IMPACT OF UNEMPLOYMENT, WAGES AND COST OF LIVING ON HOUSE PRICES IN DIFFERENT STATES IN US

A Panel Data Analysis

Karthikeya Nalluri, Miles Marimbire

Introduction

This project involved Unemployment Rate data, Average Hourly Earnings, Home Price Index, and Regional Price Parities: Housing from data ranging from the years 2017 to 2021, all the data is state level based and we have collected the data across all states within the United States. The data was collected from the website (https://fred.stlouisfed.org/). The kind of application we sought to implement was to see how the Unemployment rate, average hourly earnings, and Regional Price Parities of each state impacts the Housing price index. We first came across this question after reading the article, “Is Local Unemployment Related to Local Housing Prices?” by the Economic Research Federal Reserve Bank of ST. Louis. The main premise of the article was to determine whether regional labor and housing markets are linked with one another. Throughout the article they noticed that there was a high negative correlation between the percent change in county house prices with the change in the county unemployment rate. This meant that the states with large decreases in housing prices experienced a much higher rate of unemployment. This lit up a spark of curiosity within our minds to investigate this assumption and verify the validity of this claim. The next article we looked at was, “The Effects of Housing Price on Unemployment Rate and Stock Market” and this section details how when China’s real estate market development section increased, so did employment as they assumed that new real estate would lead to more jobs being created. Therefore, concluding that the relationships between housing prices and unemployment rates have an inverse relationship and that is if one goes up the other will typically go down. So, our goal in the research paper is to investigate how unemployment affects housing prices but we also look at how wages impact it as well and the cost of living. We believe this is a unique point of research as most of the literature we have covered looks at how unemployment as well as things such as the stock market or other financial securities effect it whilst we examine how consumers cost of living and affordability affects the prices.

Data

Now that we have described the objective of our research, we will be providing an in-depth explanation of our data set. Below is a descriptive statistics table of each of our variables.

Unemployment Rate	Average Hourly Earnings	Regional Price Parities: Housing	Housing Pricing Index
Min. ~ 2.100	Min. ~ 16.81	Min. ~ 55.27	Min. ~ 221.7
1st Qu. ~ 3.400	1st Qu. ~ 22.95	1st Qu. ~ 72.33	1st Qu. ~ 332.5
Median. ~ 4.183	Median. ~ 25.75	Median. ~ 81.90	Median. ~ 402.3
Mean. ~ 4.686	Mean. ~ 25.69	Mean. ~ 90.70	Mean. ~ 432.0
3rd Qu. ~ 5.508	3rd Qu. ~ 27.92	3rd Qu. ~ 108.59	3rd Qu. ~ 501.2
Max. ~ 13.700	Max. ~ 37.75	Max. ~ 166.07	Max. ~ 956.8
SD. ~ 1.792204	SD. ~ 4.108468	SD. ~ 24.81322	SD. ~ 133.9269
Observations. ~ 245	Observations. ~ 245	Observations. ~ 245	Observations. ~ 245

All our variables were collected from the Fred website. The ranges are from 2017 until 2021. Below is a description of each variable:

Unemployment Rate in State
DOWNLOAD
Observation:
Feb 2023: **4.3**(+ more) Updated: Mar 27, 2023
Units:
Percent,
Seasonally Adjusted
Frequency:
Annual

Average Hourly Earnings of All Employees, Manufacturing in State
DOWNLOAD
Observation:
Mar 2023: **31.83**(+ more) Updated: Apr 7, 2023

Units:
Dollars per Hour,
Seasonally Adjusted

Frequency:
Annual

Regional Price Parities: Services: Housing for State
DOWNLOAD
Observation:
2021: **88.982**(+ more) Updated: Dec 15, 2022

Units:
Index,
Not Seasonally Adjusted

Frequency:
Annual

All-Transactions House Price Index for the State
DOWNLOAD
Observation:
Q4 2022: **623.66**(+ more) Updated: Feb 28, 2023

Units:
Index 1980:Q1=100,
Not Seasonally Adjusted

Frequency:
Annual

Before getting into the models, we wanted to check if the data is balanced or unbalanced. Hence, we test and conclude that the data is balanced.

```
Balanced Panel: n = 49, T = 5, N = 245
```

Here n= 49 as we are considering data for 49 states in US and T= 5 as we considering time periods from 2017 to 2021. Hence the total number of observations is 245.

We took the log for Housing price Index (HPI), Regional Price Parities : housing (RPP), Average Hourly Earnings of all employees(WAGE).

