Individual Project

1 Introduction

Although the price change of any particular house is difficult to predict, the overall house price change is predictable. Over the past few decades, academic interest in real estate market modeling and forecasting has expanded (Case and Shiller, 1990; Zhou, 1997; Barras, 2009; Brooks and Tsolacos, 2010).

This project compares the forecasting performance of two types of time series models: ARIMA, Holt's Linear Trend Model and compared them with baseline models. In conclusion, ARIMA model performs better in out-of-sample forecasting.

2 Data Analysis

The time series I'm going to analyze is S&P/Case-Shiller MA-Boston Home Price Index. Since it is a seasonally adjusted time series, there's no seasonality in this time series. Figure 1 is the raw time series plot.

S&P/Case-Shiller MA-Boston Home Price Index

Source: S&P Dow Jones Indices LLC

Release: S&P/Case-Shiller Home Price Indices

Units: Index Jan 2000=100, Seasonally Adjusted

Frequency: Monthly

Length: From 1987-01 to 2020-01

Figure 2 is the ACF plot and figure 3 is the partial ACF plot. From figure 2 and figure 3, we can see that ACF decays slowly while PACF drops rapidly. Note that the PACF plot has a significant spike only at lag 1. We can see a very obvious trend in this seasonally adjusted data.

3 Model Selection

3.1 Splitting the data

For this topic, I care about a long term trend of Boston Home Price Index. When dealing with these topics, a train/validation structure is better, and I can predict for a period of time. So I chose to use a train/validation structure.

I split the whole time series into two parts: Training part (From 1987-01 to 2013-12) and validation part (From 2014-01 to 2020-01).

3.2 ARIMA model

ARIMA is a very popular technique for time series modelling. It describes the correlation between data points and takes into account the difference of the values.

First of all, I ran a Dickey-Fuller test to see whether the time series is stationary or not. Since the null hypothesis cannot be rejected, so random walk plus drift exists, which make sense given that there is a growth rate. So I add a first difference and use an ARIMA model. The model with only one order of differencing assumes a constant average trend--it is essentially a fine-tuned random walk model with growth--and it, therefore, makes relatively conservative trend projections.

I use the auto-arima function with the BIC to find the best p and q in the ARIMA model. Also, I plot a linear trend for comparison. The RMSE is 0.482 for training set and 7.943 for validation set. Figure 4 shows the predicted result of ARIMA(d=1) model.

3.3 Holt's Linear Trend Model:

This method takes into account the trend of the dataset. Since our time series has an obvious trend, I tried Holt's linear trend model. Empirical evidence indicates that these methods tend to over-forecast, especially for longer forecast horizons. Motivated by this observation, Gardner & McKenzie (1985) introduced a parameter that "dampens" the trend to a flat line sometime in the future. The RMSE is 0.540 for training set and 23.940 for validation set. Figure 5 shows the damped holt's method and its predicted result.

3.4 Other Baseline Models:

Finally, we use some baseline forecasting models to see how our models work and to find out the best model.

Since this time series has an obvious trend and with no seasonality, it's reasonable to have a try on linear forecast. The RMSE is 0.540 for training set and 23.940 for validation set. Figure 6 shows linear regression model and its prediction result.

I also use some simple model such as naïve method, drift method which is shown in figure 7. The blue line represents naïve method, and the red line represents drift method.

After that, I created a table and compared all the validation RMSEs. So it's clear that ARIMA(d=1) is the best model to forecast MA-Boston Home Price Index with a RMSE only 7.943.

Model	Validation_RMSE
ARIMA (d=1)	7.943
Linear	9.091
Drift	18.924
Holt filter	23.940
Naive	32.034

Table 1 Results of validation RMSEs

3.5 Diebold/Mariano test:

Here we use Diebold/Mariano test to verify whether ARIMA is the best model. I compared ARIMA with all other models, and Table 2 shows the p-value of these Diebold/Mariano tests.

Model	p.value
Holt filter	0.0013903
Drift	0.0000000
Naive	0.0000000
Linear	0.0000000

Table 2 Results of Diebold/Mariano (ARIMA vs Model)

As shown in the table, p-values are all very small, so we can say that it is statistically significant that ARIMA model outperforms all other models.

