Housing Prices

DS 6373 Time Series Final Project
Aaron Abromowitz and Alex Thibeaux

Video Presentation of this Deck:
https://youtu.be/Yl1-x-Ha8Xw

INTRODUCTION

Accurate prediction of housing prices is important for homeowners, potential homeowners, and investors

OUR TEAM

AARON ABROMOWITZ

Systems Engineer, Raytheon

ALEXANDRA THIBEAUX

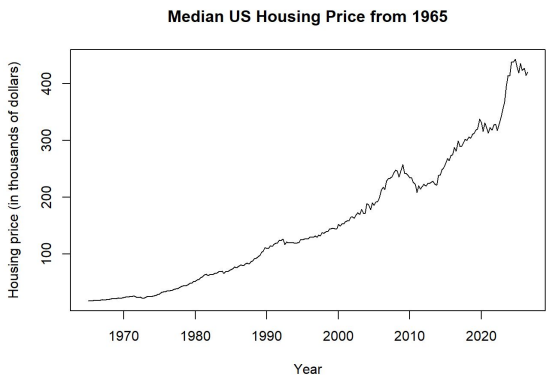
Data Analyst, Southern Methodist University

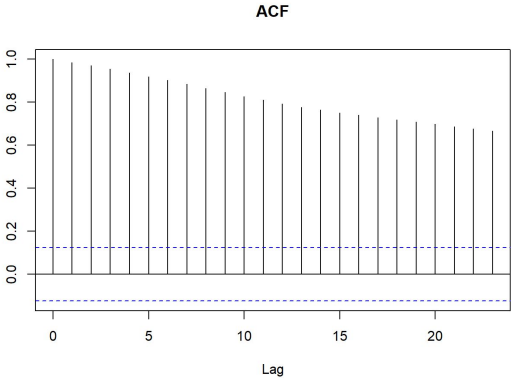
TABLE OF CONTENTS

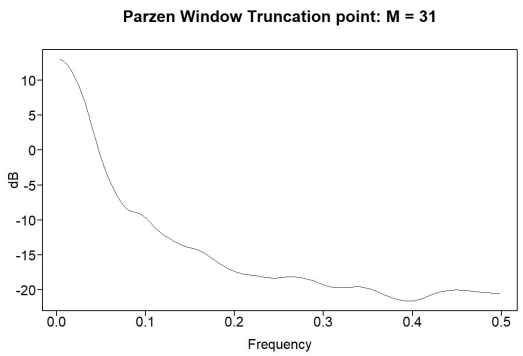
Variable of Interest: Median Housing Sale Price

- Data points are for quarter year
- Realization plot shows that Median housing price increasing
- ACF shows slowly dampening autocorrelations
- Parzen Window shows a frequency at 0

- Evidence of non-stationarity
- Median housing prices changes with time
- Variation of median housing price changes with time