Since the data is of a panel data from 2017 to 2021, we can use both dynamic model and static model. We started with dynamic model first as we have data for various periods and calculate lags for variables. The following is the dynamic model specification.

Dynamic Model specification:

$\text{Log(HPI)}_{it} = \gamma\ \text{Log(HPI)}_{it-1} + \beta_1(\text{Unemployment rate})_{it} + \beta_2(\text{Unemployment rate})_{it-1} + \beta_3\text{log(RPP)}_{it} + \beta 4\text{log(RPP)}_{it-1} + \beta_5\text{log(WAGE)}_{it} + \beta_6\text{log(WAGE)}_{it-1} + d_t + \alpha_i + \varepsilon_{it}$

Where,

Log(HPI)_{it} is the log of Housing Price Index in state i at time period t

$\text{Unemployment rate}_{it}$ is Unemployment rate of state i at time period t

log(RPP)_{it} is the log of Regional Price Parities of housing in state i at time period t

log(WAGE)_{it} is log of Average hourly earnings of all employees in state i at time period t.

dt is the time dummies where dt = $\beta_7 2019 + \beta_8 2020 + \beta_9 2021$

α_i is the unobserved heterogeneity.

ε_{it}- is the Idiosyncratic error.

For simplicity, we assume that all the explanatory variables are exogenous.

We tried using a dynamic model for the given panel data. Hence, we have data for five time periods, we used to take lag of each dependent variables and running two step Arellano and Bond estimation rather than Anderson-Hsiao 2SLS estimation.

The following is the output for two-step Arellano and Bond estimation.

```
Twoways effects Two-steps model Difference GMM

Call:
pgmm(formula = LHPI ~ lag(LHPI) + Unemployment.Rate + lag(Unemployment.Rate) +
    LRPP + lag(LRPP) + LWAGE + lag(LWAGE) | lag(LHPI, 2:99),
    data = data.panel, effect = "twoways", model = "twostep")

Balanced Panel: n = 49, T = 5, N = 245

Number of Observations Used: 147
Residuals:
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
-0.097822 -0.009764 -0.001413 -0.001425  0.008551  0.047826

Coefficients:
                          Estimate  Std. Error z-value             Pr(>|z|)
lag(LHPI)               1.41688461  0.17780629  7.9687 0.0000000000000001604 ***
Unemployment.Rate       0.00086119  0.00112137  0.7680              0.44250
lag(Unemployment.Rate)  0.00124933  0.00151709  0.8235              0.41022
LRPP                    0.01096973  0.08403081  0.1305              0.89614
lag(LRPP)               0.02819532  0.09733794  0.2897              0.77207
LWAGE                   0.06080693  0.07397089  0.8220              0.41105
lag(LWAGE)             -0.03576621  0.07691911 -0.4650              0.64194
2019                   -0.02017793  0.01128620 -1.7878              0.07380 .
2020                   -0.03461432  0.01976855 -1.7510              0.07995 .
2021                    0.01552379  0.02665076  0.5825              0.56024
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Sargan test: chisq(5) = 19.60223 (p-value = 0.0014837)
Autocorrelation test (1): normal = 1.578933 (p-value = 0.11435)
Autocorrelation test (2): normal = -3.039575 (p-value = 0.0023691)
Wald test for coefficients: chisq(7) = 121.0132 (p-value = < 0.0000000000000000222)
Wald test for time dummies: chisq(3) = 157.4987 (p-value = < 0.0000000000000000222)
```

Here, we can observe that the value of γ is 1.41688 which is greater than 0.8. In Monte Carlo, 0.8 is treated as threshold and since the value of γ is greater than 0.8, there means there is evidence of weak Instrumental Variable (IV)

From the output results, we can observe Sargan Test results, Autocorrelation Test(1) and (2) test results.

Sargan test is the test for Instrumental Variables (IV)
In this test, the Null Hypothesis H_0 is Moment condition ((instruments used in the instrumental variables (IV) regression model are uncorrelated with the error term)

The moment conditions in dynamic panel data models are often referred to as "orthogonality conditions." This is because, in order to estimate the parameters of the model, the sample moments of the data must be orthogonal (uncorrelated) to the parameters. If the moment conditions are satisfied, the GMM estimator produces consistent and efficient estimates of the parameters.

Here, the p-value for Sargan test is 0.0014837 which is less than 0.05. Hence, we reject the Null Hypothesis and conclude that there is no moment condition, and the model is not reasonable.

