**Analyzing the Real Estate Market: Are United States’ Home Values Impacted by Changes in Price Levels?**

By Genna Barge

***Abstract—***

This analysis utilizes panel data for the 50 U.S. States and Washington D.C. over two six-month periods, January 2021-July 2021, and July 2021-January 2022. The theory under examination throughout this paper is whether there is a relationship between median home values across the United States as it relates to price level, or merely inflation. Results from a quadratic and interaction model estimated using Ordinary Least Squares techniques suggest that the marginal effect of price level on home values is dependent on vacancy levels. An increase in the number of vacancies when the rate is low leads to an increase in terms of home value, but in states with higher existing vacancy rates, an increase in the vacancy rate will cause a decrease in terms of the home value. Other control variables include the unemployment rate, real GDP per capita, the vacancy rate, and population change.

Periods of persistently high inflation can have adverse implications for consumers and their consumption, particularly as it is related to inelastic goods and the proportional shift in both their supply and demand as price changes. As has been concluded in previous research, a small change in price level results in a significant shift for the demand of houses (Shwab 1982). While many researchers have analyzed this impact over both the long and short run, I strive to study how price level influences home value over a period of two six-month intervals. Stemming from my curiosity of real estate, as well as my concern for persistent inflation, this paper investigates whether an increase in prices influences home values, utilizing real GDP per capita, the unemployment rate, population growth, and vacancy rates as control variables.

Growing price levels and inflation could impact consumers’ decisions, and subsequently, the demand for houses: higher inflation may lead consumers to spend a higher proportion of their disposable income on necessities and inelastic goods, thus lowering the demand for houses. In context with recent events, the COVID-19 pandemic may have altered home valuations. I hypothesize that this global event sparked periods of tight market control, as well as periods of heightened inflation, which led consumers to stay at home and spend their disposable income on home improvement projects, leading to increased construction and heightened home values. Thus, I am curious to study the longevity of these projects on home values, as I hypothesize that values will start to converge toward their long run mean in this period. My initial prediction follows closely in the conclusions of Shwab (1982) in that I predict that an increase in inflation raises home values in the short run, but ultimately, they fall over time.

1. **Literature Review**

Research has been conducted to investigate the specific determinants of fluctuations in housing prices and values, however, no definitive conclusion has been reached due to the presence of varying independent and control variables. Shwab (1982) uncovers many critical findings in his analysis of home values in the United States as they are related to the expected rate of inflation. By developing a utility function to estimate the impact of income and inflation on consumption, he ultimately computed the derivative for housing demand as it relates to future inflation to study consumer response to price level increases. Shwab concluded that a relationship does exist between price levels, future expectations, and housing demand and valuation: consumers must adjust their consumption and housing plans so that they satisfy both their budget and borrowing constraints. Furthermore, Shwab determined that an increase in the rate of inflation raises the real cost of housing in the short run, but ultimately results in a decrease of home values in the long run due to increased vacancy rates and outside policy. In addition, Shwab determined that the elasticity value of the price level as it is related to home values is almost exactly unit elastic at -0.966, meaning that a small change in overall price level will result in a large change for the demand of houses, however, this demand will fall over time.

Similarly, Head, Llyod-Ellis, and Sun (2014) note that home values and market prices are more volatile than is predicted by the standard asset pricing model. This is ultimately a result of shocks in both inflation, as well as city-specific income changes. In their analysis, they estimated a vector autoregressive model that utilizes city-level data. Their findings were of similar stature to Shwab in that they discovered in the short run, as price level increases, home values rise as a response. Similarly, they concluded that in the long run, these values will fall to around their mean. Their analysis also concluded that there is a negative relationship between vacancies and home values. Furthermore, they argue that periods of tight market control are typically mitigated by a reduction in vacancies, which is due to housing sector reallocation. Overall, their analysis lends me to include vacancy rates as a control variable in my model.

Valadez (2011) contributes similar conclusions in his research, ultimately providing that as price levels rise, regulation of the repacking of debt, as well as the prominence of a lack of transparency in the mortgage industry escalates home values. Additionally, Valadez analyzes in great depth the relationship between house prices and GDP per capita by using quarterly changes in the level of GDP compared with changes in the housing price index. From his analysis, Valadez notes that there exists a linear relationship between the house price index and real GDP, as well as an inverse relationship between home values and vacancies. This research lends many critical findings to my analysis in that I will build on Valadez’s finding that there exists a linear relationship between the variable of GDP and home values.

As Farber (2012) argues, the current duration of unemployment is substantially longer than those studied in previous, weaker labor markets, leading to very little cyclical variation. As the duration of unemployment lengthens, worker mobility declines as there exists a lower demand for durable goods. Farber’s main conclusion reveals the fact that housing prices are affected by both unemployment and the presence of vacancies, both of which have the ability to affect worker mobility, housing accessibility, and ultimately influence house values and prices. Farber ultimately concludes that periods of high unemployment will lead to a decrease in home values and prices, while the inverse will increase house prices. From this, I hypothesize that as unemployment falls, inflation will rise, subsequently increasing home values.

1. **Development of a Testable Hypothesis**

Consistent with the findings of both Shwab (1982) and Head, Llyod-Ellis, and Sun (2014), I expect the coefficient associated with price level to be negative in nature. While an increase in inflation resembles that there has been an increase in the price of goods across a period, which will put upward pressure on the price of inelastic goods, I hypothesize this period under study to be the long run because of the COVID-19 pandemic where consumers undertook home improvement projects and battled through periods of high inflation associated with housing booms. Therefore, I theorize that across this span, home values will fall and settle around their mean value producing a negative sign.

