 4 | Month == 7 | Month == 10, Day == 1)
inflation_drop = inflation%>%
  select(-Month, -Day, -Federal.Funds.Upper.Target, -Federal.Funds.Target.Rate, -Federal.Funds.Lower.Ta
inflation_drop = inflation_drop[,c(5,4,2,3,1)]

housing_demand = housing_demand[1:69,]
```

```r
housing_demand$DATE = as.Date(housing_demand$DATE)

housing = merge(housing_demand, inflation_drop, by = "DATE")
head(housing)
```

```
##         DATE UNEM_RATE MORTGAGE    GDPC1 FORECLOSURES    HPI Inflation.Rate
## 1 2000-01-01      4.03     8.26 12935.25         1.95 101.34            2.0
## 2 2000-04-01      3.93     8.32 13170.75         1.97 103.67            2.3
## 3 2000-07-01      4.00     8.02 13183.89         2.09 105.79            2.5
## 4 2000-10-01      3.90     7.62 13262.25         2.23 108.27            2.5
## 5 2001-01-01      4.23     7.01 13219.25         2.34 110.48            2.6
## 6 2001-04-01      4.40     7.13 13301.39         2.41 112.20            2.6
##   Effective.Federal.Funds.Rate Real.GDP..Percent.Change. Year
## 1                         5.45                       1.2 2000
## 2                         6.02                       7.8 2000
## 3                         6.54                       0.5 2000
## 4                         6.51                       2.3 2000
## 5                         5.98                      -1.1 2001
## 6                         4.80                       2.1 2001
```

```r
#Summary statistics for HPI to see bounds of data frame
summary(housing$HPI)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   101.3   138.8   150.1   151.7   171.6   187.4
```

#Results:

```r
#Creating first model without the election year dummy
housing_fit = lm(HPI ~ UNEM_RATE + MORTGAGE + Real.GDP..Percent.Change. + Inflation.Rate +FORECLOSURES,
summary(housing_fit)
```

```
##
## Call:
## lm(formula = HPI ~ UNEM_RATE + MORTGAGE + Real.GDP..Percent.Change. +
##     Inflation.Rate + FORECLOSURES, data = housing)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -32.669 -14.673  -1.438  11.760  36.313
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 2.668e+02  2.638e+01  10.112 9.69e-15 ***
## UNEM_RATE                  -3.220e+00  3.331e+00  -0.967    0.337
## MORTGAGE                   -1.497e+01  2.767e+00  -5.412 1.07e-06 ***
## Real.GDP..Percent.Change. -9.462e-01  1.006e+00  -0.941    0.351
## Inflation.Rate             6.303e-04  6.949e+00   0.000    1.000
## FORECLOSURES               -2.588e+00  1.810e+00  -1.430    0.158
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 19.03 on 62 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.3901, Adjusted R-squared:  0.3409
## F-statistic:  7.93 on 5 and 62 DF,  p-value: 7.96e-06
```

This first model shows mostly negative correlations for all variables involved in the multivariate regression. This is to be expected, as higher unemployment, mortgage rates, and foreclosures per 100 would all signal difficult economic times and in turn lower housing costs. The only value that shows as statistically significant, at least by 0.1, is mortgage rate, as these are seemingly heavily correlated. In addition, the inflation feature sees only a small increase in HPI (0.115) for every 1% increase in the rate. It also has a p-value of 1, so due to it's statistical insignificance and very small impact on the estimate, we can exclude this from further analysis entirely.

```
#Create a dummy variable based on whether the current year is an election year (i.e. 2000, 2004, etc.)
housing$election_year = 0
housing$election_year[housing$Year %in% c(2000,2004,2008,2012,2016)] = 1
```

```
#Creating the model with the dummy variable for election year
housing_fit_election = lm(HPI ~ UNEM_RATE + MORTGAGE + Real.GDP..Percent.Change. +FORECLOSURES + electi
summary(housing_fit_election)
```