The Auto correlation tests(both first and second order) are for auto correlation among residuals:

Auto correlation test(1)
The null hypothesis for this test is,
H_0: the residuals are uncorrelated with their own one period lagged values.

Here, from the output, we can observe that the p-value is 0.11435 which is greater than 0.05. Hence, we fail to reject null hypothesis and it means the residuals are uncorrelated.

Auto correlation test(2)
The null hypothesis for this test is,

Ho: the residuals are uncorrelated with their own lagged values two periods ago

Here, from the output, we can observe that the p-value is equal to 0.0023691 which is less than 0.05. It means we reject the null hypothesis and conclude that the residuals are correlated with their own lagged values two periods ago.

In the dynamic model, we can see that orthogonality and relevancy conditions are not met and lag values of explanatory variables doesn’t seem to be significantly affecting the dependent variable, so we decided to try dynamic model excluding lagged dependent variable.

Dynamic Model (excluding lag dependent variable)

The following is the dynamic model regression excluding lagged dependent variable.

$\text{Log(HPI)}_{it} = \beta_1(\text{Unemployment rate})_{it} + \beta_2(\text{Unemployment rate})_{it-1} + \beta_3\log(\text{RPP})_{it} + \beta 4\log(\text{RPP})_{it-1} + \beta_5\log(\text{WAGE})_{it} + \beta_6\log(\text{WAGE})_{it-1} + d_t + \alpha_i + \varepsilon_{it}$

```
Coefficients:
                        Estimate  Std. Error z-value              Pr(>|z|)
Unemployment.Rate      0.00024051  0.00095351  0.2522                0.8009
lag(Unemployment.Rate) 0.00131947  0.00161169  0.8187                0.4130
LRPP                   0.20087418  0.14081088  1.4266                0.1537
lag(LRPP)              0.19360393  0.10464378  1.8501                0.0643 .
LWAGE                 -0.00039040  0.11642942 -0.0034                0.9973
lag(LWAGE)            -0.09781452  0.09835182 -0.9945                0.3200
2019                   0.04940827  0.00739275  6.6833      0.00000000002336 ***
2020                   0.10269151  0.01542704  6.6566      0.00000000002802 ***
2021                   0.22894301  0.02514454  9.1051 < 0.0000000000000022 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Sargan test: chisq(6) = 21.85442 (p-value = 0.0012867)
Autocorrelation test (1): normal = 2.014564 (p-value = 0.04395)
Autocorrelation test (2): normal = 2.142124 (p-value = 0.032183)
Wald test for coefficients: chisq(6) = 7.306775 (p-value = 0.29341)
Wald test for time dummies: chisq(3) = 231.4601 (p-value = < 0.000000000000000222)
```

The Sargan test is p-value = 0.0012867 which is less than 0.05. Hence, reject null hypothesis and conclude that the model is not reasonable or good.

Autocorrelation test(1) p-value = 0.04395 <0.05. Hence, reject null hypothesis and conclude that the residuals are correlated with their own first lag.

Autocorrelation test(2) p-value = 0.032183 < 0.05. Hence, reject null hypothesis and conclude that, the residuals are correlated with their own second lag.

In the dynamic model, we can see that orthogonality and relevancy conditions are not met and lag values of explanatory doesn't seem to be significantly (both including and excluding lag dependent variable as explanatory variables) affecting the dependent variable so we decided to go with static model.

Static Model Specification

In static model, we tried to estimate various estimators like pooled OLS, Random Effect Estimator (RE), First Differenced(FD) estimator, Fixed Effects (FE) estimator and got the output for these estimators.

Pooled OLS Estimator:

```
> pols<-plm(LHPI ~ Unemployment.Rate + LRPP + LWAGE + factor(year), data=data.panel, model="pooling")
> summary(pols)
Pooling Model

Call:
plm(formula = LHPI ~ Unemployment.Rate + LRPP + LWAGE + factor(year),
    data = data.panel, model = "pooling")

Balanced Panel: n = 49, T = 5, N = 245

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-0.370622 -0.119205 -0.012649  0.118573  0.529816

Coefficients:
                    Estimate Std. Error t-value              Pr(>|t|)
(Intercept)        2.3702634  0.2313773 10.2441 < 0.0000000000000022 ***
Unemployment.Rate -0.0293380  0.0074824 -3.9209             0.0001155 ***
LRPP               0.9157582  0.0459343 19.9363 < 0.0000000000000022 ***
LWAGE             -0.1377354  0.0768389 -1.7925             0.0743245 .
factor(year)2018   0.0429182  0.0327935  1.3087             0.1918899
factor(year)2019   0.1123486  0.0330379  3.4006             0.0007889 ***
factor(year)2020   0.2469161  0.0394971  6.2515        0.000000001875 ***
factor(year)2021   0.3050013  0.0340823  8.9490 < 0.0000000000000022 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    20.753
Residual Sum of Squares: 6.1829
R-Squared:      0.70207
Adj. R-Squared: 0.69327
F-statistic: 79.7834 on 7 and 237 DF, p-value: < 0.000000000000000222
```