In line with other conclusions drawn in the chosen literature, I have considered GDP per capita, unemployment, vacancy rates, and population growth as control variables. I hypothesize that GDP per capita will have a positive relationship with home values, while vacancy rates will have a negative relationship with home values and prices as analyzed and concluded by Valadez (2011). Moreover, I expect that as a state experience an influx of people, there will be an increase in demand, which will positively impact home values. Furthermore, I theorize a positive relationship between unemployment and home values as Farber (2012) notes due to the level and availability of disposable income. The independent variables discussed above, as well as the expected signs of their marginal influences, second derivatives, and relevant cross derivatives are reported in **Table 1.** I predict that the variable representing unemployment will have a nonlinear effect, meaning that its marginal effect will depend on the level of unemployment in a state, as indicated by its positive cross derivative. As unemployment falls, I theorize that this change will drive home values up, but not at a constant pace. At low levels of unemployment, an increase in the unemployment rate will lead to an increase in home values that is inherently more positive than that of an increase in unemployment at high levels of this variable.

In line with my previous theory as well as Valadez’s (2011) conclusions, I believe that there is indeed an inverse relationship between vacancies and price levels. Furthermore, I hypothesize that the marginal influence of price levels on home values changes as the level of vacancies changes. I have also included a second interaction term between vacancies and population change based on conclusions reached by Head, Llyod-Ellis, and Sun (2014). These scholars found that a change in population density induces increased construction, which slows the growth in vacancies and eventually leads to its reversal, which will subsequently impact home values through vacancy’s marginal influence.

Models 1, 2, and 3 are reported in **Table 2.** Model 1 follows closely in the conclusions contributed by Valadez (2011) in that it adopts a linear relationship between home values and GDP per capita. Model 2 further expands the linear relationship provided in the first model by incorporating an interaction term between vacancies and population growth. I have also included an estimate of the function of the natural log of home values, as in computing my summary statistics, this value was highly skewed, justifying the usage of logs. Vacancies are logged as well to control for the skewness of that variable across states and the response of those states in limiting further vacancies. Model 3 includes an additional interaction term between vacancies and price level, as addressed in the preceding paragraph. Across all models I did not control for state-specific, or time-specific effects that account for unobservable differences across all observations. Subsequently, this means that these differences are most likely included in the error term.

1. **Data and Statistical Assumptions**

This analysis utilizes panel data for the 50 U.S. States and Washington D.C. over two six-month periods, January 2021-July 2021, and July 2021-January 2022, contributing to a sample size with 102 observations. The variables I have used correspond to the economic theory and relevant proxies proposed in Section II. All variables were collected on a state-level basis and are from reputable sources such as the U.S. Census Bureau, the U.S. Economic Committee, the Bureau of Economic Analysis, as well as Zillow. Data and their relevant sources are displayed in **Table 3.** Summary statistics for each variable are reported in **Table 4,** with **Table 5** containing a summary for each period separately. All variables are expressed in terms of their gross growth rate from the prior period. Statistics of note include the increase in the mean gross growth rate of the price level across the second period. Interestingly, the only variables exhibiting skewness are vacancies, price level, and home values, as their mean and median values differ significantly. The standard deviations are relatively low for most included variables, demonstrating that there is little dispersion in this data set.

1. **General Empirical Results**

Utilizing Ordinary Least Squares (OLS) regression techniques, I estimated three regressions corresponding to the models constructed in Section II. All models have been corrected for heteroskedasticity, as encouraged by Stock and Watson (2017) with robust standard errors and other results displayed in **Table 6.** Model 3 indicates the model with the strongest explanatory power, as both the R2, and the adjusted R2 are the highest across all models. Model 3 also has the smallest AIC across all models, further cementing it as the strongest model reported. Both interaction terms are significant at the 10% and 5% significant levels, which further proves their inclusion in this analysis. **Table 7** presents the result of the Ramsey Reset Test for each model. The null hypothesis that the models were well-specified was rejected for all tested models, leading to the conclusion that they are mis-specified. Model 3 does have the lowest F-statistic, however, making it a better fit model than the other two reported.

The Variance Inflation Factors (VIFs) of the independent variables in Model 3 contributed various conclusions. The main variable of interest, the price level, as well as GDP per capita, and the population change all had a VIF below 5 meaning that multicollinearity is not an issue for these variables. However, in the case of the unemployment rate, the squared term for unemployment, the natural log of vacancy, and the two interaction terms, their VIFs were greater than 5, so I conclude that multicollinearity is not a serious concern, as I predicted these variables to be highly correlated with one another.

Price level, while highly significant in Models 1 and 3, does not nearly have the level of explanatory power that I hypothesized, or Shwab (1982) concluded in his analysis. The negative sign associated with price level is in line with my prediction, as well as Shwab’s discourse: Shwab concluded that an increase in the expected rate of inflation raises the value of homes in the short run, but ultimately lowers their value in the long run due to outside policy. In terms of the coefficient associated with vacancy, it is both negative, and statistically significant at the 10% and 5% significance levels, as hypothesized by Head, Llyod-Ellis, and Sun (2014) who determined a negative correlation between house vacancies and home value growth. The included interaction terms also proved to be significant at the 10% and 5% levels as well. Interestingly, the coefficients for both unemployment, and its squared term, are not significant at any level, although unemployment is negative in nature, which is what I originally theorized.

I identified 12 outliers in Model 3 using studentized residuals, and studentized DFFITS, which were all measured at the 5% significance level. As a check for robustness, I estimated Model 3 without these outliers corresponding with the studentized residuals and reported the results in column 4 of **Table 6.** The removal of these outliers increased not only the value for R2 and the adjusted R2 significantly but made the coefficient for both vacancy and the interaction term between vacancy and population growth significant at the 1% level. Interestingly, the coefficient associated with price level lost significance in this check for robustness. None of the signs across all four estimated models changed and were in line with my theory. I determined that removing these 12 outliers was too big of a sacrifice, especially in a data set comprised of only 102 observations. Additionally, the removal of these outliers weakens the significance of the main variable under observation and could potentially limit the robustness of my results. Therefore, I have decided to include them in my analysis.