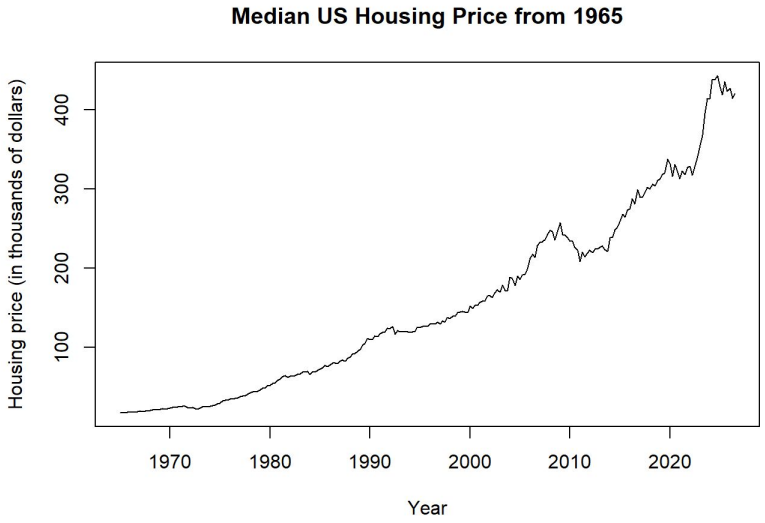


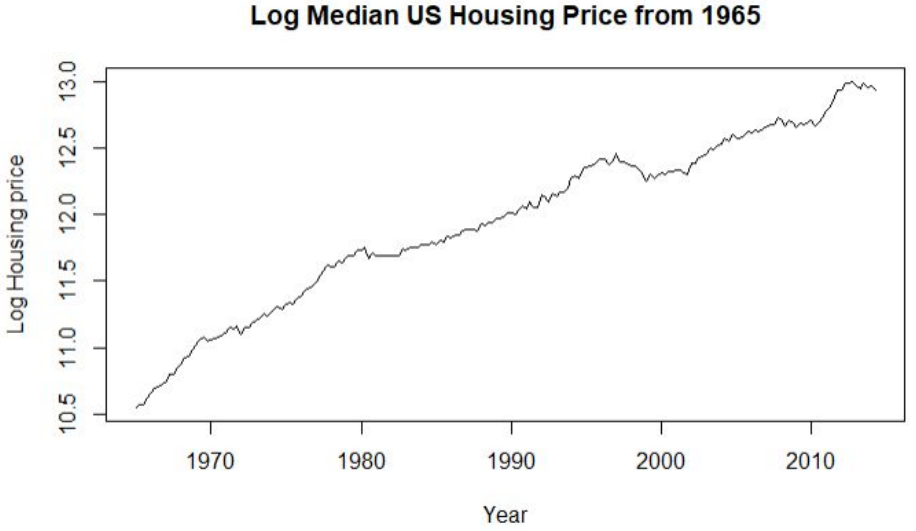




Log of Mean Housing Price

- Log smoothed out variation
- Used log Median Housing Price in models





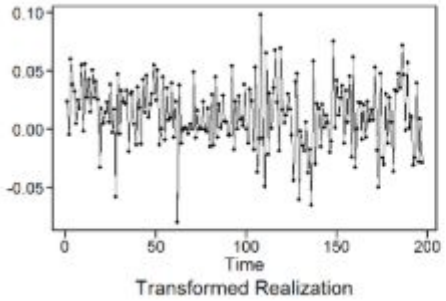
Evaluation Metrics

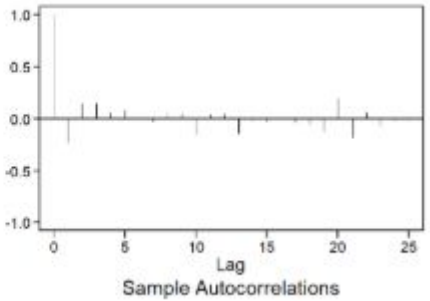
- rwRMSE (rolling window root mean squared error, log Median Housing Price)
- ASE (average square error, Median Housing Price)
- Short term forecast (1 year / 4 quarters)
- Long term forecast (5 years / 20 quarters)

Short term rwRMSE	
Long term rwRMSE	
Short term ASE	
Long term ASE	

ARIMA Model Selection

- Difference Filter removes the low frequency
- ARIMA(1,1,2) model is chosen





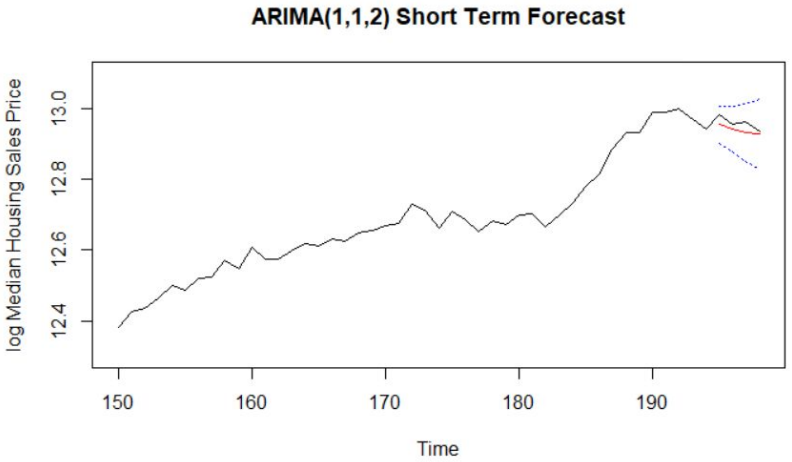
p	q	aic
1	2	-7.188559
2	2	-7.180263
5	2	-7.174131
3	2	-7.170135
6	2	-7.167230