From the above output of Pooled OLS estimator, we can observe that all the variables are significant except logWAGE at 5 percent significance level.

If unemployment rate increases by 1 unit, then HPI will decrease by 2.93%

If RPP increases by 1 % then HPI will increase by 0.916%

We used Breusch-Pagan test (BP test) and found the below results.

```
> bptest(pols)

        studentized Breusch-Pagan test

data:  pols
BP = 5.3873, df = 7, p-value = 0.6128
```

We use BP test to check for unobserved heterogeneity. The null hypothesis for BP test is the error variances are equal to 0. Here, we can observe that the p-value is 0.6128 which is greater than 0.5. Hence, we fail to reject the null hypothesis and conclude that pooled OLS is preferred estimator for the model.

But finalizing pooled OLS as the best estimator, we tried to estimate through other estimators and formulated the below combined output.

Table 1 LHPI - pooled OLS RE FD FE

	Dependent variable: LHPI			
	(1)	(2)	(3)	(4)
Unemployment.Rate	-0.029***	-0.001	-0.0003	-0.001
	(0.007)	(0.002)	(0.001)	(0.002)
LRPP	0.916***	0.602***	0.112***	0.424***
	(0.046)	(0.056)	(0.043)	(0.070)
LWAGE	-0.138*	-0.060	-0.013	-0.098*
	(0.077)	(0.056)	(0.051)	(0.058)
Constant	2.370***	3.412***	0.069***	
	(0.231)	(0.302)	(0.002)	
Year Dummies	Yes	Yes	Yes	Yes
Observations	245	245	196	245
R2	0.702	0.923	0.718	0.940
Note:	*p<0.1; **p<0.05; ***p<0.01			

We can observe that all the estimators have 245 observations except FD estimator as we take first difference and hence the 49 variables will be eliminated.

All the variables have significant effect on log HPI only in pooled OLS estimator.

From the above combined output, for Random Effect (RE) Estimator, we can observe that unemployment rate and wage do not have any significance, but the other variables have significance on log HPI. If RPP increases by 1%, then HPI will increase by 0.601%

From the above combined output, for First Differenced(FD) estimator we can observe that RPP has significance on HPI as it is significant at 5% level. If RPP increases by 1% then HPI will increase by 0.112%.

From the above combined output, for Fixed Effects (FE) estimator we can observe that unemployment rate and WAGE are not significant, but RPP is significant at 5% level. If RPP increases by 1% then HPI will increase by 0.424%

Hence after considering all the estimators, we conclude that pooled OLS estimator to be the best estimator.

Conclusion

After considering different models like dynamic models (including and excluding lagged dependent variable) and static model, we can say that unemployment rate, Regional House Parities: Housing, Average hourly earnings of all employees in the state have an impact on House price index. Considering the different static model estimators, we concluded that pooled OLS estimator provides best results when compared to other estimators. The unemployment rate and average hourly rate has a negative impact and regional price parities: Housing has positive impact on Housing Price Index.

References

Employment research. (n.d.). Retrieved April 25, 2023, from <https://research.stlouisfed.org/publications/employment-research/is-local-unemployment-related-to-local-housing-prices>

Sun N, 'The Effects of Housing Price on Unemployment Rate and Stock Market' , 'International Journal of Trade, Economics and Finance, Vol. 12, No. 5',

Irandoust, M. (n.d.). House prices and unemployment: An empirical analysis of causality. Retrieved April 26, 2023, from <https://www.researchgate.net/publication/325596404_House_Prices_and_Unemployment_An_Empirical_Analysis_of_Causality>

Irandoust, M. (n.d.). House prices and unemployment: An empirical analysis of causality. Retrieved April 26, 2023, from <https://www.researchgate.net/publication/325596404_House_Prices_and_Unemployment_An_Empirical_Analysis_of_Causality>