Macroeconomic Factors that Affect the Price of Housing in the United States: Evidence from a State-Level Panel Data Analysis

Séan Ronan

Department of Economics and Finance
School of Business and Economics
SUNY Plattsburgh
Srona002@plattsburgh.edu
Sean_ronian22@yahoo.com

May 2023

Abstract

Due to all the changes in the economy recently the goal of the study is to see how specific macroeconomic factors have affected the housing market in the past to gain some insight into where the overall housing market may trend in the future. This paper offers evidence from a state-level panel data analysis from 1975-2021 on the effects that macroeconomic factors have on home prices in the United States. This analysis focuses on state and country level variables to determine their effect on the housing price index. The results produced by this model indicate that increases in the unemployment rates, the annual supply of new homes, and the federal funds effective rate are connected to decreases in the housing price index. While increases in state minimum wages and inflation can be attributed to increases in home prices. All the variables in the model are statistically significant at the 1% level and the model has an R-squared of .89.

Keywords: Panel-data, Housing Market, Supply and demand, Unemployment, Interest Rates, Inflation, Economic Growth
1. Introduction

In the United States and worldwide, the housing market has experienced significant changes in recent decades. As shown in Figure 1 it is a dynamic and complex system that varies significantly from state to state. From the housing crisis in 2008 to the rise in suburbanization in the 1970s. The housing market has been molded by a wide range of social, economic, and political factors. Homes make up a large majority of household wealth for Americans which is why this is such a heavily studied topic. The housing market has also been used to gauge the health of the economy in a variety of studies I discuss (Ozdemir, 2020). Therefore, the question at hand for this study is “How do macroeconomic factors affect the price of homes in the United States”. To do this, I have created a state-level panel data model on the housing price index of each state with data from 1975 to 2021 built around the law of supply and demand.

In recent years, there has been a lot of uncertainty about the future economic health of countries around the world, mainly due to problems that were caused by the COVID-19 pandemic, political tensions, and conflicts between countries that have affected trade. COVID-19 caused a variety of issues for economies around the world from stock markets crashing, and high unemployment, to limited manufacturing and trade (Edwards et al., 2022). In the United States, to
dampen these effects of COVID-19, there was a release of stimulus money from the federal government and federal reserve. The FED started a massive campaign of quantitative easing and extremely low-interest rates to help kickstart the economy at the same time the government provided stimulus checks to the unemployed. This created an enormous increase in the money supply of around six trillion dollars (Cryan, 2022), and ultimately helped the economy to recover, the stock market to boom, and unemployment levels to decrease. Along with other factors it created an increase in the demand for homes around the United States at the same time not many homes were being built due to the pandemic. This resulted in a lot of demand that was met with limited supply and thus caused home prices to increase dramatically as seen in Figure 2 (Scorse et al., 2015). Although home prices were not the only thing that increased in price, we also saw record-high inflation all over the country. Which has now been met with aggressive interest rate hikes in recent months by the FED in its attempt to tame this inflation.

![Figure 2](image)

The higher interest rates are likely to slow the country's economic growth and affect the housing market and economy overall (Ozdemir, 2020). As there have been a lot of changes in
macroeconomic indicators and policy choices recently. I am curious as to how these changes have affected the housing market in the past to gain a better understanding of how they may affect the market this time around.

2. Literature Review

Over time, there have been countless studies on factors and policies that affect home and real estate prices in markets around the world. Studies cover a variety of different aspects that affect home prices. Take the study called “Early effects of COVID-19 pandemic-related state policies on housing market activity in the United States” (Yörük, 2022) for example. This study took daily and weekly data from 2020 on 100 metropolitan areas to investigate the effect of state-level policies on housing market activity. The study concluded that the closure of non-essential businesses in certain states was associated with a decrease in new home listings. This can be related to my study because when non-essential businesses closed it caused a lot of people to become unemployed and the unemployment rate is a key variable for this analysis. It shows that when the non-essential businesses were closed there was a decrease in the supply of new homes which could be associated with the increase in home prices experienced during the pandemic.

Another study I came across called “Impact of surf breaks on home prices in Santa Cruz, CA” looks at the effects of location and other area-specific characteristics that affect home prices in Santa Cruz California (Scorse et al., 2015). Although this isn’t highly related to my topic it shows the effects of supply and demand on home prices. The study concluded that homes right next to surf breaks with characteristics such as views of the beach and within proximity of the beach caused the prices of these homes to be on average $106,000 more than a home a mile away. Thus, emphasizing the idea that when there is a high demand for homes and there is a low supply it causes prices to increase. Which is directly related to my underlying theory, the law of supply
and demand. Another study that isn’t completely related to my topic but provides some insight into the functions of the housing market is a study called “Cointegration Analysis of Regional House Prices in the U.S.” (Zohrabyan et al., 2008). This study used quarterly data from 1975 - 2006 to investigate the dynamic relationship of home prices in the United States. To do this they split the United States into 9 regions and studies the effects of each region on each other to determine if some regions influenced the price of homes in other regions. The study found that regions that have higher levels of activity can influence nearby regions that have lower levels of economic activity. This can be related to my study because it helps to show the effects of economic activity on home prices.

There is a variety of studies that look at different aspects of the housing market from social factors, location, market trends, and policy decisions that affect various areas. Key insights can be drawn from them and implemented into other studies to create sound conclusions. For my model I have taken some of the results from the previous studies and implemented them into my study. As the studies above are not highly related to mine, I have chosen to base most of my model on studies that are focused on the effects macroeconomic variables have on the housing market. These studies have helped determine the variables I used and support the results produced from my model.