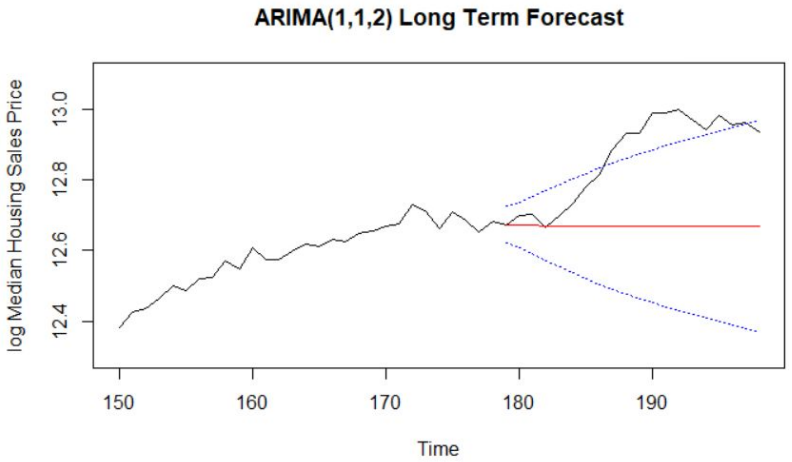
$$(1 - 0.69B) * (1-B) * (\log(X(t)) - 0.012) = (1 - 0.92B + 0.4B^2) * a(t)$$

$$\text{var}(a(t)) = 0.0007$$

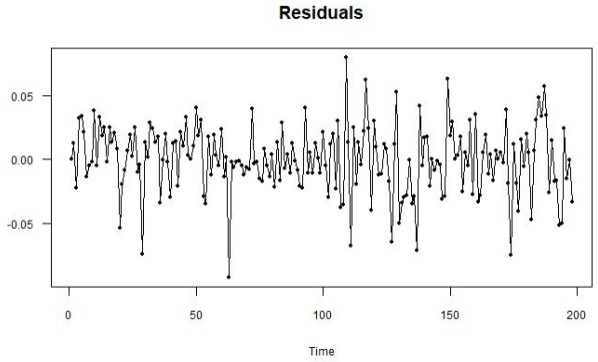
ARIMA Model Metrics

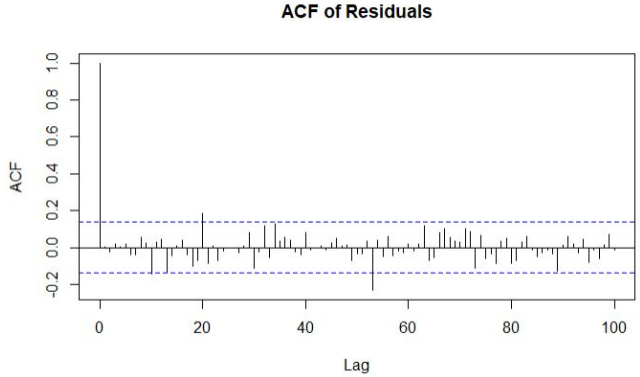
Short term rwRMSE	0.037
Long term rwRMSE	0.133
Short term ASE	84.4 M
Long term ASE	7.61 B