To account for state-wide influences and time-fixed effects, I utilized panel techniques in order to re-estimate Model 3 for a check of robustness. Omitted variables are of concern to many researchers, so in drawing on panel techniques, I strove to check if these variables were of concern to my model. To determine the correct fixed effects to use, I used hypothesis testing of each estimated panel model, leading to the conclusion that two-ways fixed effects were the correct functional form to utilize. Results of this technique are reported in column 5 of **Table 6.** While the R2 and the adjusted R2 increased significantly in this model, the significance of the coefficient price level switched signs and lost significance. These results imply a multitude of conclusions: while the panel model’s inclusion of fixed effects have helped control for all state and time-specific variables, it has inherently taken away too many degrees of freedom and has caused my results to become insignificant. The decreased significance of price level, its irrelevant sign, and its confidence including both positive and negative values, led me to accept my OLS results as correct.

While the coefficient for price level is significant, I tested for endogeneity concerns in order to determine if there was measurement error in this variable, causing attenuation bias.

While the source that I obtained this metric from is under governance of the United States, they utilize state-level personal consumption measures instead of the Bureau of Labor Statistics’ consumer expenditure survey. While both metrics measure spending, personal consumption expenditures consistently suggest higher aggregate spending levels than the consumer expenditure survey and is derived using government surveys of businesses instead of a survey of households. This disparity leads me to consider a concern of endogeneity in my model because of possible measurement error.

Utilizing the rank of price level of as an instrumental variable, I satisfied the inclusion criterion in that I obtained an F-statistic greater than 10, and a t-statistic significant at all levels. Upon testing that the coefficient associated with the endogenous portion of price level was zero by utilizing the Hausman test, I failed to reject the null, suggesting that price level was exogenous to begin with and therefore, endogeneity is not a concern in this analysis.

1. **Interpretation of the Empirical Results**

Model 3 results illustrate price level’s marginal effect on home values. The coefficients for both the negative, linear representation of price level, as well as the interaction term between price level and vacancy are statistically significant at the 10% and 5% levels. Interestingly, the coefficient for price level has gained additional significance in Model 3 at the 1% level, further justifying the inclusion of the interaction term. The marginal effect of price level can be better visualized in **Graph 1.** The impact of price levels on home value changes depending on whether the gross growth rate of the price level is above or below its median value of 1.080. Although both effects are positive, if the gross growth rate of the price level is below the median, price level’s impact on the home value is significantly more positive than when the price level is about the median value. Furthermore, this graph portrays the effect of price alone, meaning no other variables are controlled for.

**Graph 2** similarly illustrates price level’s marginal effect on home values holding other variables constant. As depicted, price level’s impact on home values is dependent on the gross growth rate of the vacancy rate. An increase in the number of vacancies when the rate is low leads to an increase in terms of home value (as a gross growth rate), but in a state with a previously high vacancy rate, an increase in the vacancy rate will cause a decrease in terms of the home value. However, in the middle of the graph where the curve intersects the x-axis, price level’s effect on home values is ambiguous as it could be either positive or negative depending on the specific sample tested.

While price level is highly significant in Model 3, it does not nearly have the level of explanatory power as hypothesized, even when an interaction term is included. In his analysis, Shwab (1982) determined the elasticity value of the price level to be -0.966, contributing to the conclusion that a small change in overall price level will ultimately result in a noticeable change in home values. In order to test whether price levels are a key determination in the valuation of homes across the United States, I utilized a subset F test in which my null hypothesis was that the restricted model was correct—price levels are not a key driver. As the results display in **Table 8,** I rejected the null in favor of the alternative hypothesis at a significance level of 5% indicating that price levels do determine home prices and valuations.

The coefficient associated with vacancy is both negative, and statistically significant at the 5% level, as hypothesized by Head, Llyod-Ellis, and Sun (2014). While they did not include any interaction terms with vacancy, the interaction term between vacancy and price level proved to be significant at the 10% level. I conducted a hypothesis test to examine whether the number of vacancies within a state are a factor in determining home values, as I theorized that vacancy rates would alter the marginal effect of price level on home values. Conducting a subset F test in which my null hypothesis was that vacancy rates did not impact home values, I concluded that vacancy rates do determine home prices and valuations at the 5% significance level, which justifies the addition of vacancy levels and the included interaction terms. Conclusions are reported in **Table 9.**

Interestingly, the interaction term between vacancies and population change is positive, as the coefficient for population growth is positive and significant at all levels across all models. These results are complementary to my theory, as I hypothesized that as a state experienced population growth, there would be an upward pressure on the demand for houses. In terms of vacancy’s impact on population growth’s marginal effect, I predict that in states with a large number of vacancies, population influx may spur construction and therefore, diminish the effect of population on home values.

1. **Summary and Conclusion**

This analysis expands upon prior literature by estimating a model that includes logged variables to account for high skewness, and interaction terms between vacancy and price level, as well as vacancy and population growth. My results confirmed my theory that the marginal influence of price level is dependent on vacancy levels. An increase in vacancy levels in states that exhibit low levels of vacancies inherently increases home values. However, in states with an existing heightened level of vacancies, there is a decrease in terms of home value. Similarly, the impact of price level on home values is dependent on whether the gross growth rate of the price level is above or below its median value, 1.080. Although both effects are positive, if the gross growth rate of the price level is below the median, price level’s impact on the home value is significantly more positive than when the price level is about the median value.

Consistent with the findings of both Shwab (1982) and Head, Llyod-Ellis, and Sun (2014), I hypothesized that the coefficient associated with price level would be negative in nature. Although price level is highly significant and negative in signage in Model 3, it does not nearly have the level of explanatory power as hypothesized or concluded by Shwab (1982). Although Head, Llyod-Ellis, and Sun (2014) predict this coefficient to be positive in the short run, my conclusions lend to the theory that periods of persistently high inflation cause consumers to spend a higher proportion of their disposable income on goods of necessities, thus lowering the demand for houses, and their subsequent value.