I have found a few complex studies that shed light on how macroeconomic factors affect home prices. In a paper written by Dicle Ozdemir called “Time-Varying Housing Market Fluctuations: Evidence from the U.S. Housing Market” (Ozdemir, 2020). The study focused on the effects of the housing market and credit markets on U.S. business and interest rate cycles. Although this study is more focused on the effects that the housing market has on the economic health of the U.S. it helps to support the idea that the housing market and the economy are very
highly related. It has also been very useful in supporting my decisions for specific variables I have added to my model. The study concluded that both the credit market and the housing market contribute to whether the economy moves to a low-growth regime or remains in a high-growth regime (Ozdemir, 2020). Indicating that the housing market plays a large role in the business and interest rate cycle for the United States. The study supported my decision to include interest rates in my study as well as the results of my model.

Similarly, another study conducted called “Macroeconomic Factors Affecting Housing Prices: Take the United States as an example” (Ding, 2022). This study used time-series data from 15 years on 191 observations samples in the United States to study the macroeconomic factors affecting housing prices in the United States. The study found that economic and stock growth cause the housing price index in the United States to increase which is like the results found by Dicle Ozdemir (Ding, 2022; Ozdemir, 2020). This study of macroeconomic variables on housing prices also found that increases in mortgage rates and unemployment had a negative impact on home prices (Ding, 2022). For which I have chosen to create unemployment and the federal funds effective rate into my model to see if I could create similar results. The study help to further support my results.

A study done by Leszczyński and Olszewski (2014) called “Panel analysis of home prices in primary and secondary market in 17 largest cities in Poland” is related to the other two studies I just mentioned and help to support mine. The panel analysis found that in Poland’s primary markets employment and economic growth increased home prices and which helps to support my hypothesis that unemployment will decrease home prices. It also found that in secondary markets, loan availability increased the prices of homes. This helped to support my decision to add interest
rates into my model because loan availability decreases when interest rates rise due to the loans becoming more expensive.

Some of the findings from these studies fit alongside the theory I am using in my model which is the theory of supply and demand which was the main determinate for my variable selection. Although I was not able to include GDP per capita as a measure of economic or income growth in my regression because it became a dominant variable in regression due to the high correlation with the housing price index. This high correlation can also support the findings of these studies that economic growth is highly related to the health of the housing market. To get around this issue I chose to add the minimum wages of each state as a way of measuring wage increases and economic growth.

The variety of studies I came across and chose to focus on when analyzing my model’s results help to support my findings. The studies also help to support my variable selection and the interpretation of the variables. Although a lot of the studies weren’t specifically on the effects of macroeconomics on the housing market, they all used similar variables and ideas to support their claims which I found to be very interesting and helpful.

3. Empirical Model and Estimation

This is a state-level panel data analysis of the Housing price index of each state in the United States from 1975 to 2021. The empirical model estimated for this research question is a mixed specification model with a combination of semi-logarithmic and double logarithmic specifications. The reason this equation takes two functional forms is that I have chosen to log my dependent housing price index variable creating the semi-logarithmic relationship with all but one of the independent variables. I had to create a log variable for the annual housing supply variable to
produce statistically significant results that supported the reviewed studies and underlying theory. This created a double-logarithmic relationship between the dependent variable and the lagged log annual housing supply variable. This is why my empirical model is a mixed specification model of semi-logarithmic and double-logarithmic specifications. This was the best choice for my model as it created the highest R-squared and resulted in having the most statistically significant variables when comparing it to other function forms such as semi-logarithmic on the dependent variable, double-logarithmic, and linear equations. This will not cause any issue with the results and only strengthen them; the only difference is the interpretation.

\[ \ln_{HPI_{it}} = \beta_0 + \beta_1 (URATE_{it} \times IR_{it}) + \beta_2 Post80s_{it} + \beta_3 SMW_{it} + \beta_4 LNLASH_{it} + \beta_5 FEDR_{it} + \varepsilon_{it} \]

Above is the empirical model, which includes a variety of manipulated variables to ensure the results produced support the literature and underlying theory. Starting with my dependent variable, I used a log form for the housing price index. I did this because there is a large variation between the smallest and largest values, as can be seen in the summary statistics table, Figure 7, in the data section of the paper. Due to the large variation in the HPI variable, it caused a lot of my variables to be insignificant and decreased the R-square. Making the HPI variable into a log variable caused the results of the regression to be overall statically better and increased the R-squared. As mentioned above I also created a lagged log variable for the annual housing supply variable, lnLash. I first turned tried to lag the variable because when I used the original variable it created a positive coefficient which is not supported by the underlying theory. This solved the problem and made the variable coefficient negative, but it made the coefficient statistically insignificant. I then created a logged variable to see if that had any effect and it did. It made the
coefficient statistically significant and positive like when I used the original variable. So, I then decided to lag the logged version of the ASH variable and it created a statistically significant negative coefficient which is what I was looking for because it is supported by the theory and reviewed studies. The only difference that has been caused by doing this is that I created a double-logarithmic relationship between the ASH variable and the dependent variable. This affects the interpretation of the results so the way I will be interpreting the coefficient in the results section will be that a one percent change in the annual housing supply will cause the housing price index to have a percent change of whatever that coefficient is in the results. If the coefficient produced for the annual housing supply is negative 0.06 then a one percent increase in the annual housing supply will cause a 0.06 percent decrease in the housing price index. As for my other independent variables they will be interpreted as for every one-unit change in the non-logged variables will cause a percent change in the housing price index determined by each of those variable’s coefficients. Then there was no need for manipulation of the state minimum wage, federal funds effective rate, and post 80 variables throughout the rest of the analysis as they produced significant results that corresponded with the theory, literature, and hypotheses. As for the inflation rate and unemployment rate variables they needed to be manipulated into an interacted variable. An interacted variable tries to represent the interaction effects of two variables into one independent variable. I have chosen to do this with my inflation and unemployment rate variables because according to a lot of literature, the two variables have an inverse relationship (Team, 2023). This inverse relationship is called a Philips curve, Figure 3, and it explains that in periods of high inflation when the economy is growing there is low rates of unemployment. Conversely stating that in periods of high unemployment when the economy is tightening there is going to be lower inflation due to the slowdown in the economy (Team, 2023).
Above in Figure 4 is my Hausman specification test which is used to determine whether I should use a fixed effect or random effect model. The test indicated that I could use either a fixed effect model or a random effect model as the test did not reject the null hypothesis. Stating that the model is efficient using variables that are fixed or random. To determine if I will be using a fixed or random effect regression, I compared the results of the two models.
Figures 5 and 6 are the results from my initial regressions using a fixed effect model in Figure 5 and a random effect model in Figure 6. Both regressions show promising results that would be acceptable for a model because in both models all the variables are statistically significant with p-values of 1%. As well all the variable coefficients indicate the expected results the only difference between the two models is the R-squared. The R-square is the measure of how much variance in the dependent variable can be explained by the dependent variables. The R-
squared for the fixed effect model was .89 indicating that 89% of the variance in the dependent variables can be explained by independent variables. As for the random effect model, the overall R-squared was .82 or 82%, this is less than the R-squared in my fixed effect model so I will be continuing by using a fixed effect model.

