FORECASTING U.S HOUSE PRICES

INTRODUCTION

The goal of this project is to develop an accurate forecast of U.S. house prices 4 months ahead. An accurate prediction of the house price is important to prospective homeowners, investors, and other real estate market participants, such as mortgage lenders and insurers. Therefore, the availability of a house price prediction model helps fill up an important information gap and improve the efficiency of the real estate market.

I want to decide whether to buy a house. Specifically, I am considering three prices to take the decision, i.e., the Current price of the house, the forecasting price of the house, and the future price of the house. If the forecast shows an increase above the current price, the more likely that I will buy the house now; and alternatively, if the forecast shows that house price will decline in the next period , then I won't buy a house now rather I will buy it in the future .

The key variables of interest are the average Sales Price of Houses Sold for the United States, property tax collected by the state and local government for United States and the delinquency rate on U.S consumer loans.

Result

19 out of 20 times house price is going to be between $351146.06 and $351145.94.

LOSS FUNCTION

The decision is to either buy a house now or buy a house in the future. The loss function for my model depends on both the direction of the forecast and the direction of the future house price. If I forecast an increase in the house price, I would buy the house now and if the future house price went down, then I will lose the extra money which I paid on the house now. On the other hand, if I forecast a decrease in the house price, I would buy the house in the future and if the future house price went up I am loosing the money which I would have gained of selling the house which I bought now, at a higher price. Thus, the loss function is determined by the decision I take and the actual change in the house price. If I forecasted an increase in house price and I bought a house now for $200000, I will lose $50000 if the future house price is decreasing $150000.I will also lose $50000 if I forecasted a decrease in house price and didn't buy a house for $200000 now and future house price went up to $250000. becauseI could have bought a house now for $200000 and still sell it for $250000 in the future.
Mathematically, loss function can be summarized in by the following indicator function:
Loss:

$-(D1(Pf_{t+1}>P_t)\ D2(Pt_{+1}> P_t)(P_{t+1}-P_t))+D1(Pf_{t+1}>P_t)\ D3(P_{t+1}<Pt)(P_t-P_{t+1})+D4(P_t>Pf_{t+1})D2(P_{t+1}>Pt)(P_{t+}1-Pt)+D4(P_t>Pf_{t+1})*0*D3(P_{t+1}<P_t)$

Pf_{t+1}=forecasted house price
pt=current house price
pt+1=future house price

CASE 1

Pf_{t+1}>pt and pt>pt+1

So D1=1

D3=1

Current house price	$200000
Forecasted price	$250000
Future price	$150000

=-(0)+1*1*(200000-150000)+0+0

=$50000

CASE 2 Pf_{t+1}<pt and pt<pt+1

So D4=1 and D2=1

Current house price	$200000
Forecasted price	$150000
Future price	$250000

=-(0)+0+1*1*(250000-200000)+0

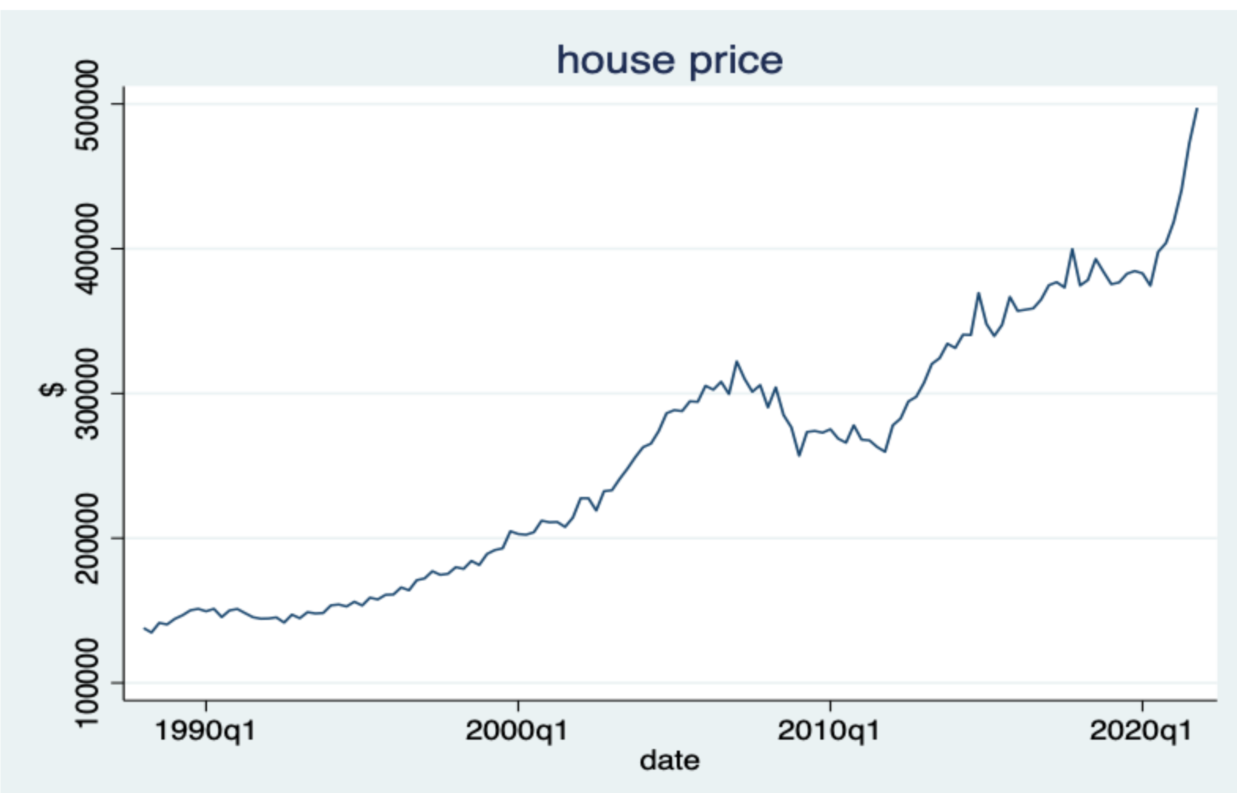
=$50000

Numerical example of loss function

	Future house price>current house price	Future house price<current house price
Buying house	0	$50000
Not buying house	$50000	$0

