Market factors affecting housing prices

Subtitle

E. Whitehead, Mohamed XX, Alejandro XX, & M. Troeger

ECON 4580/6030 - Data Analysis in Economics and Finance
Dr. Rui Fan

Department of Economics
Rensselaer Polytechnic Institute
Troy, NY, USA
11 Nov. 2024

Contents

Realtor Data

To see the effect of various market features on the median listing price of houses, we use Resedential Data from Realtor.com which includes various housing market features listed by county Federal Information Processing Series (FIPS) codes and by date. We extract `year` from the `date` field and, for parsimony, deselect all time variables with the exception of `year` and summarize the dataset by mean over `year` and `fips`. This greatly reduces dimensionality and aids computation.

The data set contains 10,176 `NA` values from 4,761 observations (approximately 16.9% of all observations) which predominantly belong to rural counties where such information is difficult to obtain or is simply unavailable. Consequently, we elide these variables in coming analysis. Even with 16.9% of observations removed, Figure 1 indicates that most of the country is still represented (counties with data are colored red):

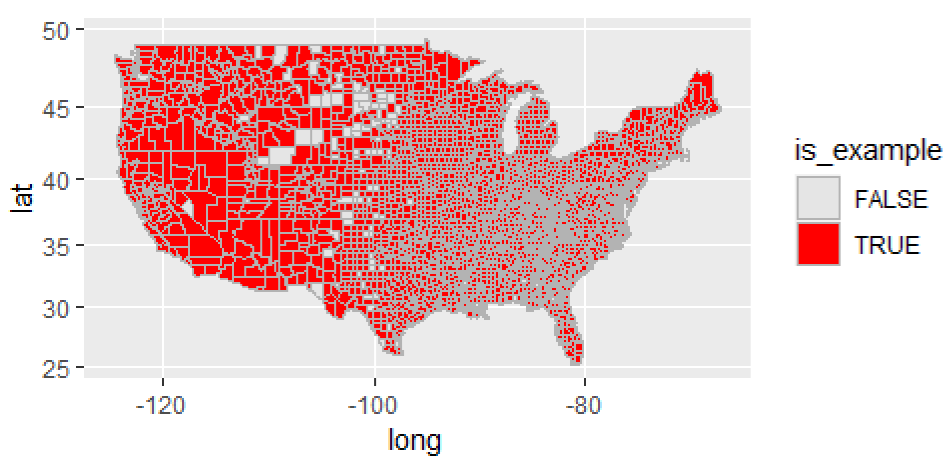

Figure 1: Counties Represented by Realtor with `NA`s Elided

We further augment the data set by adding annually averaged consumer price index (CPI) as a measure of inflation. Table 1 visually summarizes the adjusted data set:

Variable	μ	σ	**Min**	**Max**
fips	30642.6	14957.9	01001	56045
year	2020.3	2.54931	2016	2024
median_listing_price	270042.4	196290.3	35750	4780104.1
active_listing_count	337.33	920.55	1	21778.8
median_days_on_market	71.108	23.526	4	283
new_listing_count	154.409	422.650	0	8615.33
price_increased_count	11.18	47.42	0	1106.67
price_reduced_count	97.23	317.94	0	9120
pending_listing_count	165.70	498.57	0	11212
median_listing_price_per_square_foot	143.27	99.41	20	1877.11
median_square_feet	1874.96	326.90	576	4844.09
average_listing_price	377056.26	353401.63	35750	11998834.08
total_listing_count	501.50	1361.03	1	30349.33
pending_ratio	0.541	0.472	0	5.581
cpi	0.031	0.0219	0.012	0.08

Table 1: Realtor Data Set Summary

Realtor Data Processing

Before we begin our analysis, we start with an exploratory ordinary least squares (OLS) regression on the dependent variable `median_listing_price` to identify variables with a variance inflation factor (VIF) greater than 10, and then deselect those variables as well as those that *cheat* by also reflecting pricing information. In so doing, we remove `active_listing_count`, `total_listing_count`, `pending_listing_count`, `median_listing_price_per_square_foot`, and `average_listing_price`. We further subset the data to exclude entries for 2024, which is incomplete and does not have an annual average CPI value.