5. Data
Throughout my analysis, I have tested a variety of different variables to use in my final analysis which brought my initial independent variable count from 15 to 6. Overall improving the regression model and clearing out some of the clutter. I knew I was going to have to eliminate some of my variables due to them being extremely correlated which can cause problems for the model results. The reason I started with more variables was to enable me to run multiple regressions to determine the best combination of variables. One specific variable I had to remove which went against some of my studies was my gross domestic product per capita variable which was a state-level independent variable. I was anticipating this variable to be correlated with the housing price index variable, but it was too highly correlated with the dependent variable. Which caused it to become a dominant variable that skewed the results of my regression for which I had to remove it entirely. This makes a lot of sense according to studies I have reviewed which were primarily focused on income and economic growth as the main indicators for their models.

The housing price index is the dependent variable that measures the movement of single-family home prices in the United States. It is a broad inflationary adjusted variable with a base year of 1980 that is set equal to 100. It does not represent a specific monetary value but helps to gauge changes in home prices over time. In Figure 7 you can see that I included both the normal form of the HPI variable as well as the logged form. I did this to show the significant change in
variation between values for the variable which ultimately helped create accurate results as explained previously. As you can see the original HPI variable had a minimum value of 43.627 and a maximum value of 956.93. This is a large difference in variation that can skew results thus why I logged the variable. Turning the variable into a log variable significantly decreased this variation and caused the minimum value to drop to 3.776 and the maximum value to drop to 6.864. It also caused the standard deviation to decrease from 138.829 to 0.588 and the mean to decrease from 236.3 to 5.3. The variable also has 2350 observations indicating it is a large data set. Larger data sets tend to produce better results as there are more values for the model to work from.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. Dev</th>
<th>Min</th>
<th>Max</th>
<th>Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPI</td>
<td>236.322</td>
<td>138.829</td>
<td>43.627</td>
<td>956.93</td>
<td>N = 2350</td>
</tr>
<tr>
<td>ln_hpi</td>
<td>5.298</td>
<td>0.588</td>
<td>3.776</td>
<td>6.864</td>
<td>N = 2350</td>
</tr>
<tr>
<td>lnl_ash</td>
<td>1.771</td>
<td>0.248</td>
<td>1.361</td>
<td>2.370</td>
<td>N = 2300</td>
</tr>
<tr>
<td>post80</td>
<td>0.894</td>
<td>0.308</td>
<td>0</td>
<td>1</td>
<td>N = 2350</td>
</tr>
<tr>
<td>FEDR</td>
<td>4.740</td>
<td>3.916</td>
<td>.08</td>
<td>16.390</td>
<td>N = 2350</td>
</tr>
<tr>
<td>IR</td>
<td>3.728</td>
<td>2.760</td>
<td>-0.356</td>
<td>13.549</td>
<td>N = 2350</td>
</tr>
<tr>
<td>Urate</td>
<td>5.877</td>
<td>2.076</td>
<td>2.1</td>
<td>17.200</td>
<td>N = 2301</td>
</tr>
<tr>
<td>SMW</td>
<td>4.986</td>
<td>2.359</td>
<td>.75</td>
<td>13.690</td>
<td>N = 2345</td>
</tr>
</tbody>
</table>

Figure 7: Summary Statistics

My state minimum wage variable, SMW, has 2345 observations. This is only 5 less than most of the other variables, but this is because Virginia adopted a state minimum wage in 1976 and Kansas adopted a minimum wage in 1979. Some states still do not have a minimum wage law, so I had to manipulate my state minimum wage variable for some of the states. For those specific states, I inputted the federal minimum wage to ensure there wasn’t a lot of missing data. The states that do not have minimum wage laws were Alabama, Louisiana, Mississippi, South Carolina, and
Tennessee for which I implemented that federal minimum wage for those states. For the state
minimum wage variable, there is a minimum value of .75 or 75 cents and a maximum value of
13.69 or $13.69. The variation between the max and min is due to wages increasing over time, but
there is still a wide degree of variation between states as shown in Figure 7. The variable also has
a mean of $4.99 and a standard deviation of $2.36. This means that for a large portion of the
minimum wage values are $2.36 more or less than the mean of $4.99.