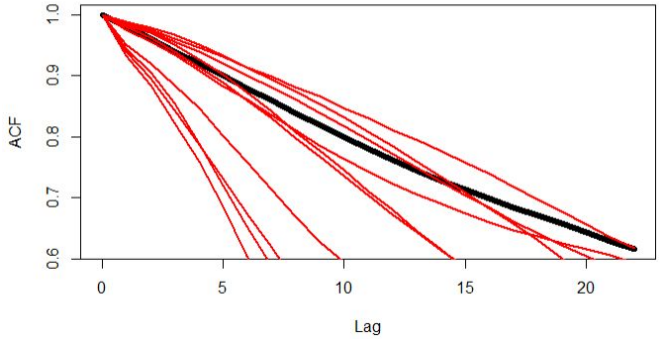


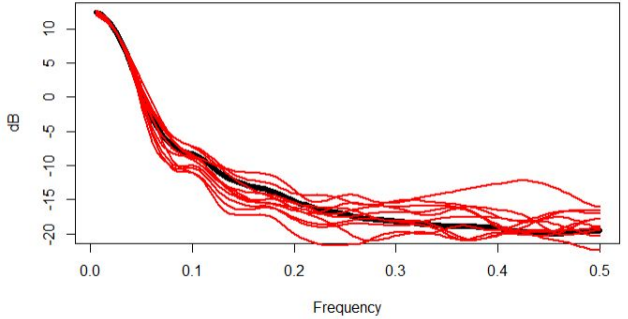


ARIMA Model further Evaluation

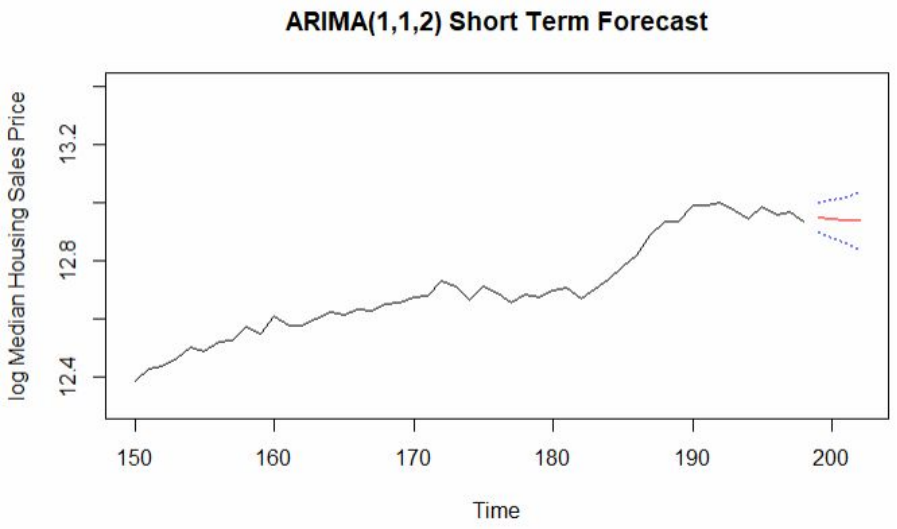


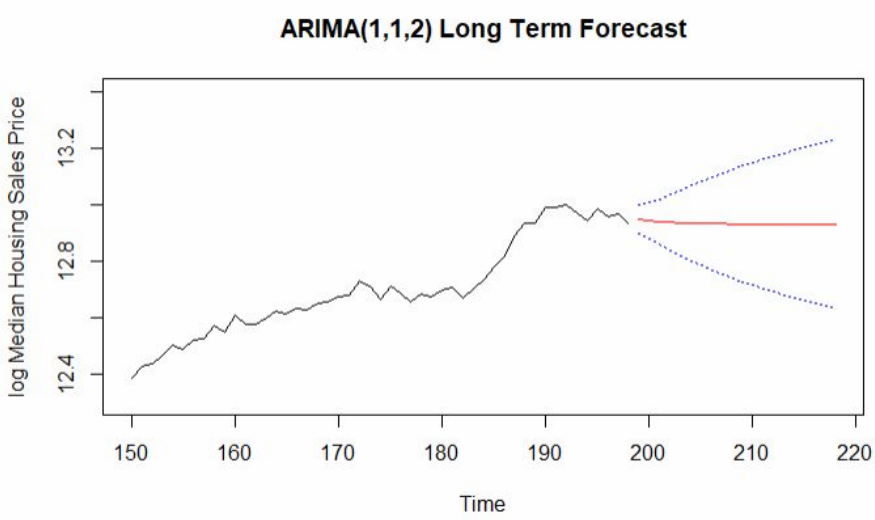






ARIMA Model Forecasts

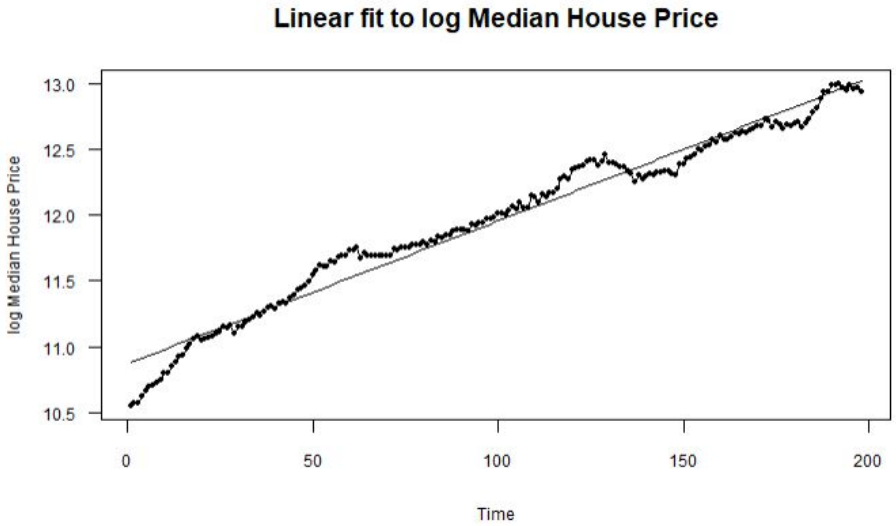




Signal Plus Noise Model Selection

- Use Line with AR noise
- Noise is modeled to be an AR(6)