4 Conclusion

From the data we get, we can say that ARIMA is the best model. It has the lowest validation RMSE, and according to Diebold/Mariano tests, it outperforms all other models.

As the results of the study suggest, ARIMA is a useful technique to assess broad market price changes. Government and central bank can use ARIMA modelling approach to forecast national house price inflation. Developers can employ this methodology to drive successful house-building programme. Investor can incorporate forecasts from ARIMA models into investment strategy for timing purposes.

Certainly, there are number of limitations attached to this particular modelling approach. Firm predictions about house price movements are also a challenge, as well as more research needs to be done in establishing a dynamic interrelationship between macro variables and the Boston housing market.

Although the research focused on Boston, the findings extend to the domestic housing market. ARIMA house price modelling provides insights for a spectrum of stakeholders. The use of this modelling approach can be employed to improve monetary policy oversight, facilitate planning for infrastructure or social housing as a countercyclical policy and mitigate risk for investors.

References

Hyndman R J, Athanasopoulos G. Forecasting: principles and practice[M]. OTexts, 2018.

Jadevicius A, Huston S. ARIMA modelling of Lithuanian house price index[J]. International Journal of Housing Markets and Analysis, 2015.

Crawford G W, Fratantoni M C. Assessing the forecasting performance of regime-switching, ARIMA and GARCH models of house prices[J]. Real Estate Economics, 2003, 31(2): 223-243.

Stevenson S. A comparison of the forecasting ability of ARIMA models[J]. Journal of Property Investment & Finance, 2007.