We further observe that a handful of counties are significantly wealthier than others, and consequently have much larger median listing prices. Thus, we programatically remove outliers within the bottom 5% and top 5% of `median_listing_price`. To account for time effects in the panel data, we encode `year` as a factor which ensures they enter our regression as a dummy variable. We then adjust each observation's value of `median_listing_price` by its respective inflation rate. Finally, we eliminate spatial effects by group demeaning the `median_listing_price` by FIPS and by introducing state factors which we favor in place of FIPS factors.

Realtor Analysis

We now specify an unrestricted OLS regression with `median_listing_price` as our dependent variable and employ backward stepwise selection to produce a model with only the most important features.

We arrive at a model of the form

$$\begin{aligned} median\widehat{Listing}Price = 44338.1 &- 622.5 \cdot medianDaysOnMarket \\ &- 5.1 \cdot newListingCount \\ &- 111.9 \cdot priceIncreasedCount \\ &- 2.4 \cdot medianSquareFeet \\ &+ 23065.3 \cdot pendingRatio \\ &+ state\ factors \end{aligned}$$

With $\bar{R}^2 = 0.2$ and $F = 93.1$. Figure 2 depicts the relationship between $medianListingPrice$ against $year$, with data points in blue and our predicted $median\widehat{Listing}Price$ in red:

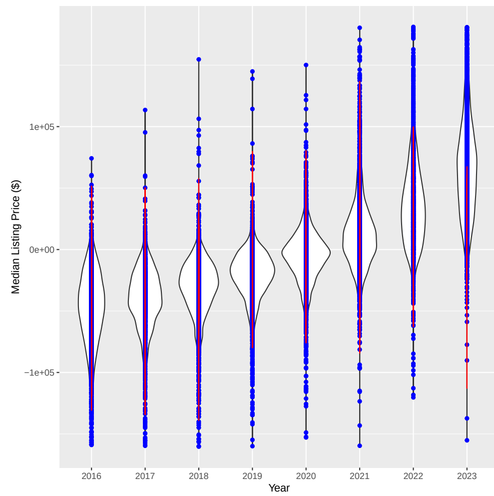

Figure 2: Median Listing Price vs. Year Violin Plot

Despite controlling for inflation, we have a strong time trend. Of interest is the relatively large and highly significant coefficient on *pendingRatio*: $\hat{\beta} = \$23,065$ with $p << 0.05$. Figure 3 shows the change in *pendingRatio* with time:

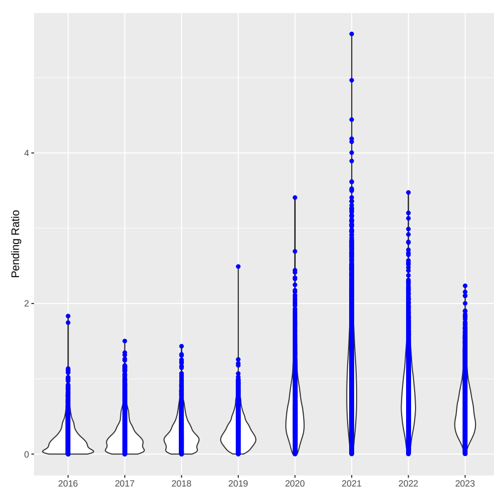

Figure 3: Pending Ratio vs. Year

As the pending ratio of housing is calculated as the share of pending listings over the share of active listings, an increase therefore indicates either a great leap in pending listings, or a major contraction in active listings. Because the bump takes place during the COVID-19 pandemic, it is likely the decrease comes from a precipitous drop in active listings as work from home became the norm and people were largely locked in place. The Realtor data for *activeListingCount* supports this, marking a continuous downward trend starting in March 2020 and having a point of inflection in February 2022. These findings are consistent with those of Yörök (2022).