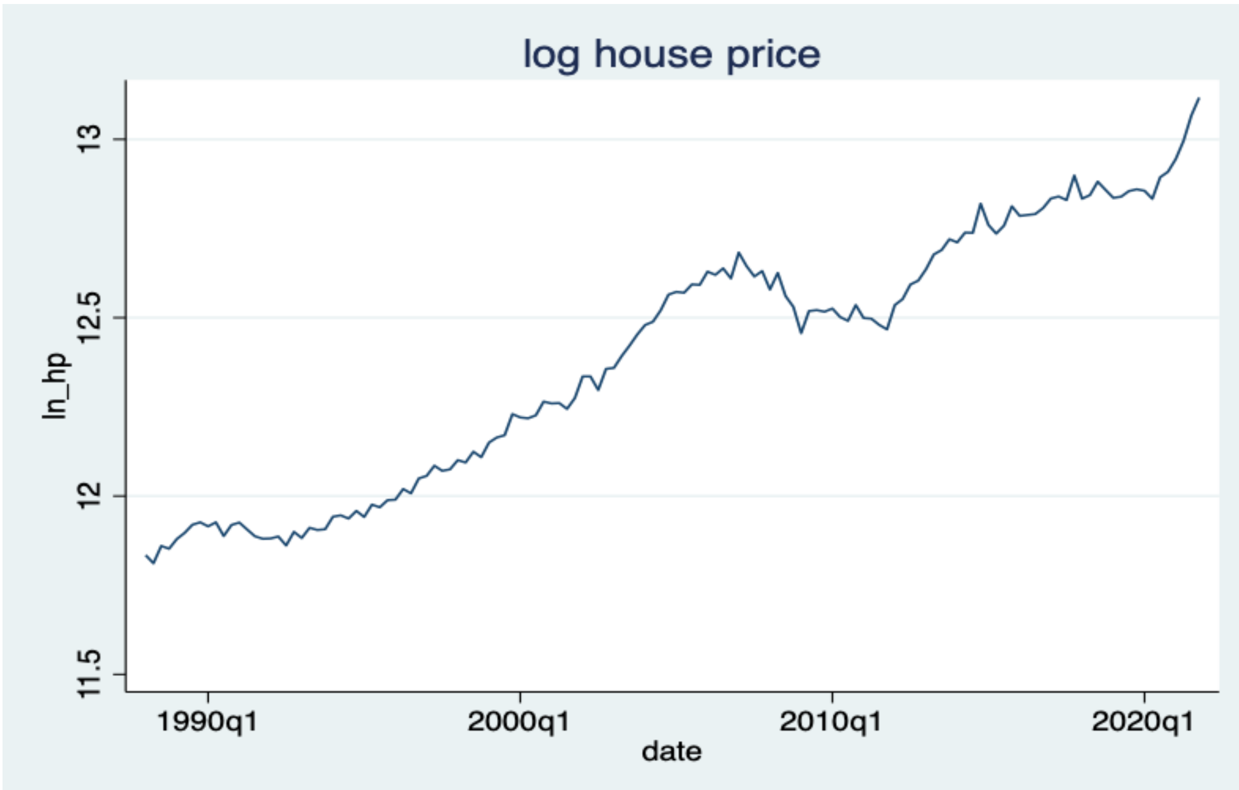
DATA &STATIONARITY&EXOGENEITY

*plotting the time series data of the house price



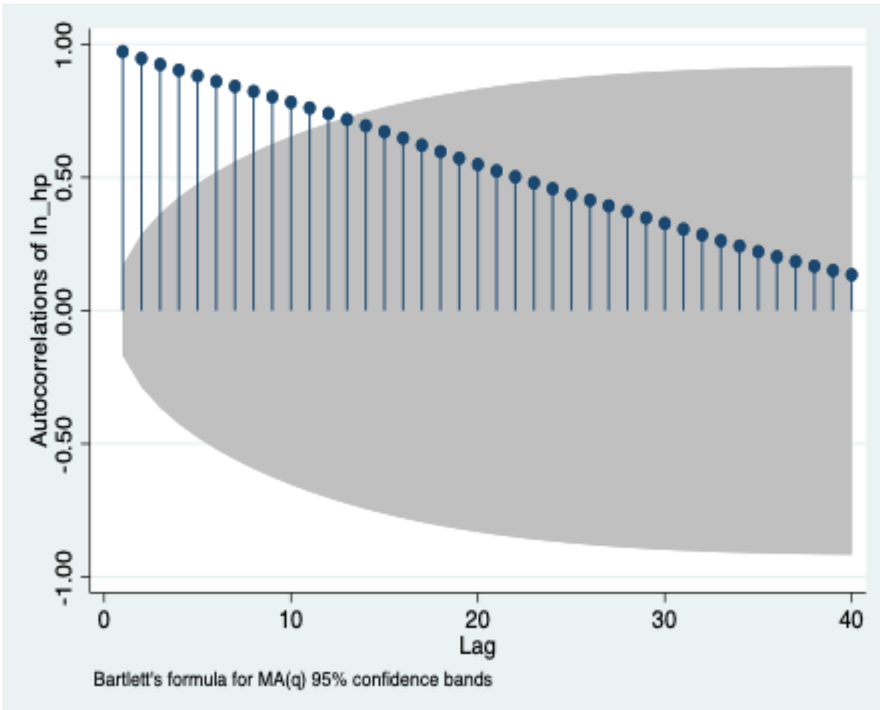
There is an exponential trend in the data, it is a hint of non-stationarity. So use log transformation.

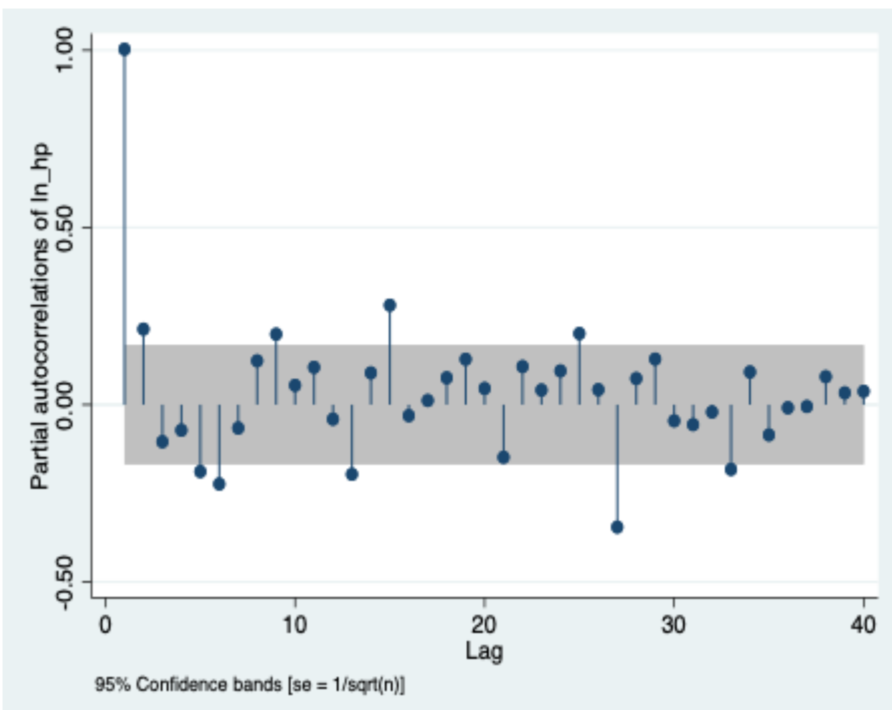
*log transformation of house price data



After the log transformation, still there is exponential trend in the data.