The results of subset F tests affirmed the importance of including price level in this analysis, as well as vacancy, and the relevant interaction terms, which was not studied previously. The presence of outliers was identified and accounted for, and using panel techniques to check for robustness, I conclude that reported OLS results are correct. As to changes in further research and analysis, my results provide reason as to why the unemployment should be excluded from a secondary study, as it, nor the squared term, were significant in Model 3. While I compiled data from two six-month periods, I believe it would be more beneficial to study a plethora of periods taken from multiple years. In addition, a beneficial variable to study might be the cost of renting, as it would be intriguing to note how inflation impacts that as well. I could have also included a dummy variable to represent the income per capita within the state itself. All these variables, while not initially included in my analysis, may offer interesting insight into different factors of determining the median sales price of homes sold within the United States. Inflation, especially as it continues to rise, has many implications for consumers: heightened price levels restrict access to goods of necessity, and as proven in this study, has an adverse effect on home values in the long run. Furthermore, this research contributes to a better understanding of the factors influencing housing prices as it relates to both their short and long-term effects and provides insight for policymakers and investors alike.

Bibliography

Farber, Henry S. 2012. "Unemployment in the Great Recession: Did the Housing Market Crisis Prevent the Unemployed from Moving to Take Jobs?" *The American Economic Review*, 102 (3): 520-25.

Head, Allen, Huw Lloyd-Ellis, and Hongfei Sun. 2014. "Search, Liquidity, and the Dynamics of House Prices and Construction." *American Economic Review*, 104 (4): 1172-1210.

Schwab, Robert M. 1982. “Inflation Expectations and the Demand for Housing.” *The American Economic Review*, *72*(1), 143–153.

Valadez, Ray M. 2011. “The Housing Bubble and GDP: a Correlation Perspective.” *Journal of Case Research in Business and Economics,* 3, 1.

Data

Bureau of Economic Analysis, Gross Domestic Product by State, 4th Quarter and Annual, retrieved from BEA, Bureau of Economic Analysis: U.S. Department of Commerce; <https://apps.bea.gov/regional/histdata/releases/0322gdpstate/index.cfm>

United States Congress Joint Economic Committee, JEC Republicans State Inflation Tracker, retrieved from JEC Senate, United States Congress Joint Economic Committee; <https://www.jec.senate.gov/public/index.cfm/republicans/state-inflation-tracker>

United States Census Bureau, Housing Vacancies and Homeownership, retrieved from Census.gov, United States Census Bureau; <https://www.census.gov/housing/hvs/index.html>

United States Census Bureau, State Population Totals and Components of Change: 2020-2022, retrieved from Census.gov, United States Census Bureau; <https://www.census.gov/data/tables/time-series/demo/popest/2020s-state-total.html>

U.S. Bureau of Labor Statistics, Regional and State Employment and Unemployment (Monthly), retrieved from BLS, U.S. Bureau of Labor Statistics; <https://www.bls.gov/bls/newsrels.htm#OEUS>

Zillow Inc, United States Home Value Index, retrieved from Zillow Group; <https://www.zillow.com/research/data/>

Tables and Figures

*Table 1: Variables and Theory Independent of Models*

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **∂Y/∂X** | **∂2Y/∂X2** | **Relevant Cross-Derivatives** |
| **PriceLevel**  Growth rate of increased price levels | + | 0 | ∂2HomeValue/(∂Vacancies∂PriceLevel)  = - |
| **UNP**  Growth rate of the unemployment rate | - | + |  |
| **GDPPC**  Growth rate of GDP per capita | + | 0 |  |
| **PopChange**  Growth rate of a state’s population | + | 0 | ∂2HomeValue/(∂Vacancies∂PopChange)  = + |
| **Vacancies**  Growth rate of the number of vacancies | - | 0 |  |

*Table 2: Models 1,2, and 3*

|  |  |
| --- | --- |
| **Model** | **Equation** |
| 1. Linear |  |
| 1. Interaction Model |  |
| 1. Quadratic and Interaction Model |  |

*Table 3: Data and Relevant Sources*

|  |  |  |
| --- | --- | --- |
| *Variable* | *Definition* | *Source* |
| **Home Value** | Measured in terms of the growth rate of the value of homes on a state-level; the raw value represents the median value of homes in a particular state | *Zillow* |
| **Unemployment Rate** | Measured in terms of the growth rate of the unemployment rate on a state-level | *U.S. Bureau of Labor Statistics* |
| **Price Level (Inflation)** | The overall increase in prices in terms of growth rate on a state-level basis | *U.S. Congress Joint Economic Committee* |
| **Housing Vacancies** | Measured in terms of the growth rate of the number of vacancies as compared to the prior period on a state-level | *U.S. Census Bureau* |
| **GDP Per Capita** | GDP Per Capita growth rate on a state-level basis | *Bureau of Economic Analysis* |
| **Population Change** | Measured in terms of the growth rate of the population as compared to the prior period on a state-level | *U.S. Census Bureau* |

*Table 4: Variable Summary Statistics*

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Units** | **N** | **Mean** | **Std. Dev** | **Minimum** | **Lower Quartile** | **Median** | **Upper Quartile** | **Maximum** |
| HomeValue | Median value of homes growth rate | 102 | 1.088 | 0.034 | 1.009 | 1.067 | 1.085 | 1.107 | 1.247 |
| PriceLevel | Increase in state-level prices growth rate | 102 | 1.373 | 0.373 | 0.616 | 1.046 | 1.080 | 1.700 | 2.087 |
| UNP | Change in unemployment rate growth rate | 102 | 0.848 | 0.121 | 0.544 | 0.754 | 0.849 | 0.949 | 1.111 |
| GDPPC | GDP per capita growth rate | 102 | 1.044 | 0.011 | 1.011 | 1.038 | 1.043 | 1.049 | 1.076 |
| Vacancies | Number of vacancies growth rate | 102 | 1.102 | 0.596 | 0.211 | 0.778 | 0.943 | 1.241 | 4.750 |
| PopChange | Growth rate of the population vs. prior period | 102 | 1.004 | 0.007 | 0.988 | 0.999 | 1.002 | 1.007 | 1.030 |

*Table 5: Variable Summary Statistics comparing January 2021-July 2021 to July 2021-January 2022*