```
Five Smallest Values of  aic
Method= mle
    p        aic
    6  -7.203842
    5  -7.198616
    7  -7.194018
    8  -7.188880
    4  -7.180527
```



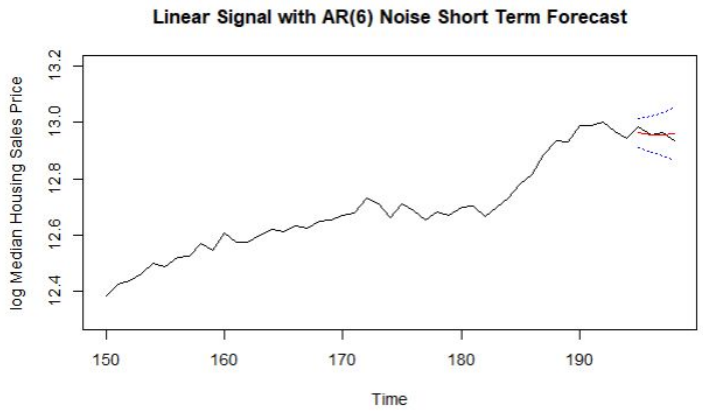
$$\log(X(t)) = 10.871 + 0.111 * t + Z(t)$$

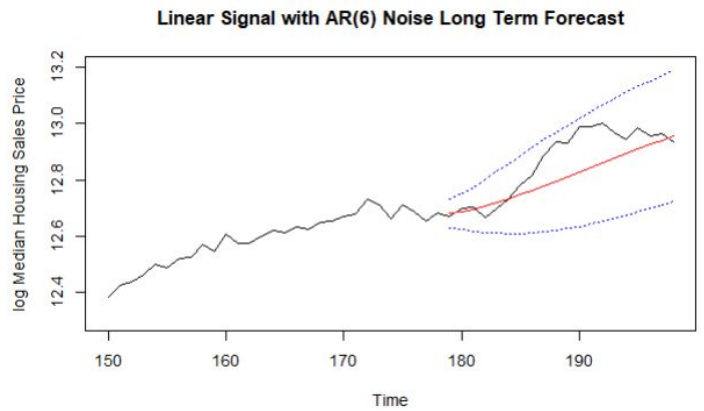
$$Z(t) = (1 - 0.71B - 0.367B^2 - 0.137B^3 + 0.062B^4 + 0.76B^5 + 0.118B^6) * a(t)$$