* Analyzing autocorrelation and partial autocorrelation of house price





ac declines very slowly => possibly non-stationary

pac: first partial autocorrelation is equal to 1 => possibly non-stationary

*Done unit root test for house price to check for stationarity.

Since there is trend, we need ADF test with trend and drift.

Lag selection: lags after 5 are not significant. So dropped them.

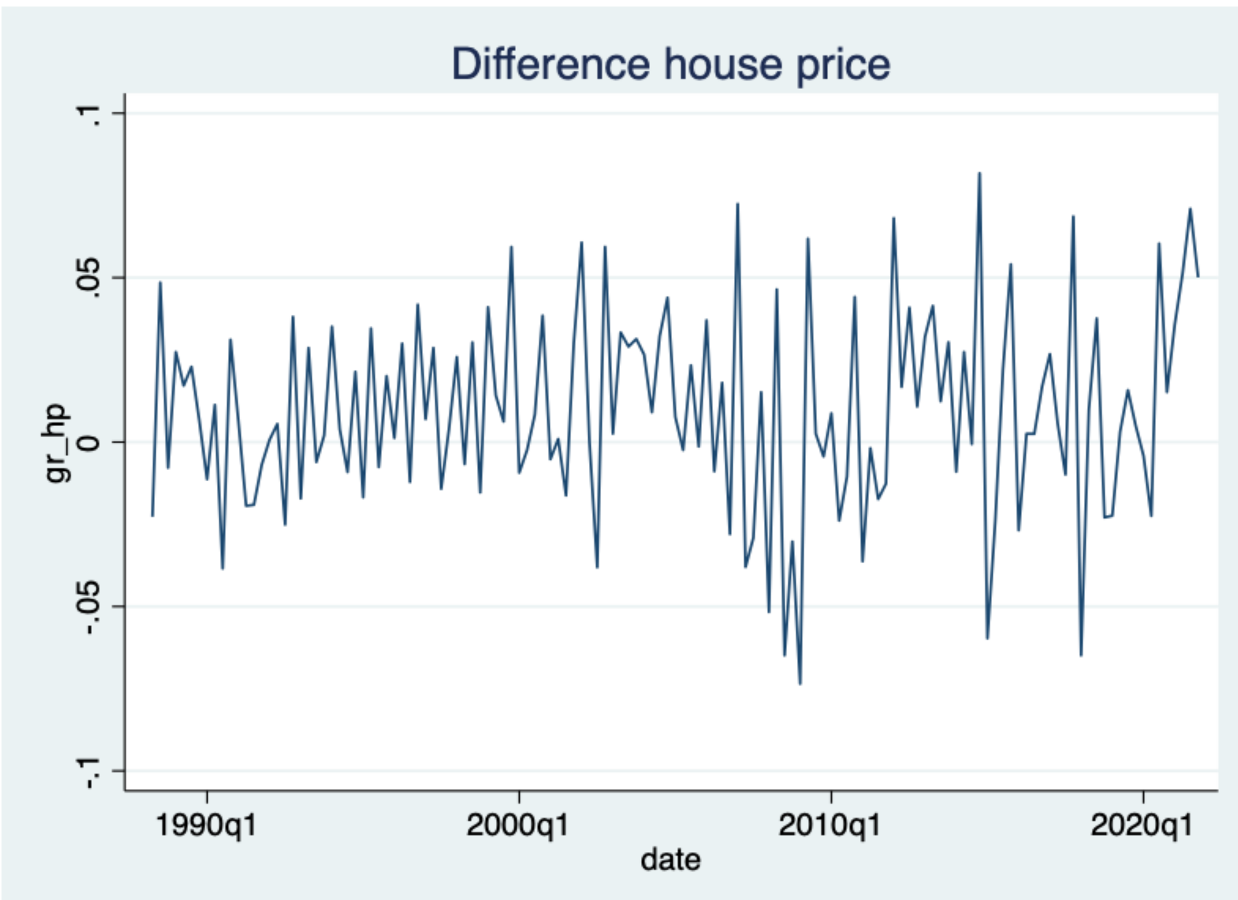
Result: cannot reject unit-root at 5% significance level.

```
Test statistic    P>|t|
-2.944           -3.446
```

evidence of non-stationarity, so taking the difference of the data.

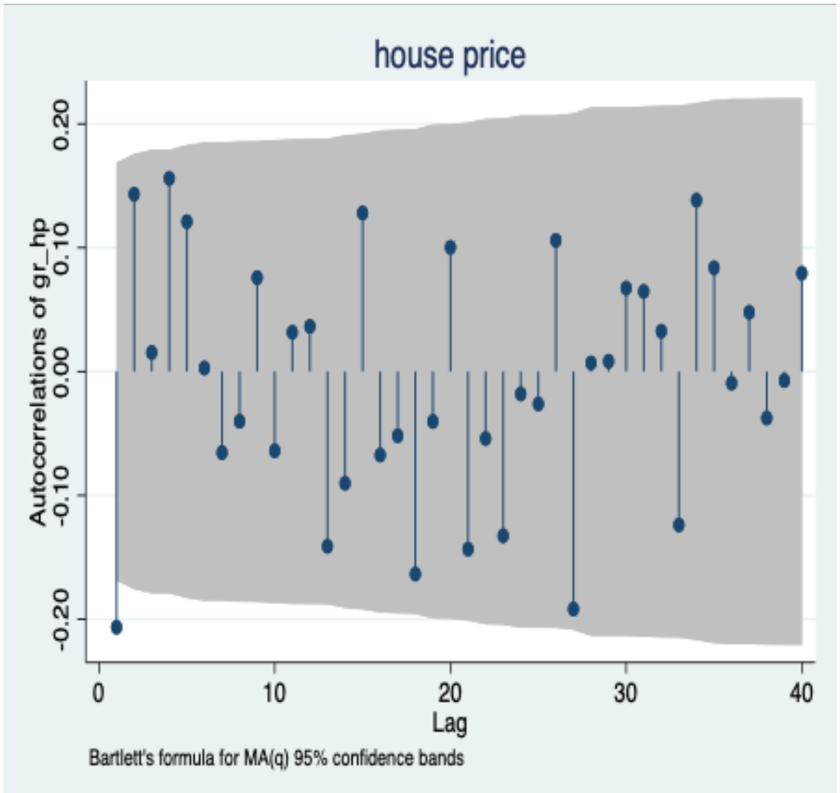
(APPENDIX page1-page2)