As for the post 1980 variable I used, it is a dummy variable that shows the impact of prices
after 1980. I did this because the dependent variable, HPI, has the base year of 1980 but retrieved
data beginning in 1975. This variable is particularly helpful because of the way the dependent
variable has been set up. As well if you look back at Figure 1 in the introduction portion of the
paper you can gain a real understanding of how much variation there was in the index after 1980.
This variable has 2350 observations as it is based on the number of years studied and is
implemented for every state. The variable’s minimum and maximum values are 0 and 1. If the
value is 1 it indicates that the variable value is after 1980 and if it is 0 it is before 1980. The
important statistic to look at for this variable is the mean because it shows how many of the values
are after 1980. Which is 89.4%, this is a significant amount and helps explains why the effect of
this variable is so significant in the final regression results.

The next variable I am going to go over is the lagged log annual housing supply variable,
LNL_ASH. This variable is the ratio of new houses for sale to new houses sold in the United States
annually. Due to the variable being logged, it has decreased the variation in the minimum and
maximum values as well as the other statistics. The minimum value is 1.361 and the maximum is
2.37. This indicates that the shortest amount of time for the new home inventory to be depleted is
1.361 years and the longest time it would take to deplete the new home supply would be 2.37 years
if no homes were built over the studied period. This variable helps to gauge the demand for new homes and the effects of periods when fewer homes are being built on the housing price index. The variable has a mean of 1.771 and a standard deviation of .248. This indicates that on average it would take 1.771 years to deplete the new home supply if no new homes are built with a standard variation of 24.8% percent of a year from that average. As shown in Figure 7 one can also see that there are only 2300 observations for this variable because this is a lagged variable. When one decides to lag a variable, it uses the value from the previous year to estimate its effect on the value of the dependent value for the current year which then excludes the most recent year’s value from its observation. Being that my data is based on 50 states it eliminates the variable’s value for all 50 states for the year 2021.

The next variable I am going to go over from the summary statistics is the federal funds effective rate or FEDR. This variable is typically measured monthly, but since this study is based on changes per year it has been changed to represent the average rate per year. The federal funds effective rate is the interest rate at which depository institutions lend funds to each other overnight to meet reserve requirements. It is a monetary tool used by the federal reserve to influence borrowing costs that can either restrain or stimulate economic activity. This rate effectively impacts the interest rates individuals receive when they take out loans for homes and other goods. Figure 7 shows us that there are 2350 observations indicating there are no missing values. The minimum value or lowest rate seen over the studied period is .08% and the highest rate represented by the maximum value in the summary statistics is 16.39%. The mean rate for the period was 4.474% with a majority of the values being between 3.916% below or above that average rate as seen in the standard deviation. This is a large standard deviation in comparison to the mean indicating there is a significant variation in rates over time.
For the inflation rate variable, IR, we can see that number of observations is 2350 in the summary statistics indicating this variable has no missing values. This variable measures the annual percent change in the cost of goods and services over time for the United States. The summary statistics indicate the largest change in prices over the period was 13.549% shown by the max. The minimum value indicates the largest decrease in prices over the period which was .356%. It is expected for this decrease in prices to be very low because it is very rare for the United States to experience the deflation of prices. The average change in prices over a year is 3.728% which is represented by the mean. The standard deviation from the mean is 2.760% indicating that over the period the annual inflation rate ranged from .968 to 6.488.

The unemployment rate variable, Urate, indicates the rate of unemployment per year for each state in the US. In the summary statistics, Figure 7, you may notice that there are 2301 observations. This is because Connecticut was the only state that recorded its unemployment rate in 1975 and the other 49 states did not measure unemployment until 1976 which is why there are 49 fewer values than some of the other variables. The highest level of unemployment represented by the maximum value in the summary statistics was 17.2% and the lowest recorded unemployment rate was 2.1% as represented by the minimum value. As for the mean, there is an average unemployment rate of 5.877% between states with a standard deviation of 2.076%. This means that on average for every state in the US the unemployment rate tends to be from 3.801% to 7.953%.
Figure 8: Correlation Matrix

Above, in Figure 8, is my correlation matrix which is used to determine the level of correlation between variables and is a helpful tool for detecting multicollinearity. Correlation matrices produce values between -1 and 1. The closer are the correlation coefficients to 1 or -1, the stronger is the correlation between the variables. If the value is closer to 0, they are less correlated to each other. The model indicates how significant the relationships are and all the relationships in my correlation are significant at the 10% level indicated by the asterisk located to the right of each value. Typically, one is looking for the relationships between variables to have a value between .7 and -.7 because when variables are highly correlated it can create poor or inaccurate results in your regression. When a variable is above .9 or below -.9 it indicates there is a very strong correlation between the variables and should be analyzed closely to make sure it does not cause problems in the regression.

I was glad to see that all but one of the variables I have chosen to use in my final regression was not close to the .9 or -.9 levels. The one highly correlated relationship is the relationship between the logged housing price index and the state minimum wage variables with a value of 0.871. This is expected as both variables increase consistently over time as shown in Figure 9 below and Figures 20 and 21 which are in the results portion of the paper. The graph below looks
at the relationship between state minimum wage, the red line, and the housing price index, the blue line. Although this correlation is high it is expectable considering they increase over time.