```
##
## Call:
## lm(formula = HPI ~ UNEM_RATE + MORTGAGE + Real.GDP..Percent.Change. +
##     FORECLOSURES + election_year, data = housing)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -37.204 -12.155   0.232  12.866  33.153
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                274.8498    19.3299  14.219  < 2e-16 ***
## UNEM_RATE                   -4.4912     2.9643  -1.515    0.135
## MORTGAGE                   -14.9019     2.6086  -5.713 3.39e-07 ***
## Real.GDP..Percent.Change.   -1.1678     0.9626  -1.213    0.230
## FORECLOSURES                -2.1324     1.7531  -1.216    0.228
## election_year               -8.5625     5.2387  -1.634    0.107
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18.63 on 62 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.4153, Adjusted R-squared:  0.3681
## F-statistic: 8.806 on 5 and 62 DF,  p-value: 2.348e-06
```

After adding the election year variable to the model, we see that it has a -8.5625 decrease in HPI for the quarter when it is an election year. However, the p-value just misses significance at the alpha = 0.1 level, so we cannot determine for certain how important this feature is to our model. It would seemingly be beneficial to have data from years prior to 2000 for more certainty on the importance of this feature. The impact on r-squared was only minor as well, moving from 0.3409 to 0.3681, which means it only plays a small part in explaining the variance on its own.

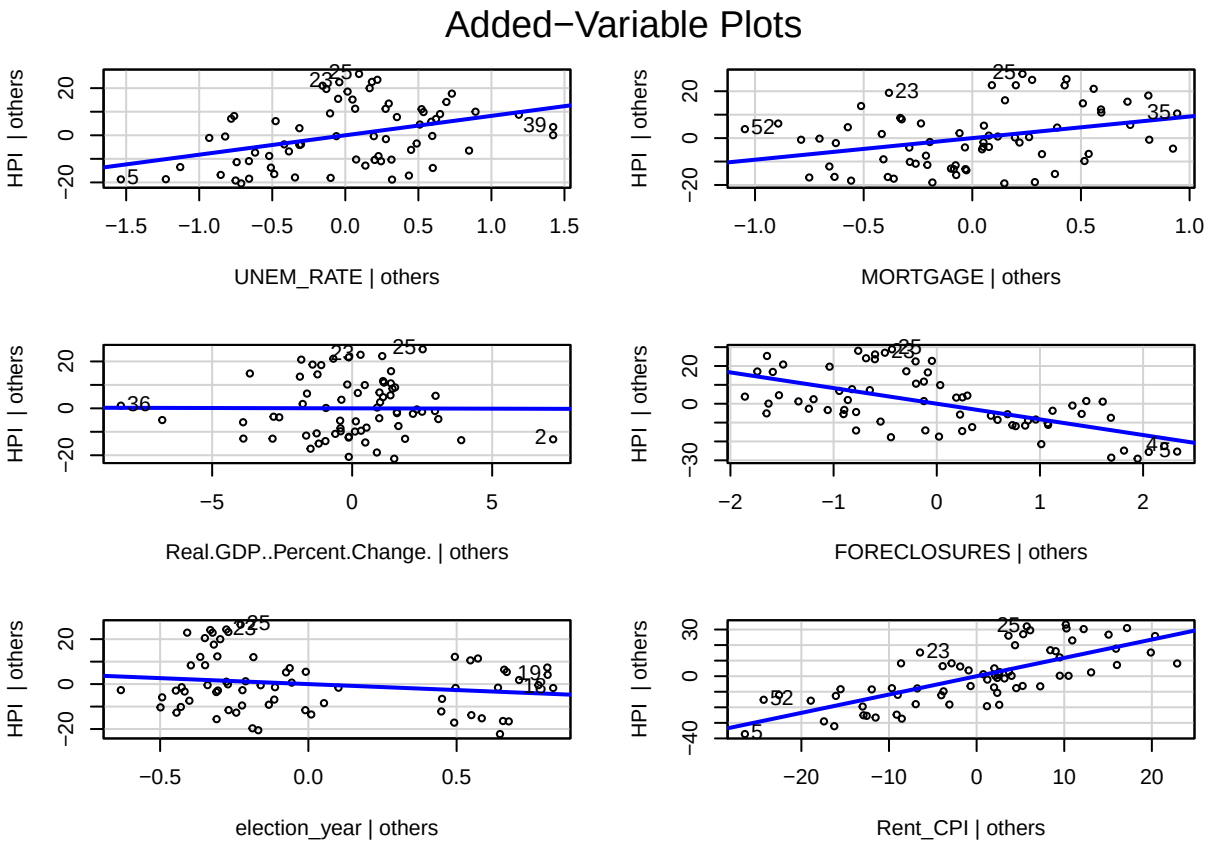
```
#Adding the home rental price variable to the data frame
rentals = read.csv('rentals.csv')
rentals$DATE = as.Date(rentals$DATE)
rentals <- subset(rentals, format(DATE, "%m") %in% c("01", "04", "07", "10"))
colnames(rentals) = c("DATE", "Rent_CPI")
housing_rentals = merge(housing, rentals, by = "DATE")
head(housing_rentals)
```