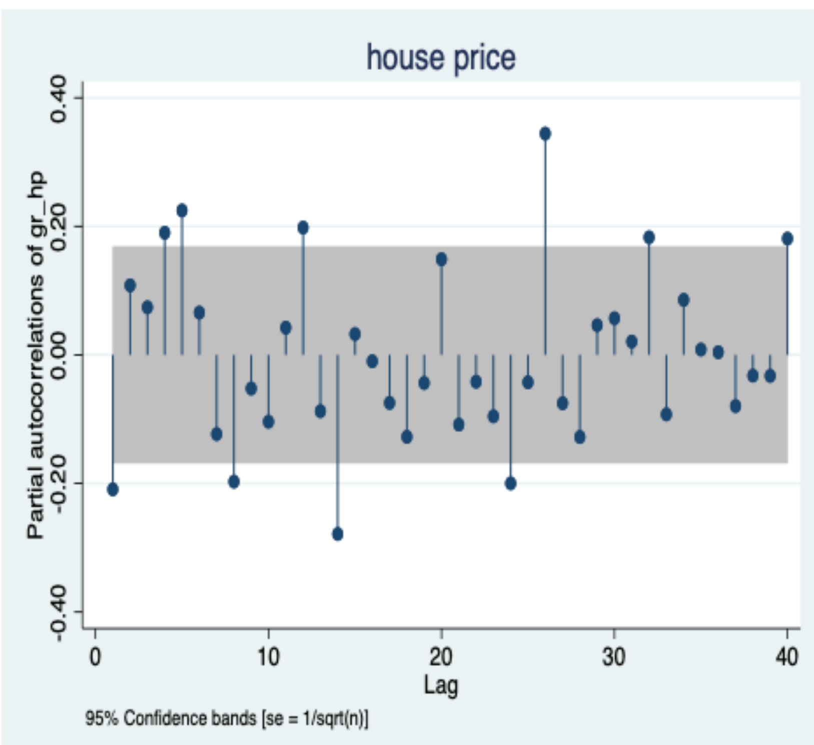
*plotting the data of Growth rate of house price



process is not as slow as it was. Data can be put it in a box. Most likely stationary.

* analyzing autocorrelation and partial autocorrelation of growth rate of house price





Quick exponential decline. both ac and pac decline quite fast, pac(1) is not 1.So both ac and pac suggest stationarity.

*Confirmed that transformed variable is stationary using unit root test.

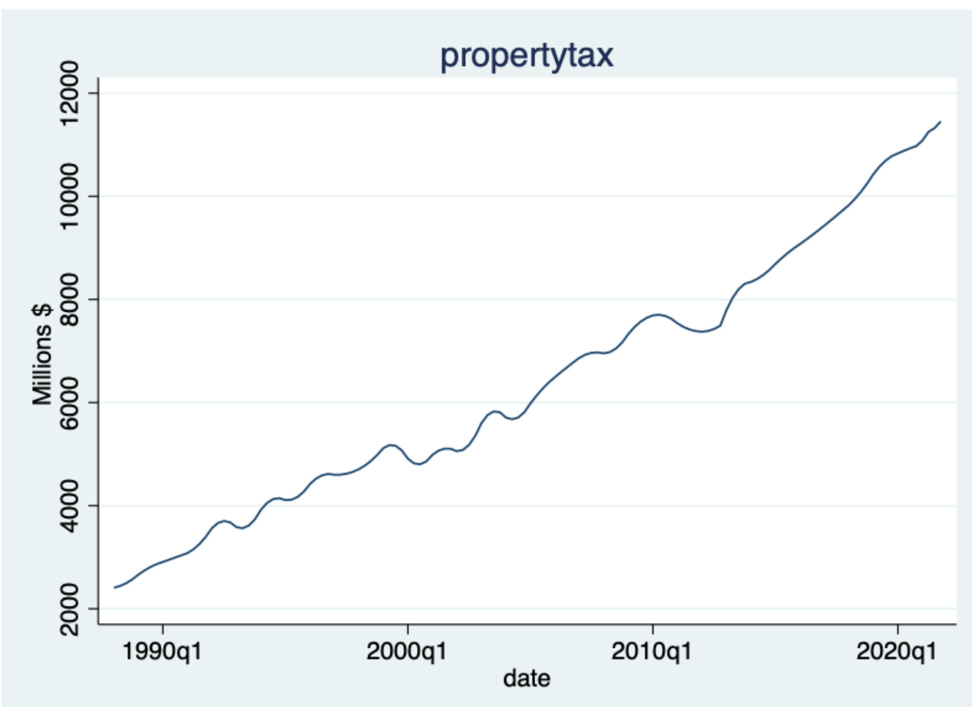
```
Test statistic    critical value
-7.716            -3.446
```

Lag selection:lags after 1 are not significant.So dropped them.

Result:can reject unitroot at 5% significance level against the alternative of stationarity.

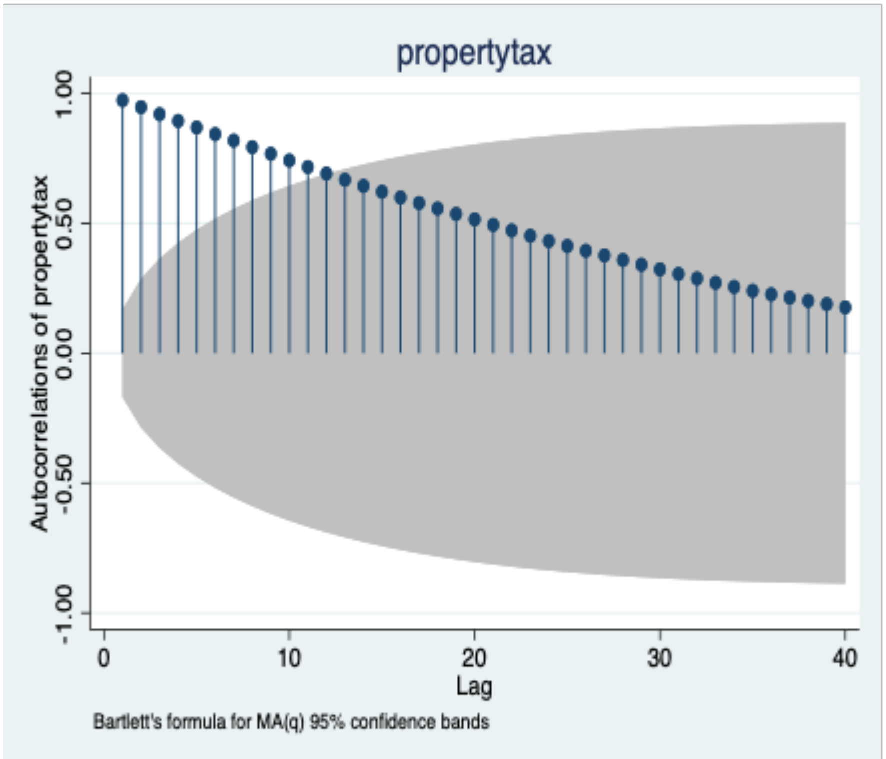
(APPENDIX page1-page2)