![Figure 9: Red = SMW, Blue = ln_hpi](image1)

![Figure 10: Blue = FEDR, Red = IR](image2)

![Figure 11: Blue = FEDR, Red = SMW](image3)

There are two other variable relationships that have a correlation above .7 and below -.7. In Figure 10 we can examine the highly correlated relationship between our inflation rate and federal funds effective rate variables. In Figure 8 we can see that the relationship resulted in a value of .747. This is high but isn’t high enough where one should remove one of the variables as they are important to the regression. These variables being highly correlated makes sense as well because in periods of high inflation the federal reserve tends to increase the federal funds effective rate to combat the inflation. Similarly, when inflation is low, the federal reserve tends to decrease the federal funds rate to incentivize lending and borrowing to spark economic activity by offering lower interest rates.

As for the relationship between the federal funds rate and state minimum wage being negatively correlated with a value of -.739 in the correlation matrix, Figure 8. We can examine
this relationship by using the graph in Figure 11. Although these variables are measured in different ways there seems to be a relationship between the two. This can be explained by the fact that in the early 80s, there was high inflation that caused the federal reserve to increase interest rates significantly, but over recent decades the federal funds effective rate has been significantly lower since then causing the FEDR variable to look like it is decreasing over time. As for the SMW variable, it increases over time because states increase the minimum wage to support individuals as prices of goods increase and they will never decrease the minimum wage. This explains the correlation between the two variables, but because they are not related in a way that will affect the regression there is no need to make any further changes to these variables.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Description</th>
<th>Hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPI</td>
<td>The real housing price index per state (inflation-adjusted base year 1980 = 100)</td>
<td>Dependent variable</td>
</tr>
<tr>
<td>post80</td>
<td>A dummy variable that indicates the effect on HPI after 1980</td>
<td>+</td>
</tr>
<tr>
<td>Urate</td>
<td>The Unemployment rate of each state per year</td>
<td>-</td>
</tr>
<tr>
<td>IR</td>
<td>The inflation rate of the US per year</td>
<td>+</td>
</tr>
<tr>
<td>SMW</td>
<td>The state minimum wage per state; states with no minimum wage laws have the federal minimum wage implemented</td>
<td>+</td>
</tr>
<tr>
<td>ASH</td>
<td>The annual supply of new homes in the US per year</td>
<td>-</td>
</tr>
<tr>
<td>FEDR</td>
<td>The Federal funds effective rate</td>
<td>-</td>
</tr>
</tbody>
</table>

Figure 12

Above is a table of my hypothesis for the results of my model, which are based on the underlying theory and relevant studies I have reviewed. For the underlying theory of my project, I have chosen to use the law of supply and demand as it is the best for explaining my results and variables. Although another framework that I am using to build my model is the asset pricing
theory. As this has been the framework for well-recognized models such as Asset Pricing Model (CAPM), the Arbitrage Pricing Theory (APT), and the Fama-French three-factor model. These models are based on the underlying theory of supply and demand but rely on the idea that other variables affect asset prices such as interest rates and risk. As for my variable hypotheses, I believe that the unemployment rate variable would result in a negative coefficient. This can be explained by some of the studies mentioned and by the law of supply and demand because when unemployment is high, and people are losing jobs it is an indication that the economy is tightening, and people are less likely to be demanding homes. When people aren’t buying up homes it will cause the price of homes to decrease thus why I hypothesized the unemployment coefficient to have a negative effect on the dependent variable, HPI. As for the inflation rate variable, I feel it is relatively self-explanatory why I think this will increase the price of homes. This is because the inflation rate is a measure of increases in the prices of goods in all sectors around the country and when inflation is increasing so will the price of homes. These two variables play an important part in my model because they create an inverse relationship called the Philips curve.

As for my hypothesis for the effect of state minimum wages on housing prices, I think there will be a positive coefficient for its effect on the HPI variable because as there are increases in income for those that rely on minimum wage to survive it may result in them being able to move into a home instead of renting or even moving into a new home thus contributing to increases in home prices. Although this variable could be viewed as being insignificant for a model like this because one may think minimum wage-earning individuals may be less financially stable or are younger adults that aren’t likely to buy homes thus not contributing to price increases. It is a fair point but, I think that adding this variable helps to improve my model by representing increasing wages over time and wage differences between states. Take New York and California for example
which have extremely high home prices and higher minimum wage laws. Whereas places like Alabama and South Carolina that rely on the federal minimum wage law and have much lower home prices. Thus, why I think this a good variable to add as I was unable to add GDP per capita because it was a dominant variable. It also supports the idea that when people have more money, they are likely to spend more which increases prices and in this case home prices.

I have hypothesized that my annual supply of new homes variable will negatively affect the housing price index. This hypothesis is rooted in my underlying theory which states that as the supply of something increases it will cause a decrease in prices. We saw the opposite of this happen during the pandemic when there was a large demand for homes and the new housing inventory was low which is associated with the increase in prices, this can be seen in Figure 2 above. Lastly, my hypothesis for how the federal funds effective rate will affect the housing price index, I think it will be negative. This is backed by the studies I have reviewed as interest rates have been a key part of the studies. This is also supported by the theory of supply and demand. The idea is that when the interest rates are higher or increasing, like at the beginning of 2023, it makes investments less attractive due to the higher monthly payments they must pay due to the higher interest rates. Thus, it will lessen the demand for homes and other goods as people won’t want to accrue that additional interest expense.