```
##         DATE UNEM_RATE MORTGAGE    GDPC1 FORECLOSURES    HPI Inflation.Rate
## 1 2000-01-01      4.03     8.26 12935.25         1.95 101.34            2.0
## 2 2000-04-01      3.93     8.32 13170.75         1.97 103.67            2.3
## 3 2000-07-01      4.00     8.02 13183.89         2.09 105.79            2.5
## 4 2000-10-01      3.90     7.62 13262.25         2.23 108.27            2.5
## 5 2001-01-01      4.23     7.01 13219.25         2.34 110.48            2.6
## 6 2001-04-01      4.40     7.13 13301.39         2.41 112.20            2.6
##   Effective.Federal.Funds.Rate Real.GDP..Percent.Change. Year election_year
## 1                         5.45                       1.2 2000             1
## 2                         6.02                       7.8 2000             1
## 3                         6.54                       0.5 2000             1
## 4                         6.51                       2.3 2000             1
## 5                         5.98                      -1.1 2001             0
## 6                         4.80                       2.1 2001             0
##   Rent_CPI
## 1    181.1
## 2    182.3
## 3    183.9
## 4    186.1
## 5    188.2
## 6    190.2
```

```
#Creating the multivariate model with the rental price variable
rentals_fit = lm(HPI ~ UNEM_RATE + MORTGAGE + Real.GDP..Percent.Change. +FORECLOSURES + election_year +
summary(rentals_fit)
```

```
##
## Call:
## lm(formula = HPI ~ UNEM_RATE + MORTGAGE + Real.GDP..Percent.Change. +
##     FORECLOSURES + election_year + Rent_CPI, data = housing_rentals)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -21.439 -10.101  -1.459   9.203  25.261
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                -185.23161   55.94259  -3.311  0.00156 **
## UNEM_RATE                     8.23691    2.52371   3.264  0.00180 **
## MORTGAGE                      9.21133    3.36121   2.740  0.00804 **
## Real.GDP..Percent.Change.    -0.02839    0.67173  -0.042  0.96643
## FORECLOSURES                 -8.29852    1.40265  -5.916 1.62e-07 ***
## election_year                -5.31781    3.60219  -1.476  0.14501
## Rent_CPI                      1.17818    0.13920   8.464 7.03e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.74 on 61 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.7311, Adjusted R-squared:  0.7046
## F-statistic: 27.64 on 6 and 61 DF,  p-value: 1.124e-15
```

```
#Create a plot that shows fitted line on relationship of variables when all other predictor features ar
avPlots(rentals_fit)
```



The final model that incorporates the current rent CPI shows a positive correlation between the two variables, stating that a one unit change in rent CPI results in estimated increase of 1.17818 in HPI. This would make sense, as when renting a house it costly, it would also in turn make buying a home more costly as the markets would move to meet one another's direction. This feature is significant at a near-zero level, while also making a number of other large estimators significant as well. This also changed the sign on the estimators for mortgage and unemployment rate, meaning an increase in unemployment rate will cause an increase in the HPI. The change means that these variables are highly correlated with rent costs, so by excluding it from earlier models, there was some bias in the results.

#Conclusions:

After interpreting the results, it can be concluded that election years do have a negative correlation to the estimated cost of purchasing a single family home, although we can not say at what significance level. It is also seen now that the price of rental properties and homes plays the largest significance in our models in determining the HPI for the US. This makes intuitive sense in that when renting costs increase, buying costs will increase unilaterally. Using this information, it would seemingly make sense to purchase a home during the election years if one knows they will be financially stable in upcoming years despite the possible political and economic uncertainty. To carry out further research, the first possible task would be to find accurate data for years preceding 2000 to find if this conclusion is held through a greater number of observations. Some fo the data could have been heavily impacted by the economic recession of 2008 and caused some trends that are not as common, so added data could allow for greater normalization. It would also be useful to carry out this analysis by region, as HPI is an aggregated variable, so it may be possible that some areas see positive shifts during election years while others see the same correlation. This could be due to the demographic of the region or the local election outcomes as well. In the future, it may also be useful to explore other models that fit the data as well, as a more advanced machine learning model may allow us to optimize the features used to predict well into the future.