Appendix

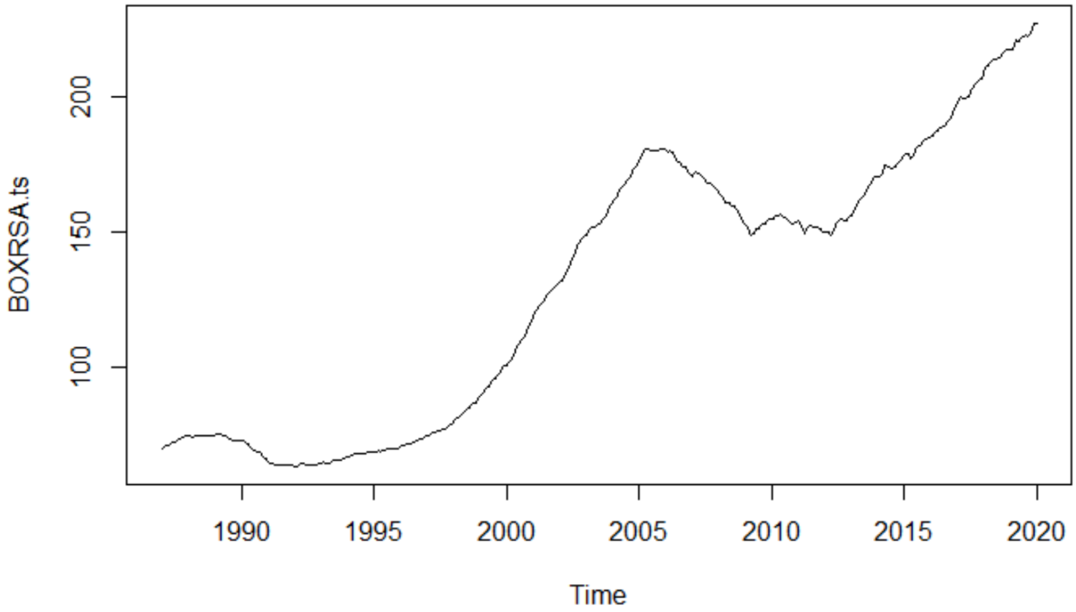

Figure 1 Simple time plot

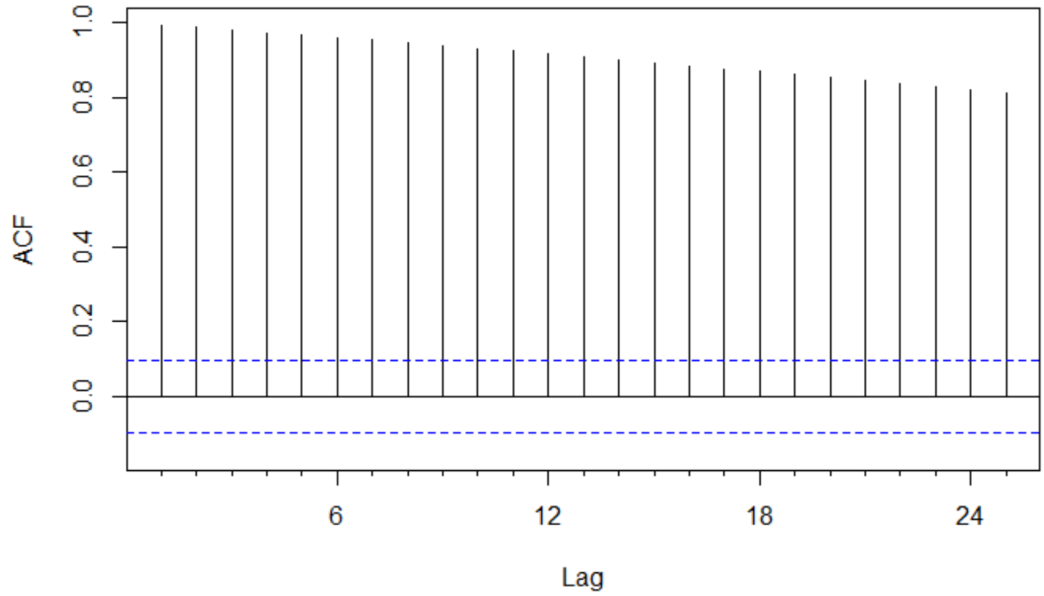

Figure 2 ACF plot

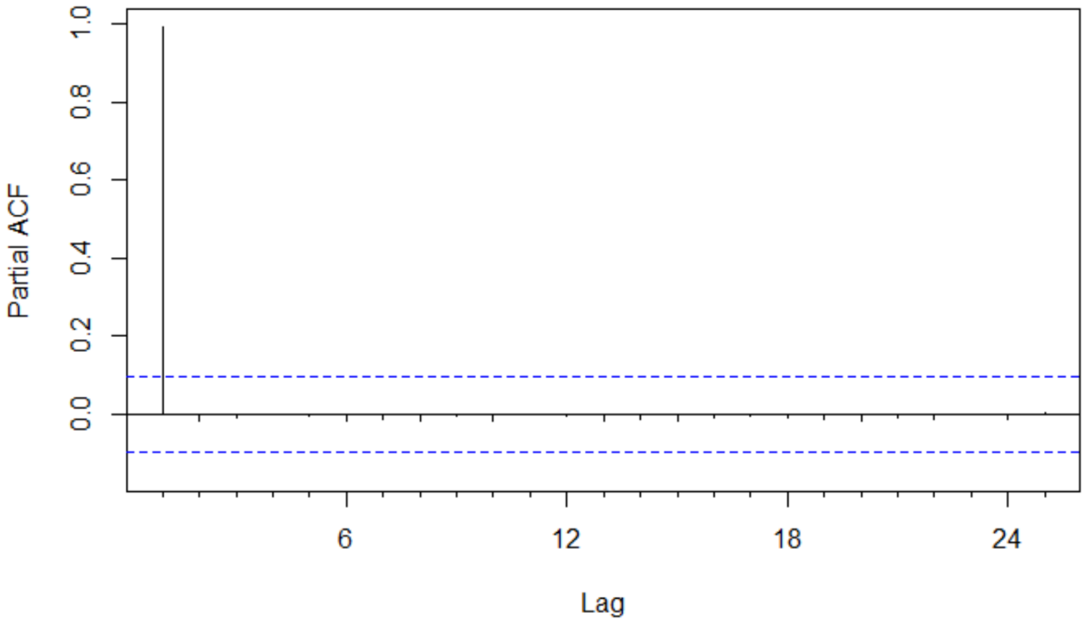