5. Empirical Results
On Figure 13 are the results from my empirical model. After the analysis of my final regression results, I go into the robustness checks and tests needed to create these results. The equation uses robust standard errors as explained below. Due to the format created when using the asdoc command in Stata, it eliminates two important statistics created by the regression, so I will begin
my result analysis by explaining them. Starting with the rho statistic which indicates the positive persistence in the error terms across time within each entity. For my regression the rho value was .4429 or 44.3% this suggests a moderate positive correlation in their error terms of the observations. The rho of .4429 is relatively low for a model which is a good thing that helps indicate the model is specified correctly and there aren’t omitted variables. Typically, a model with a rho of .7 or below is acceptable making this model very good for determining the effects on the housing price index. The other statistic I wanted to mention that is not included in the image above is the corr(u_i, Xb) which measures the correlation for the whole model. The correlation indicated by this metric for my model is 0.0353 indicating there is not a lot of correlation between variables that may skew results. This is good because when variables are too highly correlated with each other they can cause inaccurate results.

![Regression results table](image)

Figure 13

Now diving into the results shown above in Figure 13 starting with the R-squared. The R-square of the model indicates the total variation in the dependent variable that can be explained by the independent variables. My model produced a R-squared of 0.89 indicating that the independent
variables explain 89% of the variance in the dependent variable. The other 11% of the variation is explained by the error term. The R-squared for this model indicates the model has a strong and effective fit for the data used. This is a key statistic when using regressions which is why I aimed to keep my R-squared high when choosing and manipulating variables throughout my analysis. It is also important to note that when the R-squared is very close to 1 it could indicate there are variables in the equation that are too highly correlated with the dependent variable and could affect the accuracy of the model. Being that my R-squared is less than .95 helps to support the idea that this model is accurate with correct specifications.

I will now begin to break down the results of each variable shown in Figure 13 starting with the interacted unemployment and inflation rate variable, c. The variable has produced a p-value of 0.007 indicating it is statistically significant at the 10%, 5%, and 1% levels. Which means it does influence home prices, the logged housing price index. In the regression results, you can see that the effect of each independent version of the variable is displayed. Both of which are statistically significant to the interpretation. The unemployment rate portion of the interacted variable causes the logged housing price index to decrease by .025% with a robust standard error of .006%. This means that every time the unemployment rate increases by one the logged homing price index decreases by 0.25%. The inflation rate variable on its own causes the logged housing price index to increase by .044% with a robust standard error of .011. This means that every time the inflation rate increases by 1 unit the logged housing price index increases by .044%. Then when examining the interacted variable’s coefficient, the results indicate that when there is a 1 unit increase in the relationship between inflation and unemployment it causes a .005% decrease in the logged housing price variable with a robust standard error of .002. This may not be what one may expect because as explained above inflation causes a larger effect on the HPI than the
unemployment rate independently. The reason the interaction between these variable coefficients causes a negative effect instead of a positive effect is because if you look at the summary statistics in Figure 7 you can see that the unemployment rate tends to be higher than the inflation rate. This causes the regression to estimate that the unemployment rate has a larger effect on the housing price index when it is interacted with the inflation rate but is smaller for a one-unit increase. The individual results of these variables support my underlying theory, reviewed studies, and hypothesis.

The next variable I am going over is the post 80 dummy variable. As explained earlier this variable is used to show the effects on home prices after 1980 which is primarily based on how the housing price index variable is set up. The variable has a P-value of 0 indicating it is statistically significant at the 10%, 5%, and 1% levels emphasizing its importance to the model. The variable also produced a positive coefficient which lines up with my hypothesis. The post80 variable causes a .546% increase in the logged housing price variable with a standard error of .029. This means that after 1980 there was a .546% increase in the logged housing price index in comparison to the index from 1975 to 1980. This aligns with my hypothesis and contributes to the overall accuracy of the model and the estimated coefficients.

I had similar results with my state minimum wage variable, SMW. The regression created a P-value equal to 0 indicating it is statistically significant at the 10%, 5%, and 1% levels. The coefficient created by the model is .135 which means there is a significant effect caused by state minimum wages. It can be interpreted as for every $1 increase in the state minimum wage it causes a .135% increase in the logged housing price index. The robust standard error for this coefficient is .006 which means the .135% effect on the HPI can vary from .129% to .141%. These results
align with my hypothesis and theory. It also aligns with the reviewed studies because I am using the state minimum wage to help estimate the effect of wage increases on home prices.

The results from the lagged log annual housing supply variable also align with my hypothesis and the underlying theory for the model. The coefficient for this variable has a P-value of .003 which means it is statically significant at all levels and is important to ensuring the accuracy of the model. As well the coefficient has a value of negative .057 with a robust standard error of .018. This means that when the annual housing supply increases by 1% in the previous year it causes the logged housing price index to decrease by .057% which is a variation of .018%. These results are directly correlated with the theory of supply and demand because when there is a large supply of new homes inventory it causes a decrease in home prices because it decreases the demand for new homes thus allowing prices to decrease.

The last variable in my regression is the federal funds effective rate which has a P-value of 0. This indicates the variable is statistically significant at all levels and is important that it is included in the model. It also creates a negative effect on the logged housing price index which supports the reviewed studies. The coefficient for the FEDR variable is negative .037 with a robust standard error of .004. This means that every time the federal funds effect rate increases by 1 unit, percent, it creates a drop in the logged housing price index of .037% with a variation of .004%.