\*\*The Summary statistics for the first period (Jan. 2021-Jul. 2021) are on the top row, while the summary statistics for the second period (Jul. 2021-Jan. 2022) are on the bottom row.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Units** | **N** | **Mean** | **Std. Dev** | **Minimum** | **Lower Quartile** | **Median** | **Upper Quartile** | **Maximum** |
| HomeValue | Median value of homes growth rate | 51  51 | 1.095  1.080 | 0.038  0.029 | 1.014  1.009 | 1.074  1.061 | 1.089  1.076 | 1.110  1.095 | 1.247  1.149 |
| PriceLevel | Increase in state-level prices growth rate | 51  51 | 1.047  1.700 | 0.011  0.252 | 1.032  0.616 | 1.041  1.613 | 1.046  1.701 | 1.053  1.951 | 1.107  2.087 |
| UNP | Change in unemployment rate growth rate | 51  51 | 0.906  0.789 | 0.093  0.117 | 0.628  0.544 | 0.843  0.710 | 0.918  0.769 | 0.965  0.863 | 1.111  1 |
| GDPPC | GDP per capita growth rate | 51  51 | 1.044  1.044 | 0.012  0.010 | 1.011  1.018 | 1.036  1.040 | 1.041  1.044 | 1.052  1.047 | 1.076  1.070 |
| Vacancies | Number of vacancies growth rate | 51  51 | 1.183  1.021 | 0.711  0.445 | 0.455  0.211 | 0.796  0.745 | 1  0.933 | 1.292  1.191 | 4.750  2.250 |
| PopChange | Growth rate of the population vs. prior period | 51  51 | 1.004  1.004 | 0.007  0.007 | 0.988  0.991 | 0.999  0.999 | 1.002  1.002 | 1.007  1.009 | 1.030  1.019 |

*Table 6: General Empirical Results*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model Estimates** | | | | | |
|  | **Dependent Variable: HomeValues (1), LnHomeValues (2, and 3)** | | | | |
| **OLS** | | | | **PLM** |
| **Parameter** | **(1)** | **(2)** | **(3)** | **(3) – No Studentized Residual Outliers** | **(3) – Utilizing Panel Techniques** |
| **PriceLevel** | -0.020370\*\*  (0.007254) | -0.016894\*  (0.006578) | -0.01934\*\*  (0.00659) | -0.011518\*  (0.005741) | 0.022\*  (0.013) |
| **UNP** | -0.041710 .  (0.022733) | -0.036639 .  (0.020439) | -0.09483  (0.21216) | -0.179228  (0.169645) | -0.001  (0.178) |
| **GDPPC** | 0.209895  (0.233205) | 0.211097  (0.209160) | 0.21861  (0.20657) | 0.306933 .  (0.167403) | 0.069  (0.178) |
| **PopChange** | 3.216845\*\*\*  (0.344092) | 2.713610\*\*\*  (0.318552) | 2.67096\*\*\*  (0.31541) | 2.685468\*\*\*  (0.264928) | 1.671\*  (1.001) |
| **Vacancy** | 0.004004  (0.004190) |  |  |  |  |
| **LnVacancy** |  | -1.5700368\*  (0.687886) | -1.46937\*  (11.08) | -2.094763\*\*\*  (0.601868) | -2.031\*\*\*  (0.691) |
| **LnVacancy\*PopChange** |  | 1.567017\*  (0.685502) | 1.50176\*  (0.67774) | 2.106808\*\*\*  (0.597550) | 2.082\*\*\*  (0.698) |
| **LnVacancy\*PriceLevel** |  |  | -0.02537\*  (0.01189) | -0.009837  (0.010136) | -0.033\*  (0.017) |
| **UNP2** |  |  | 0.03329  (0.12646) | 0.092547  (0.101066) | -0.010  (0.108) |
| **Constant** | -2.300987\*\*\*  (0.396704) | -2.805995\*\*\*  (0.362122) | -2.74331\*\*\*  (0.3662) | -2.831651\*\*\*  (0.298620) |  |
| ***R2*** | 0.5042 | 0.5198 | 0.5424 | 0.6597 | 0.8993 |
| **Adjusted *R2*** | 0.4784 | 0.4895 | 0.503 | 0.628 | 0.7579 |
| **AIC** | -457.0411 | -478.4813 | -479.3931 | -491.5526 | -635.8488 |
| ***F* statistic** | 19.53 | 17.14 | 13.78 | 20.84 | 6.36 |
| **Observations** | 102 | 102 | 102 | 95 | 102 |
| \*\*\*, \*\*, \*, and . denote significance at .1%, 1%, 5%, and 10% levels, respectively. | | | | | |

*Table 7: Ramsey Reset Tests for Models 1,2, and 3*

|  |  |  |  |
| --- | --- | --- | --- |
| **Ramsey Reset Test** | **(1)** | **(2)** | **(3)** |
| F statistic/P-value | 12.84/1.175e-05 | 8.5468/0.0003913 | 6.9666/0.00153 |
| Stated Conclusion | Reject H0 | Reject H0 | Reject H0 |

*Table 8: Subset F Test for Price Levels*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Res. d.o.f.** | **d.o.f.** | **F-statistic** | **Pr(>F)** |
| Restricted Model (1) | 95 | --- | --- | --- |
| Unrestricted Model (1) | 93 | 2 | 3.6259 | 0.03045\* |
| Conclusion | Reject H0 when α=0.05 | | | |
| \*\*\*, \*\*, \*, and . denote significance at .1%, 1%, 5%, and 10% levels, respectively.  The model has been corrected for heteroskedasticity. | | | | |

*Table 9: Subset F Test for Vacancy Levels*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Res. d.o.f.** | **d.o.f.** | **F-statistic** | **Pr(>F)** |
| Restricted Model (1) | 96 | --- | --- | --- |
| Unrestricted Model (1) | 93 | 3 | 2.8307 | 0.04263\* |
| Conclusion | Reject H0 when α=0.05 | | | |
| \*\*\*, \*\*, \*, and . denote significance at .1%, 1%, 5%, and 10% levels, respectively.  The model has been corrected for heteroskedasticity. | | | | |