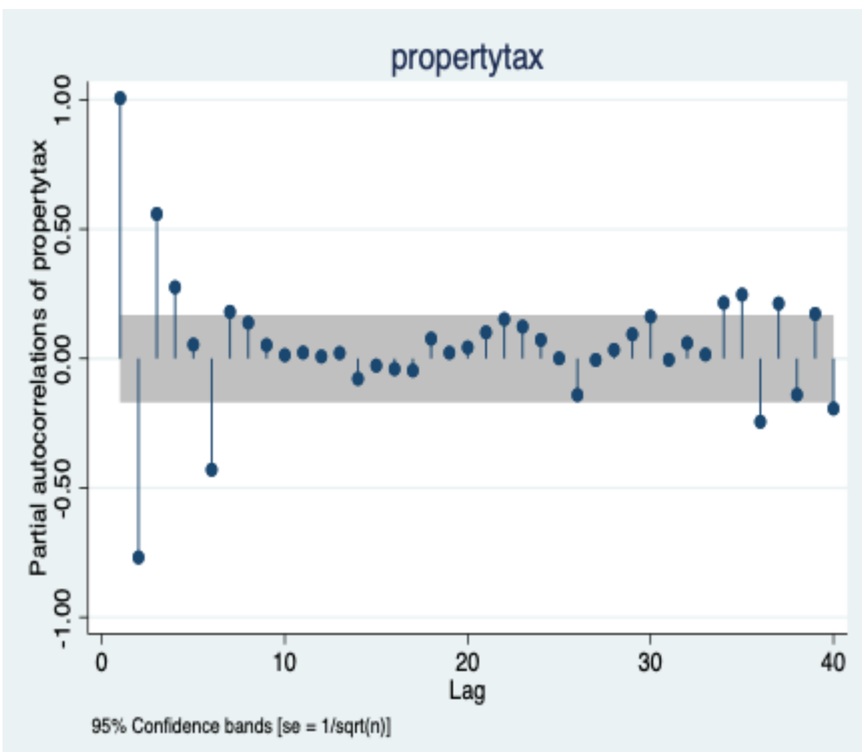
*plotting the time series data of the property tax



There is a linear trend in the data ~likely non-stationary

* analyzing autocorrelation and partial autocorrelation of property tax





ac declines very slowly => possibly non-stationary

pac: first partical autocorrelation very close to 1 => possibly non-stationary.

*Since there is evidence of stationary, Testing for unit root.

Since there is a linear trend, we need ADF test with the trend and drift.

Selecting the lags:lags after 5 are not significant. So dropped them.

Result: cannot reject unit-root at 5% significance level.

Test statistic	critical value
-1.250	-3.446critical

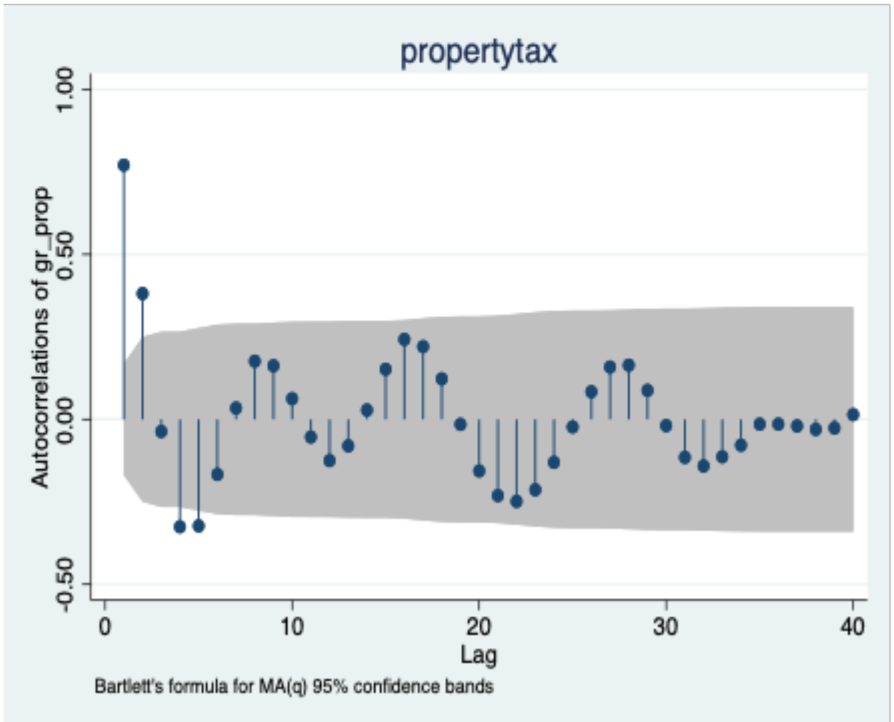
(APPENDIX page3-page5)

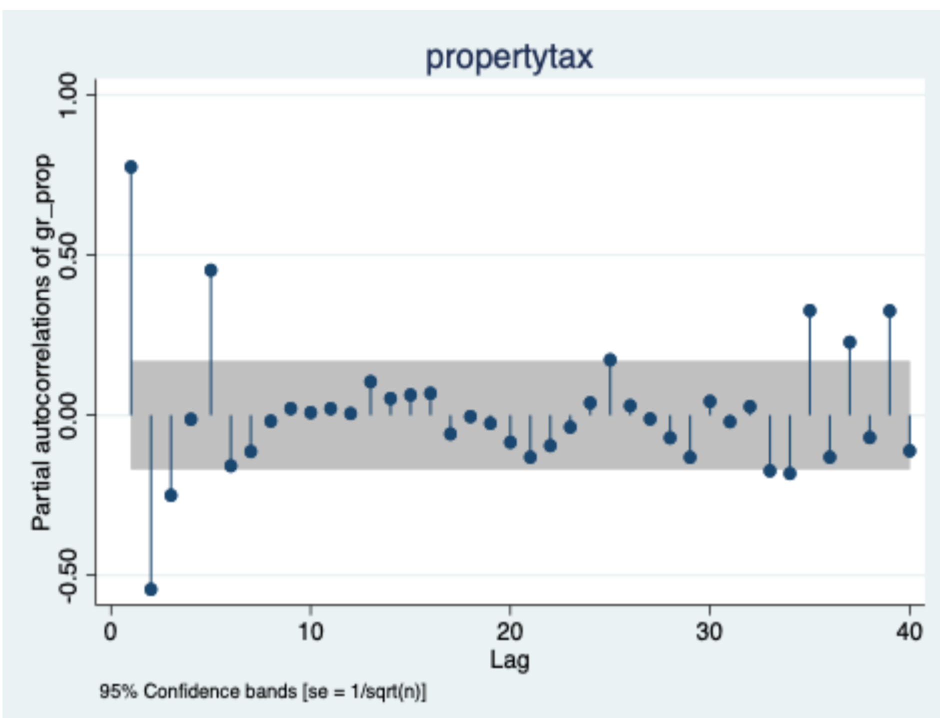
Still evidence of non-stationarity, so taking the difference of the data.

*plotting the differenced data of property tax

There is no trend in the data.