The last statistic in the model results I will go over is the constant term or the intercept. The constant term has a P-value of 0 indicating it is statistically significant at all levels and is accurate. The coefficient for the constant term is 4.509. This means that when all the independent variables are equal to 0 the logged HPI variable is equal to 4.509. To retrieve these results and ensure they are accurate there is a handful of robustness checks that need to be tested for. Below are the steps I took to ensure the accuracy of my model.
After performing my Hausman test and determining which type of model to carry on with I then went on to check for multicollinearity. Multicollinearity is when two or more variables in a regression are highly correlated with another variable in the equation and is a violation of the classical assumptions used for regressions. One can detect highly correlated variables using the correlation matrix above as explained earlier but it is important to run a variance inflation factor test, Figure 14. To use this test, you cannot run the panel data regression function xtreg in Stata first, you must use the function reg for normal regressions and then use the VIF command. The test produced a mean VIF of 6.634 which isn’t amazing because it is above 5 but is below 10 so it is acceptable. The reason for the correlation present in the test is mainly because of the high correlation between the inflation rate and the interacted variable unemployment and inflation which is expected because the interacted variable is based upon the inflation variable. The test shows a VIF of 15.266 for inflation and 16.192 for the interacted variable. Other than the significantly high correlation between those two variables one can notice that there is another variable that has a VIF above 5 which is the federal funds effective rate with a value of 5.335. This is also expected and is not an issue for the model, but it is worth pointing out. The value is high because interest rates are shown to be highly correlated with home prices as mentioned in the

<table>
<thead>
<tr>
<th>Variance inflation factor</th>
<th>VIF</th>
<th>1/VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urate</td>
<td>3.564</td>
<td>.281</td>
</tr>
<tr>
<td>IR</td>
<td>15.266</td>
<td>.066</td>
</tr>
<tr>
<td>c.Urate#c.IR</td>
<td>16.192</td>
<td>.062</td>
</tr>
<tr>
<td>post80</td>
<td>1.781</td>
<td>.561</td>
</tr>
<tr>
<td>SMW</td>
<td>2.894</td>
<td>.346</td>
</tr>
<tr>
<td>lnl ash</td>
<td>1.406</td>
<td>.711</td>
</tr>
<tr>
<td>FEDR</td>
<td>5.335</td>
<td>.187</td>
</tr>
<tr>
<td>Mean VIF</td>
<td>6.634</td>
<td></td>
</tr>
</tbody>
</table>
studies and relates to the inflation rate as discussed when going over the correlation matrix. Although a variance inflation factor test with a mean VIF above 5 isn’t perfect it is much lower than some of the other VIF results from previous regressions I ran when trying to determine which model was best to use. There is expected to be a higher correlation between variables for a model like this compared to models on different topics. The reason for that is that macroeconomic indicators move in similar directions, have time trends, and are related because they are measures of the economy thus creating correlation.

![Figure 15](image)

**Figure 15**

![Figure 16](image)

**Figure 16**

The next step in my model analysis was testing for serial and heteroskedasticity. Figure 15 is a Wooldridge test for autocorrelation in panel data, the test used to determine if a regression has serial correlation present. When serial correlation is present it indicates that there is a relationship between a variable in the regression and a lagged version of the same variable indicating it isn’t random. Serial correlation violates classical assumptions and must be situated when it is present. From the results of the Wooldridge test, we can confirm that there is serial correlation in my regression because it produces a Prob > F equal to 0. Thus, rejecting the null hypothesis that there is no serial correlation. This was expected because the variables in my equation are not randomly
created and are based upon previous changes in the economy. Before solving this issue, it is important to test for heteroskedasticity because if both are present the simple way to solve the problem is by changing the standard errors in the regression into robust standard errors. Heteroskedasticity is present when the predicted variable standard deviations are non-constant. This is a violation of the classical assumptions used for regression models. To test for heteroskedasticity in my regression I used the Modified Wald Test for groupwise heteroskedasticity in the fixed effect regression model shown in Figure 16. The test indicated that heteroskedasticity is present as it rejects the null hypothesis with a prob > chi value of 0. This indicates that I must change my regression to use robust standard errors verse using standard errors. Robust standard errors are used to solve the bias caused by heteroskedasticity and autocorrelation. In Stata the robust standard errors cause the ordinary least squares to leave u(i) in the error term alleviating the issues of serial correlation and heteroskedasticity. Below in Figure 17 shows the results of the regression after switching from standard errors to robust standard errors. Changing the standard errors to robust standard errors did not affect the model negatively at all, it only made the P-value of the interacted inflation and unemployment rate variable go from 0 to 0.007 but it is still statistically significant at the 1% level.

![Regression results](image)

Figure 17
After testing and solving the issues of serial correlation and heteroskedasticity I moved on to testing for non-stationarity and cointegration. Non-stationarity is when there is a time trend in a variable used in the regression which can cause the regression results to be unreliable and can lead to poor conclusions. On the other hand, cointegration is helpful when you have non-stationary variables in your empirical model because it indicates there is an equilibrium in the time trend that allows the regression to produce accurate and reliable results. To test for non-stationarity, I used

**Figure 18**

<table>
<thead>
<tr>
<th>Statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inverse chi-squared(100) P</td>
<td>33.7628 1.0000</td>
</tr>
<tr>
<td>Inverse normal Z</td>
<td>6.0214 1.0000</td>
</tr>
<tr>
<td>Inverse logit t(254) L*</td>
<td>5.8042 1.0000</td>
</tr>
<tr>
<td>Modified inv. chi-squared Pm</td>
<td>-4.6837 1.0000</td>
</tr>
</tbody>
</table>

P statistic requires number of panels to be finite. Other statistics are suitable for finite or infinite number of panels.