*Graph 1: Price Level’s Sole Marginal Effect on Home Values*

Chart, scatter chart

Description automatically generated

*Graph 2: Price Level’s Marginal Effect on Home Values Holding Variables Constant*

Chart, line chart

Description automatically generated

R Appendix

library(readxl)

COMBINED <- read\_excel("Downloads/ECO 205 Final Paper Data.xlsx")

View(COMBINED)

install.packages("moments")

library(moments)

install.packages("stargazer")

library(stargazer)

library(dplyr)

library(readxl)

library(sandwich)

library(lmtest)

library(sjPlot)

library(Greg)

library(car)

Project2=data.frame(UnemploymentRate = COMBINED$"Growth Rate in Unemployment Rate, Seasonaly Adjusted from January 31 to July 31 2021",

PriceLevel = COMBINED$"Growth Rate in Increase in Prices Rate from January 31 to July 31 2021",

HomeValue = COMBINED$"Growth Rate in United States Median House Prices in terms of Home Value in USD from January 31 to July 31 2021",

GDPPC = COMBINED$"Growth Rate in Current-Dollar GDP in millions of USD from January 31 to July 31 2021",

Vacancy = COMBINED$"Growth Rate in Homeowner Vacancy Rates from January 31 to July 31 2021",

PopChange = COMBINED$"Growth Rate in United States Population from January 31 to July 31 2021")

View(Project2)

Averages=sapply(Project2, mean)

print(Averages)

Stdevs=sapply(Project2, sd)

Mins=sapply(Project2, min)

Q1s=sapply(Project2, quantile, 0.25)

Medians=sapply(Project2, median)

Q3s=sapply(Project2, quantile, 0.75)

Maxs=sapply(Project2, max)

Numobs=nrow(Project2)-sapply(Project2, function(x) sum(is.na(x)))

Skewvals=sapply(Project2, skewness)

Kurtvals=sapply(Project2, kurtosis)

Summary=data.frame(Average = Averages,

Std\_Dev = Stdevs,

Minimum = Mins,

First\_Quartile = Q1s,

Median = Medians,

Third\_Quartile = Q3s,

Maximum = Maxs,

Skewnewss = Skewvals,

Kurtosis = Kurtvals,

Number\_of\_Obs = Numobs)

print(Summary)

DistributionSummaryOrganized <- t(Summary)

print(DistributionSummaryOrganized)

stargazer(DistributionSummaryOrganized, summary=FALSE,type='text')

#model 1 computation

model1 = lm(HomeValue ~ PriceLevel + UnemploymentRate + GDPPC + PopChange + Vacancy, data=Project2)

summary(model1)

AIC1=AIC(model1)

AIC1

#model 2 computation

Project2$HomeValueLn=log(Project2$HomeValue)

Project2$VacancyLn=log(Project2$Vacancy)

Project2$VacancyLnPop = Project2$VacancyLn \* Project2$PopChange

model2 = lm(HomeValueLn ~ PriceLevel + UnemploymentRate + GDPPC + PopChange + VacancyLn + VacancyLnPop, data=Project2)

summary(model2)

AIC2=AIC(model2)

AIC2

#model 3 computation

Project2$UnemploymentRate2=Project2$UnemploymentRate^2

Project2$VacancyLnPriceLevel = Project2$VacancyLn \* Project2$PriceLevel

model3 = lm(HomeValueLn ~ PriceLevel + UnemploymentRate + UnemploymentRate2 + GDPPC + PopChange + VacancyLn + VacancyLnPop + VacancyLnPriceLevel, data=Project2)

summary(model3)

AIC3=AIC(model3)

AIC3

#Correcting for Heteroskedasticity

vcov1 <- vcovHC(model1, type = "HC1")

vcov1

robust\_se1 <- sqrt(diag(vcov1))

robust\_se1

coeftable\_robust1=coeftest(model1, vcov.=vcov1)

coeftable\_robust1

tab\_model(model1,vcov.fun = vcov1, show.se = TRUE)

vcov2 <- vcovHC(model2, type = "HC1")

vcov2

robust\_se2 <- sqrt(diag(vcov2))

robust\_se2

coeftable\_robust2=coeftest(model2, vcov.=vcov)

coeftable\_robust2

tab\_model(model2,vcov.fun = vcov2, show.se = TRUE)

vcov3 <- vcovHC(model3, type = "HC1")

vcov3

robust\_se3 <- sqrt(diag(vcov3))

robust\_se3

coeftable\_robust3=coeftest(model3, vcov.=vcov3)

coeftable\_robust3

tab\_model(model3,vcov.fun = vcov3, show.se = TRUE)

#Ramsay RESET Test

resettest(model1,power=2:3)

resettest(model2,power=2:3)

resettest(model3,power=2:3)

#VIF calculation

vif(model1)

vif(model2)

vif(model3)

#Computing Outliers for Model 3

Project2Outliers=Project2

View(Project2Outliers)

#1. Leverage:

Project2Outliers$hats=hatvalues(model3)

View(Project2Outliers)

#2. Studentized residuals:

Project2Outliers$rstuds=rstudent(model3)

View(Project2Outliers)

summary(model3)

tcrit=qt(p=.05/2, model3$df.residual-1, lower.tail=FALSE)

tcrit

Project2Outliers$rstudoutliers=ifelse(abs(Project2Outliers$rstuds)>tcrit,1,0)

Project2.without.Rstud.outliers = subset(Project2Outliers, rstudoutliers!=1)

View(Project2.without.Rstud.outliers)

model3robust1=lm(HomeValueLn ~ PriceLevel + UnemploymentRate + UnemploymentRate2 + GDPPC + PopChange + VacancyLn + VacancyLnPop + VacancyLnPriceLevel, data=Project2.without.Rstud.outliers)

summary(model3robust1)

summary(model3)

AIC4=AIC(model3robust1)

AIC4

vif(model3robust1)

resettest(model3robust1,power=2:3)

#3. DFFITS:

Project2Outliers$dffitsvals=dffits(model3)

p = length(model3$coefficients)-1

n = nrow(Project2)

DFFITSthresh <- 2\*sqrt(p/n)

DFFITSthresh

Project2Outliers$dffitsoutliers=ifelse(abs(Project2Outliers$dffitsvals)>DFFITSthresh,1,0)

Project2.without.DFFITS.outliers = subset(Project2Outliers, dffitsoutliers!=1)

View(Project2.without.DFFITS.outliers)

model3robust2=lm(HomeValueLn ~ PriceLevel + UnemploymentRate + UnemploymentRate2 + GDPPC + PopChange + VacancyLn + VacancyLnPop + VacancyLnPriceLevel, data=Project2.without.DFFITS.outliers)

summary(model3robust2)

summary(model3)

#Subset F tests for Price Level

linearHypothesis(model3, c("PriceLevel", "VacancyLnPriceLevel"))

linearHypothesis(model3, c("PriceLevel", "VacancyLnPriceLevel"), white.adjust = "hc1") #Corrected for heteroskedasticity

#Subset F tests for Vacancy

linearHypothesis(model3, c("VacancyLn","VacancyLnPop", "VacancyLnPriceLevel"))

linearHypothesis(model3, c("VacancyLn","VacancyLnPop", "VacancyLnPriceLevel"), white.adjust = "hc1") #Corrected for heteroskedasticity

#Plotting Marginal Influence of Price Level

regvcov=vcov(model3)

regvcov

#plot the confidence interval for dLnHomeValue/dPriceLevel across values of dPriceLevel

curve(coef(summary(model3))[2,1]+x\*coef(summary(model3))[9,1], from=DistributionSummaryOrganized[3,2],

to=DistributionSummaryOrganized[7,2], ylim=c(-0.05, 0.05),

col='blue', xlab="Price Level (Gross Growth Rate)",

ylab="dHomeValueLn/dPriceLevel")

curve(coef(summary(model3))[2,1]\*DistributionSummaryOrganized[1,7]+x\*coef(summary(model3))[9,1]\*DistributionSummaryOrganized[1,7]+1.96\*sqrt(regvcov[2,2] + x^2\*regvcov[9,9] + 2\*x\*regvcov[2,9]), from=DistributionSummaryOrganized[3,2], to=DistributionSummaryOrganized[7,2], add=TRUE, ylim=c(-0.05, 0.05), col='red')

curve(coef(summary(model3))[2,1]\*DistributionSummaryOrganized[1,7]+x\*coef(summary(model3))[9,1]\*DistributionSummaryOrganized[1,7]-1.96\*sqrt(regvcov[2,2] + x^2\*regvcov[9,9] + 2\*x\*regvcov[2,9]), from=DistributionSummaryOrganized[3,2], to=DistributionSummaryOrganized[7,2], add=TRUE, ylim=c(-0.05, 0.05), col='red')

#can repeat with vcovHC

curve(100\*(coef(summary(model3))[2,1]+x\*coef(summary(model3))[9,1]), from=DistributionSummaryOrganized[3,8],

to=DistributionSummaryOrganized[7,8], ylim=c(-12,12),

col='blue', xlab="Gross Growth Rate of Vacancy Levels (Ln)",

ylab="(dHomeValueLn/HomeValueLn\*100%)/dPriceLevel")

curve(100\*(coef(summary(model3))[2,1]+x\*coef(summary(model3))[9,1]+1.96\*sqrt(hetvcov[2,2] + x^2\*hetvcov[9,9] + 2\*x\*hetvcov[2,9])), from=DistributionSummaryOrganized[3,8], to=DistributionSummaryOrganized[7,8], add=TRUE, col='red')

curve(100\*(coef(summary(model3))[2,1]+x\*coef(summary(model3))[9,1]-1.96\*sqrt(hetvcov[2,2] + x^2\*hetvcov[9,9] + 2\*x\*hetvcov[2,9])), from=DistributionSummaryOrganized[3,8], to=DistributionSummaryOrganized[7,8], add=TRUE, col='red')

#Scatter plot for PriceLevel values

plot(Project2$PriceLevel, Project2$HomeValueLn, xlab="Gross Growth Rate of Price Level", ylab="Gross Growth Rate of Home Value (Ln)", xlim=c(0, 2.5), ylim=c(0,0.25))

abline(v=DistributionSummaryOrganized[5,2], col="blue", lwd=3)

text(x=0.75, y=0.2, 'Median ==>')

Project2SubLow=subset(Project2, PriceLevel<DistributionSummaryOrganized[5,2])

#View(Project2SubLow)

low = lm(HomeValueLn ~ PriceLevel, data=Project2SubLow)

summary(low)

#abline(low)

segments(x0=0,y0=low$coefficients[1],x1=DistributionSummaryOrganized[5,2],y1=low$coefficients\*DistributionSummaryOrganized[5,2],col="red", lwd=2)

Project2SubHigh=subset(Project2, PriceLevel>DistributionSummaryOrganized[5,2])

#View(Project2SubLow)

high = lm(HomeValueLn ~ PriceLevel, data=Project2SubHigh)

summary(high)

#abline(high)

segments(x0=DistributionSummaryOrganized[5,2],y0=high$coefficients[1]+high$coefficients[2]\*DistributionSummaryOrganized[5,2],x1=2.5,y1=high$coefficients[1]+high$coefficients[2]\*2.5,col="green",lwd=2)

#Endogeneity Check

#Perform the first stage

Project2$rankPriceLevel=rank(Project2$PriceLevel)

model3\_multi\_s1 <- lm(PriceLevel ~ UnemploymentRate + UnemploymentRate2 + GDPPC + PopChange + VacancyLn + VacancyLnPop + VacancyLnPriceLevel + rankPriceLevel, data = Project2)

summary(model3\_multi\_s1)

coeftest(model3\_multi\_s1, vcov = vcovHC, type = "HC1")