*analyzing autocorrelation and partial autocorrelation of differenced data of property tax





Quick exponential decline.

both ac and pac decline quite fast, pac(1) is not ~1.So both ac and pac suggest stationarity.

*Confirmed that transformed variable is stationary using unit root test.

Lag selection:

lags after 4 are not significant. So dropped them. .

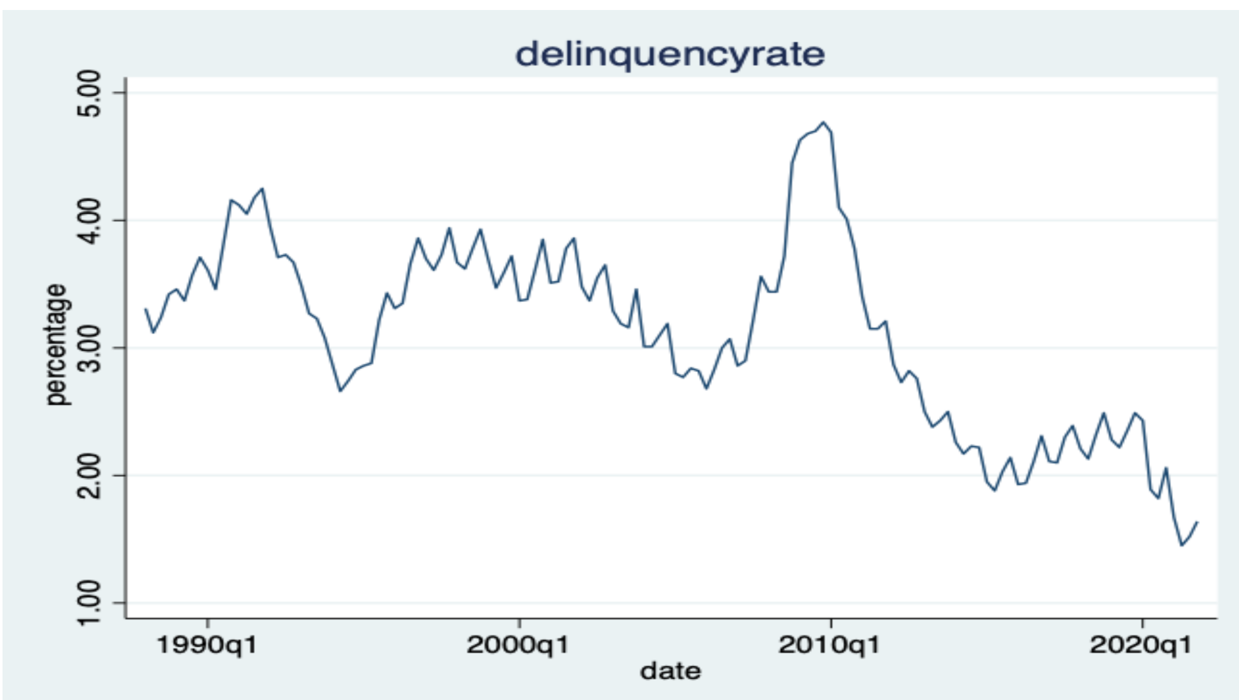
Can reject unit root at 5%significance level

Test statistic Critical value

```
-3.706          -3.446
```

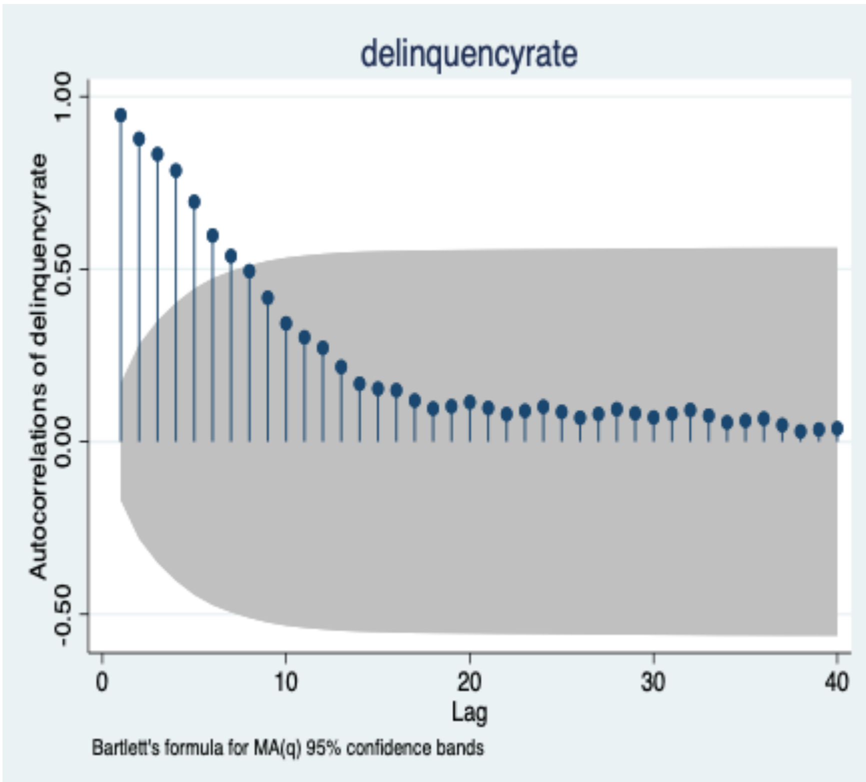
(APPENDIX page3-page5)

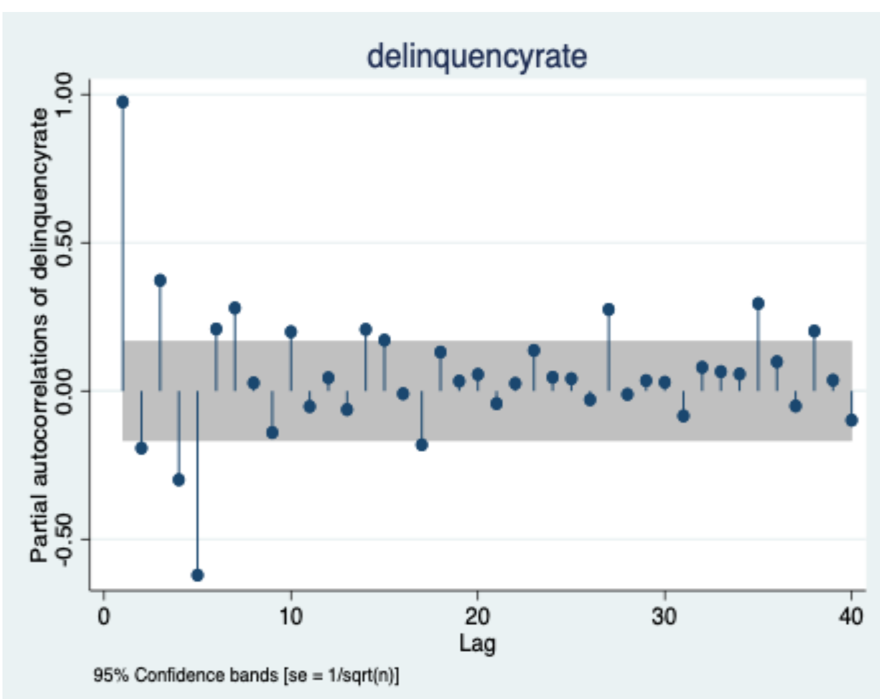
*Plotting the time series data of Delinquency rate



There is no trend. Data can be put in a horizontal box but move slowly. Possibly non-stationary.