Figure 3 Partial ACF plot

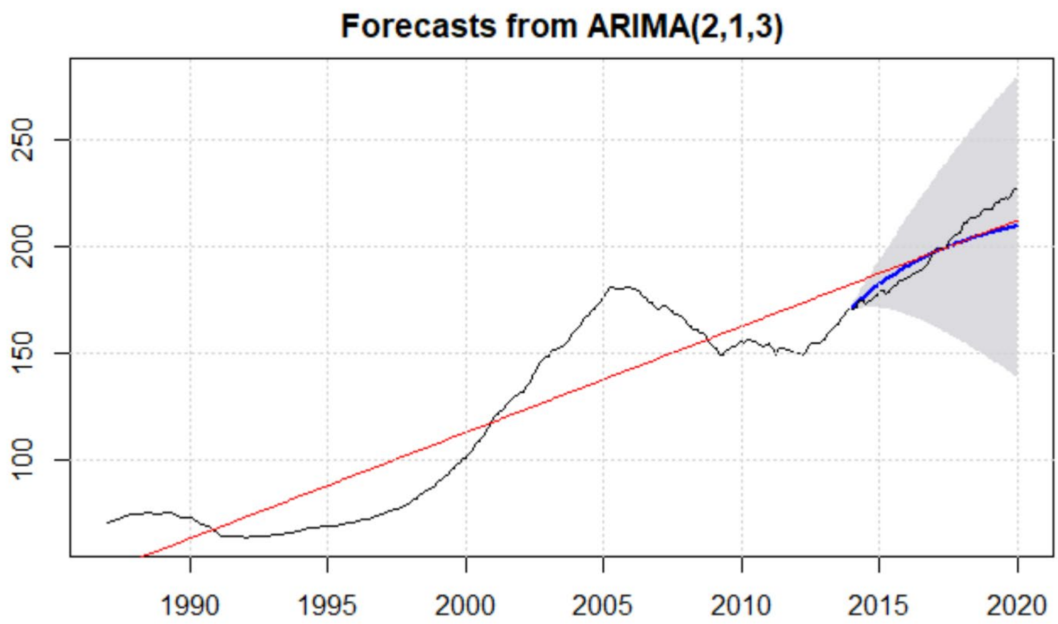

Figure 4 ARIMA(d=1) model

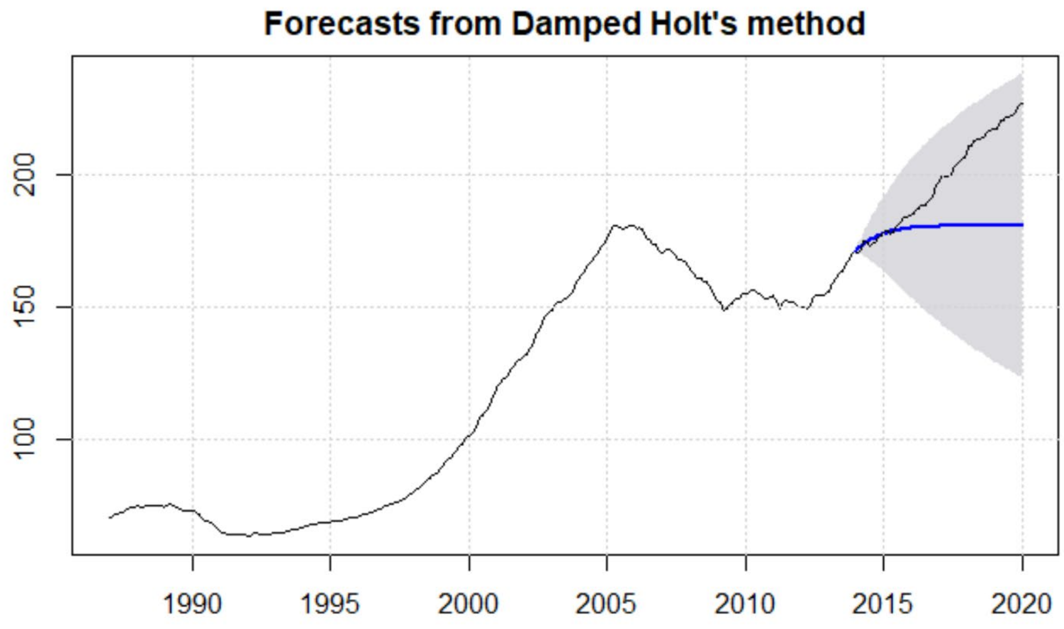