**Figure 19**

<table>
<thead>
<tr>
<th>Statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modified Dickey-Fuller t</td>
<td>-20.9792 0.0000</td>
</tr>
<tr>
<td>Dickey-Fuller t</td>
<td>-8.0319 0.0000</td>
</tr>
<tr>
<td>Augmented Dickey-Fuller t</td>
<td>-7.5328 0.0000</td>
</tr>
<tr>
<td>Unadjusted modified Dickey-Fuller t</td>
<td>-10.7805 0.0000</td>
</tr>
<tr>
<td>Unadjusted Dickey-Fuller t</td>
<td>-6.8016 0.0000</td>
</tr>
</tbody>
</table>

After testing and solving the issues of serial correlation and heteroskedasticity I moved on to testing for non-stationarity and cointegration. Non-stationarity is when there is a time trend in a variable used in the regression which can cause the regression results to be unreliable and can lead to poor conclusions. On the other hand, cointegration is helpful when you have non-stationary variables in your empirical model because it indicates there is an equilibrium in the time trend that allows the regression to produce accurate and reliable results. To test for non-stationarity, I used
the Fisher-Type unit-root test on my dependent variable which determines if there is a unit root in the variable. If there is a unit root, then non-stationarity is present which is indicated by the test in Figure 18. Being that all four tests have a p-value of 1 indicates there is a unit root in the panel, and I cannot reject the null hypothesis. In Figures 20 and 21 you can also see the presence of non-stationarity as there is a time trend in both the housing price index and state minimum wage variables between states.

To elevate the issue of non-stationarity I tested for cointegration to see if there is some type of equilibrium in the time trend. If cointegration is not present, you must manipulate your variables into difference variables to eliminate the time trend to produce accurate results. To test for cointegration in my regression I used the Kao test for cointegration which runs 4 different variations of the Dicky-Fuller test for panel data, Figure 19. The results of this test on all the variables used in my regression indicate there is cointegration. This is because the results produce a p-value of 0 which caused me to reject the null hypothesis that there is no cointegration. Due to cointegration being present alongside non-stationarity, I do not have to manipulate the regression further to produce the final regression results talked about above.

Figure 20  

Figure 21
6. Discussion and Closing Remarks

In the introduction of this paper, I introduced the research question which was “How do macroeconomic factors affect the housing price index: a state-level panel data analysis from 1975 – 2021”. I feel this model does a good job of showing the effects of macroeconomic variables on the housing price index for each state in the US. As the model has an overall R-squared of .89, and a rho of .44, and all my variables are statically significant with coefficients that are supported by the underlying theory and related studies on this topic. To get to this point I had to select variables based on other studies and the underlying theory I chose for this study. I then had to specify my empirical equation to create a model that fits the underlying theory and produces accurate results. After I had to run multiple tests to ensure the results are reliable. For some models running these tests and manipulating variables can cause results to become insignificant or go against their theories but for this model that did not happen. Thus, supporting my claim that this model sheds light on the research question at hand.

The model indicated that unemployment affects the price of homes negatively which is supported by the results of a study called “Macroeconomic factors affecting housing prices: Take the United States as an example” (Ding, 2022). This can also be supported by the underlying theory of supply and demand because as more individuals become unemployed, they tend to try and save more money than they spend thus decreasing the demand for homes. The model also indicated that the inflation rate has a positive effect on home prices. This is because inflation is a measure of price increases year after year. Homes are goods, so when inflation increases home prices will follow. The main point of the variable was to see to what degree it affected home prices. In this model, I chose to interact inflation with unemployment due to their inverse relationship, and it
created some interesting results. The model indicated that unemployment has a larger effect on the housing price index than inflation.

The model also indicated that state minimum wage increases the housing price index as well. I have decided to interpret the state minimum wage variable as a measure of wage increases over time which can be associated with economic growth. The effects of minimum wage on the housing price index can be supported by both the law of supply and demand as well as the reviewed studies. It is supported by the theory because as people have more money, they are likely to demand more goods or more expensive goods thus causing an increase in prices, in this case for homes. It relates to two studies I have reviewed earlier as well; “Macroeconomic factors affecting housing prices: Take the United States as an example” (Ding, 2022) and “Panel Analysis of Home Prices in the Primary and Secondary Market in 17 Largest Cities in Poland” (Leszczyński and Olszewski, 2014). As both studies concluded that prices of homes are directly associated with economic growth and because I have chosen to use this variable as a measure of wage growth that is associated with economic growth it is supported by these studies.

The negative effects of increasing the federal funds effective rate on home prices is supported by two of the studies I reviewed earlier. It is also supported by the framework for asset pricing I mentioned earlier. The asset pricing framework was used to create a variety of pricing models like the CAPM and suggests that interest rates and risk play a part in the pricing of assets alongside supply and demand. The framework suggests that interest rates cause the demand for assets to decrease which supports the results created by this model. Similarly, the studies “Macroeconomic factors affecting housing prices: Take the United States as an example” (Ding, 2022) study and the “Time-Varying Housing Market Fluctuations: Evidence from the U.S. Housing Market” (Ozdemir, 2020) both concluded that interest rates and mortgage rates have
negative effects on home prices. This makes supports my study because as interest rates rise people have higher interest payments on loans used to purchase homes thus decreasing the demand for homes.

Lastly, this model has concluded that increases in the annual supply of new homes have a negative effect on home prices. This is deeply rooted in the underlying theory used for this project because the law of supply and demand states that when the supply of something is increasing it causes downward price movements for that good. We can see this in our model because of the coefficient created is negatively associated with the housing price index.

In conclusion, this is an accurate empirical model that demonstrates the effects of macroeconomic variables on home prices. This model can be used to gauge the direction of the overall housing market in the United States. Which is helpful when trying to gain an understanding of the whole US economy. In the conclusion of a study on time-varying housing market fluctuations, the housing market can be used to determine the direction of the economy (Ozdemir, 2020). Although this study will not tell you what causes specific price changes between different areas around the country it will give you an idea of the direction of the housing market in the United States.
References