#Do IVreg

model3\_multi <- ivreg(HomeValueLn ~ PriceLevel + UnemploymentRate + UnemploymentRate2 + GDPPC + PopChange + VacancyLn + VacancyLnPop + VacancyLnPriceLevel | UnemploymentRate + UnemploymentRate2 + GDPPC + PopChange + VacancyLn + VacancyLnPop + VacancyLnPriceLevel + rankPriceLevel, data = Project2)

coeftest(model3\_multi, vcov = vcovHC, type = "HC1")

#Do Hausman Test

Project2$multi\_pred <- model3\_multi\_s1$fitted.values

Project2$multi\_resid <- model3\_multi\_s1$residuals

model3\_multi\_hausman1 <- lm(HomeValueLn ~ PriceLevel + UnemploymentRate + UnemploymentRate2 + GDPPC + PopChange + VacancyLn + VacancyLnPop + VacancyLnPriceLevel + multi\_pred, data=Project2)

coeftest(model3\_multi\_hausman1, vcov = vcovHC, type = "HC1")

model3\_multi\_hausman2 <- lm(HomeValueLn ~ PriceLevel + UnemploymentRate + UnemploymentRate2 + GDPPC + PopChange + VacancyLn + VacancyLnPop + VacancyLnPriceLevel + multi\_resid, data=Project2)

coeftest(model3\_multi\_hausman2, vcov = vcovHC, type = "HC1")

#Utilizing PanelData Techniques

Project2Panel=data.frame(UnemploymentRate = COMBINED$"Growth Rate in Unemployment Rate, Seasonaly Adjusted from January 31 to July 31 2021",

PriceLevel = COMBINED$"Growth Rate in Increase in Prices Rate from January 31 to July 31 2021",

HomeValue = COMBINED$"Growth Rate in United States Median House Prices in terms of Home Value in USD from January 31 to July 31 2021",

GDPPC = COMBINED$"Growth Rate in Current-Dollar GDP in millions of USD from January 31 to July 31 2021",

Vacancy = COMBINED$"Growth Rate in Homeowner Vacancy Rates from January 31 to July 31 2021",

PopChange = COMBINED$"Growth Rate in United States Population from January 31 to July 31 2021",

State=COMBINED$"State",

Period=COMBINED$"Period")

Project2Panel$State=as.factor(Project2Panel$State) #this function converts any variable to a categorical variables - it considers each unique entry in the column as its own category

str(Project2Panel)

#defining variables

Project2Panel$HomeValueLn=log(Project2Panel$HomeValue)

Project2Panel$VacancyLn=log(Project2Panel$Vacancy)

Project2Panel$VacancyLnPop = Project2Panel$VacancyLn \* Project2Panel$PopChange

Project2Panel$UnemploymentRate2=Project2Panel$UnemploymentRate^2

Project2Panel$VacancyLnPriceLevel = Project2Panel$VacancyLn \* Project2Panel$PriceLevel

#model 3 computations

panel\_pind3 <- plm(HomeValueLn ~ PriceLevel + UnemploymentRate + UnemploymentRate2 + GDPPC + PopChange + VacancyLn + VacancyLnPop + VacancyLnPriceLevel, data = Project2Panel, index=c("State", "Period"), model="pooling")

summary(panel\_pind3)

panel\_stateind3 <- plm(HomeValueLn ~ PriceLevel + UnemploymentRate + UnemploymentRate2 + GDPPC + PopChange + VacancyLn + VacancyLnPop + VacancyLnPriceLevel, data = Project2Panel, index=c("State", "Period"), model="within", effect="individual")

summary(panel\_stateind3)

panel\_timeind3 <- plm(HomeValueLn ~ PriceLevel + UnemploymentRate + UnemploymentRate2 + GDPPC + PopChange + VacancyLn + VacancyLnPop + VacancyLnPriceLevel, data = Project2Panel, index=c("State", "Period"), model="within", effect="time")

summary(panel\_timeind3)

panel\_twowaysind3 <- plm(HomeValueLn ~ PriceLevel + UnemploymentRate + UnemploymentRate2 + GDPPC + PopChange + VacancyLn + VacancyLnPop + VacancyLnPriceLevel, data = Project2Panel, index=c("State", "Period"), model="within", effect="twoways")

summary(panel\_twowaysind3)

#reporting standard errors

rob\_seind <- list(sqrt(diag(vcovHC(panel\_pind3, type = "HC1"))),

sqrt(diag(vcovHC(panel\_stateind3, type = "HC1"))),

sqrt(diag(vcovHC(panel\_timeind3, type = "HC1"))),

sqrt(diag(vcovHC(panel\_twowaysind3, type = "HC1"))))

stargazer(panel\_pind3, panel\_stateind3, panel\_timeind3, panel\_twowaysind3,

type="text",

header= FALSE,

se=rob\_seind,

title='Panel models with home value')

#F-test to determine best effects

pFtest(panel\_stateind3, panel\_pind3)

pFtest(panel\_timeind3, panel\_pind3)

pFtest(panel\_twowaysind3, panel\_stateind3)

pFtest(panel\_twowaysind3, panel\_timeind3)

pFtest(panel\_twowaysind3, panel\_pind3)

#OLS with factor (state and period)

model3olsplm <- lm(HomeValueLn ~ PriceLevel + UnemploymentRate + UnemploymentRate2 + GDPPC + PopChange + VacancyLn + VacancyLnPop + VacancyLnPriceLevel + factor(State) + factor(Period), data = Project2Panel)

summary(model3olsplm)

AIC3plm=AIC(panel\_twowaysind3)

AIC3plm

vif(model3olsplm)

#Ramsay RESET (w/ factor model) and Subset F-Tests on Panel model

resettest(model3olsplm)

linearHypothesis(panel\_twowaysind3, c("PriceLevel", "VacancyLnPriceLevel"), white.adjust = "hc1") #Corrected for heteroskedasticity

linearHypothesis(panel\_twowaysind3, c("VacancyLn","VacancyLnPop", "VacancyLnPriceLevel"), white.adjust = "hc1") #Corrected for heteroskedasticity