Figure 5 Damped Holt's method

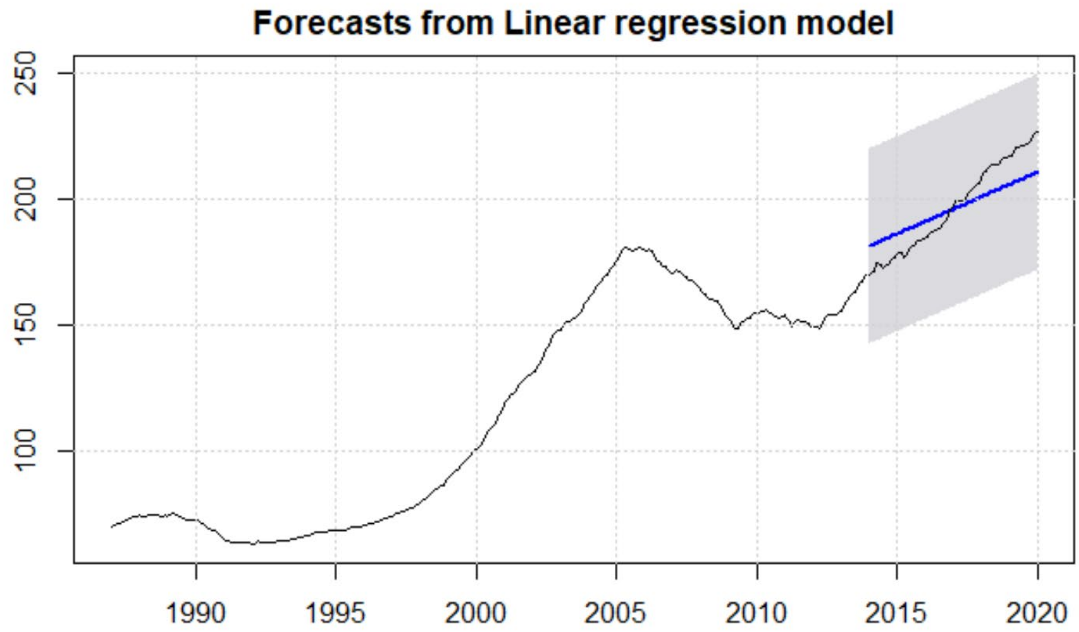

Figure 6 Linear regression method

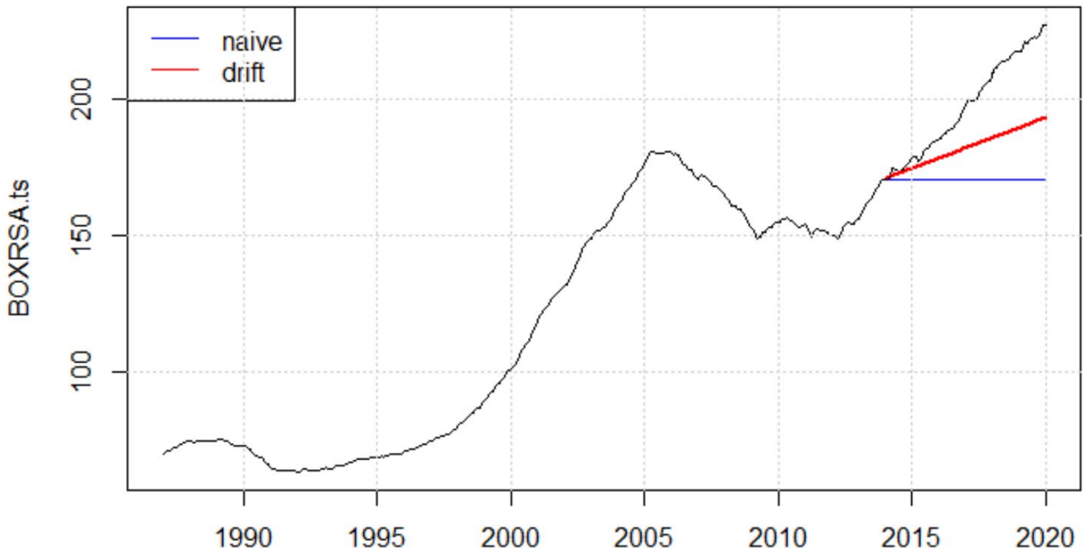

Figure 7 Other baseline methods