*analyzing autocorrelation and partial autocorrelation of Delinquency rate





ac declines very slowly => possibly non-stationary

pac: first partical autocorrelation very close to 1 => possibly non-stationary Testing for unit root.

*Unit root test

Since there is no trend, we need ADF test with no constant.

Selecting the lags:

lags after 6are not significant.So dropped them.

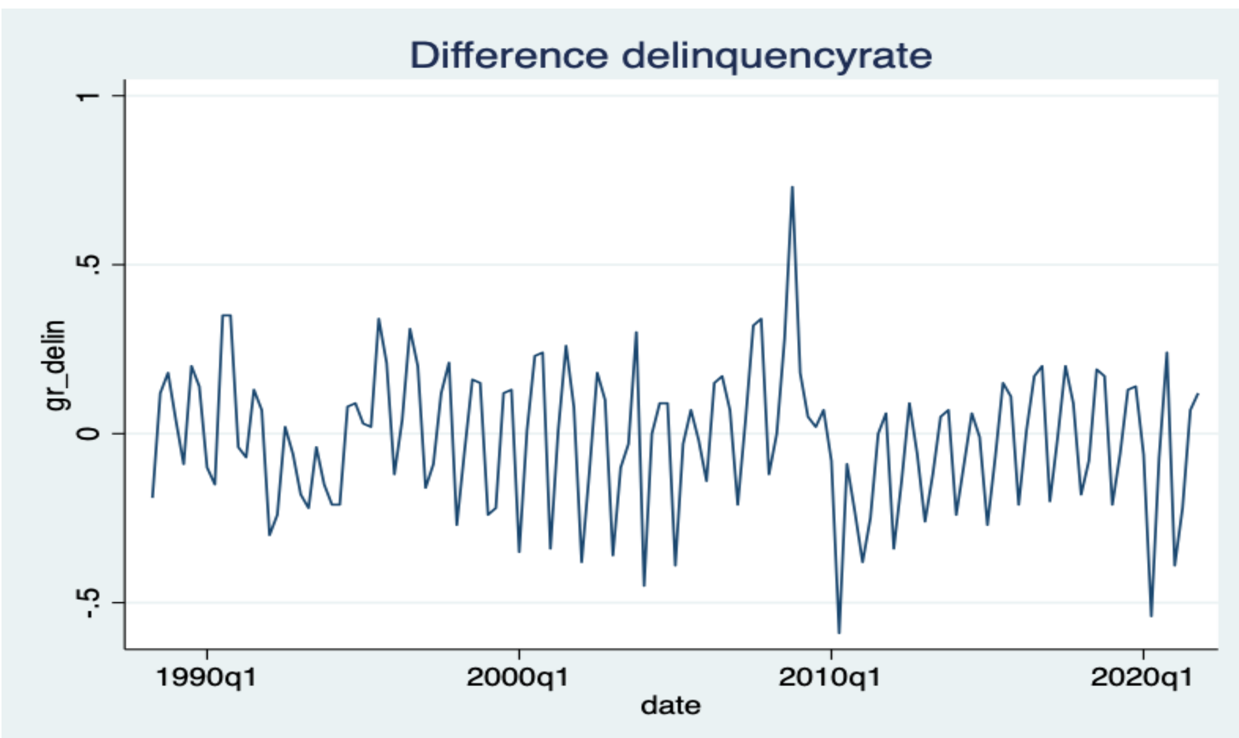
Result:cannot reject unitroot at 5% significance level.

(APPENDIX page5-page7)

Test-statistic	critical value
-1.082	-1.950

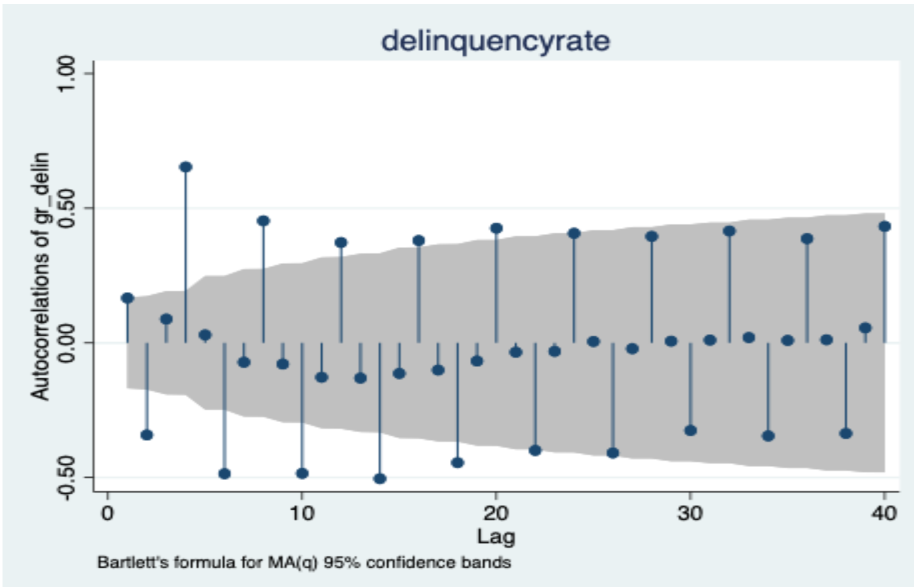
Still evidence of non-stationarity, so taking the difference of the data

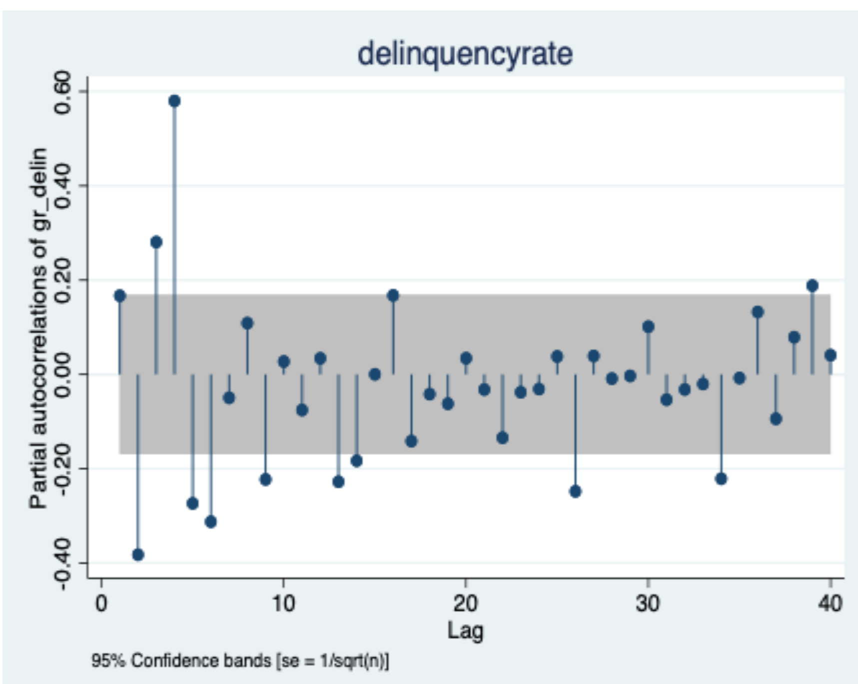
*plotting the first difference data of Delinquency rate



Data is frequently moving. Crosses mean often , suggesting stationary

*analyzing autocorrelation and partial autocorrelation of differenced Delinquency rate





Quick exponential decline.

both ac and pac decline quite fast, pac(1) is not ~1.So both ac and pac suggest stationarity.

*confirming that transformed variable is stationary using unit root..

Lag selection:

lags after 5 are not significant. So dropped them.

Can reject unit root at 5%significance level

Test statistic	Critical value
-4.403	-1.950

(APPENDIX page5-page7)

6.SUPPORTING VARIABLES AND MAIN VARIABLE

My supporting variables are property tax and the delinquency rate of consumer loans. Both these variables help me to forecast house prices. When the Govt keeps on increasing the property tax it is less likely that people will buy property. Then the demand for houses will decline which in turn increases the house price. In the same way as the delinquency rate on consumer loans increase, banks stop providing mortgage loans to people, which will reduce the demand for houses in turn increases the house price.

7.TEST FOR EXOGENEITY USING GRANGER CASUALITY TEST

Granger causality Wald tests

Equation	Excluded	chi2	df	Prob > chi2
gr_hp	gr_pt	3.8379	6	0.699
gr_hp	gr_del	4.3196	6	0.634
gr_hp	ALL	7.8895	12	0.794
gr_pt	gr_hp	9.5051	6	0.147
gr_pt	gr_del	13.827	6	0.032
gr_pt	ALL	19.54	12	0.076

```
|--------------------------------------+--------------------------|
|          gr_del              gr_hp |  18.782     6      0.005  |
|          gr_del              gr_pt |  1.8836     6      0.930  |
|          gr_del              ALL |  20.615 12     0.056  |
+-------------------------------------------------------------------+
```

From this table, both gr_pt and gr_del can be excluded from the gr_hp equation as the p-value is not significant. This means both gr_pt and gr_del do not cause gr_hp.

gr_hp can be excluded from the gr_pt equation as the p-value is not significant, but gr_del can't be excluded from the gr_pt equation as the p-value is significant This means gr_hp does not cause gr_pt but gr_del causes gr_pt.

gr_hp cannot be excluded from the gr_del equation since the pvalue is significant,so gr_hp causes gr_del whereas gr_pt can be excluded from the gr_del equation as the p value is not significant.

The result shows that gr_del is connected to gr_pt and gr_hp is connected to gr_del.

FORECAST

1.Table that summarizes univariate and multivariate forcasts.

UNIVARIATE MODELS				
MODEL	LOSS	DM	P VALUE	RESIDUALS
Arma(0,0)	-5529.63			0.0370
Arma(0,1)	-2166.667	1.06	0.301	0.0156

Arma(1,0)	-2166.667	1.06	0.301	0.0020
Arma(1,1)	-1862.963	1.29	0.208	0.0018
Arma(2,0)	-3503.704	0.98	0.336	0.0003
Arma(0,2)	-3466.667	0.80	0.432	0.0004
Arma(2,2)	-1185.185	1.34	0.192	0.3708
Arma(3,0)	-4192.593	0.62	0.539	0.0001
MULTIVARIATE MODELS				
VAR(1/3)	-659.2593	1.56	0.130	
VAR(1/4)	-4488.889	0.66	0.514	
VAR(1/5)	-2992.593	1.15	0.260	

Residuals result (APPENDIX page7-page8)

2.

Considering the loss function, arma(0,0) is my best model because it gives the smallest loss even though the residuals are not white noise.

3. Looking at the Diebold Marino test results for all models, test statistic>critical value means we have to accept the null that all models are statistically the same as

the best model which means my best model is not statistically better than other forecasting models.
(APPENDIX page9-page11)

4. According to my best forecasting model arma(0,0) the prediction interval is
upper 95% bound =351146.06
lower 95% bound =351145.94
this can be interpreted as 19 out of 20 times house price is going to be between \$351146.06 and \$351145.94.
(APPENDIX page12)

5.My best model is arma(0,0) so there is nothing to compare.

6.Loss interval

Let the actual future value be \$351146.Upper bound forecasted with my best model is \$351146.06 and take current house price as \$351146.02.
So this is a case of fp>pt and pt>pt+1
So D1=1 and D3=1

$=-(D1(Pf_{t+1}>P_t)\ D2(Pt_{+1}> P_t)(\ P_{t+1-}\ P_t))+\ D1(Pf_{t+1}>P_t)\ D3(\ P_{t+1}<Pt)(\ P_t\ -P_{t+1})+\ D4(\ P_t >Pf_{t+1})D2(P_{t+1}>Pt\ )(\ P_{t+}1-Pt\ )+\ D4(\ P_t> Pf_{t+1})*0*D3(\ P_{t+1}< P_t\ )$

=-(1*0*(351146-351146.02))+1*1*(351146.02-351146)+0*0*(351146-351146.02)+0*1*0
=0+.02+0+0
=.02

Let the actual future value be $351148.Upper bound forecasted with my best model is $351146.06 and take current house price as $351147

So this is a case of pt>fp and pt+1>pt

So D4=1 and D2=1

=-(0*1*(351148-351147))+0*0*(351147-351148)+1*1*(351148-351147)+1*0*0

=0+0+1+0

=1

Let the actual future value be $351145.Lower bound forecasted with my best model is $351145.94and take current house price as $351145.16

So this is a case of fp>pt and pt>pt+1

So D1=1 and D3=1

=-(1*0*(351145-351145.16))+1*1*(351145.16-351145)+0*0*(351145-351145.16)+0*1*0

= .16

Let the actual future value be $351146.Lower bound forecasted with my best model is $351145.94and and take current house price as $351145.98

So this is a case of pt>fp and pt+1>pt

So D4=1 and D2=1

=-(0*1*(351146-351145.98))+0*0*(351145.98-351146)+1*1*(351146-351145.98)+1*0*0

=0.02