$$\text{var}(a(t)) = 0.0007$$

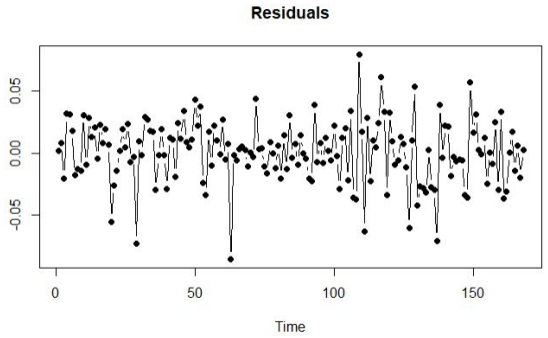
Signal Plus Noise Model Metrics

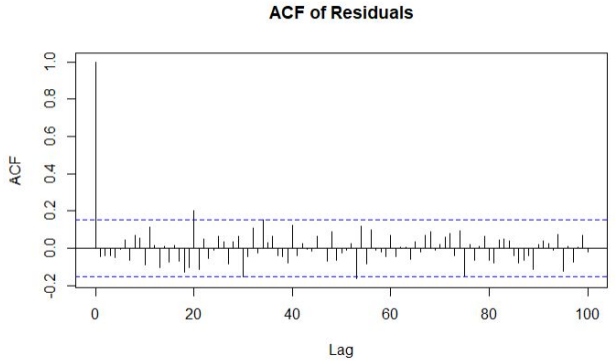
Short term rwRMSE	0.033
Long term rwRMSE	0.076
Short term ASE	50.9 M
Long term ASE	1.10 B

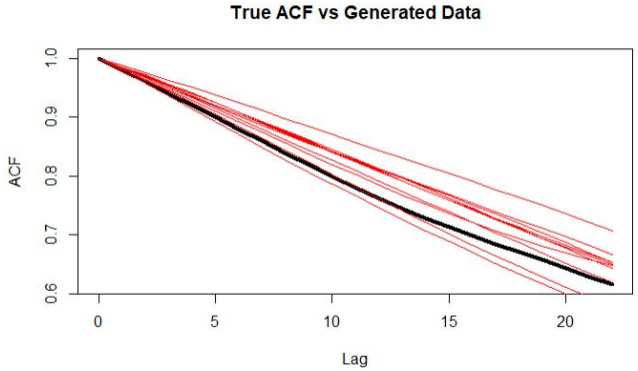


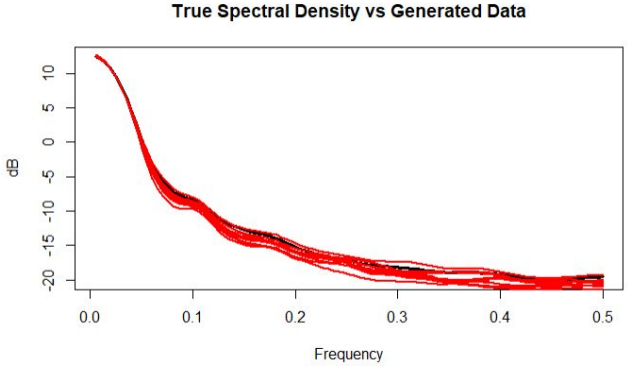


Signal Plus Noise Model further Evaluation

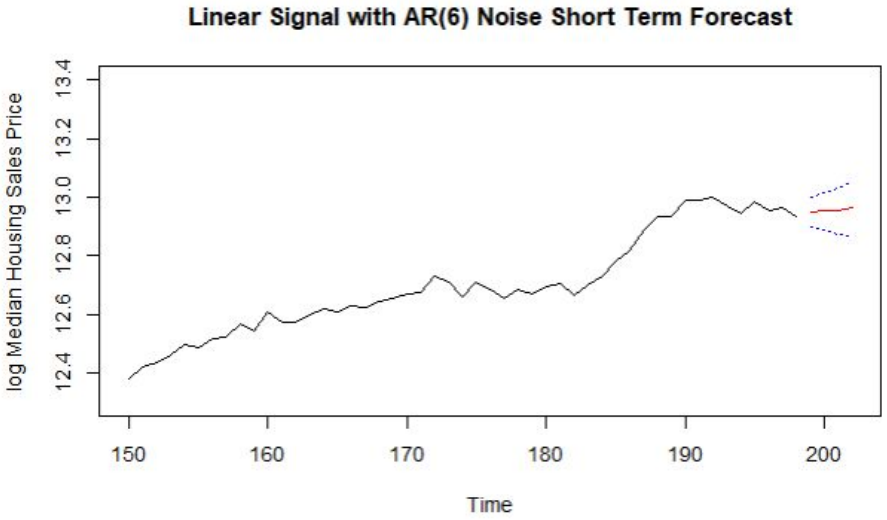


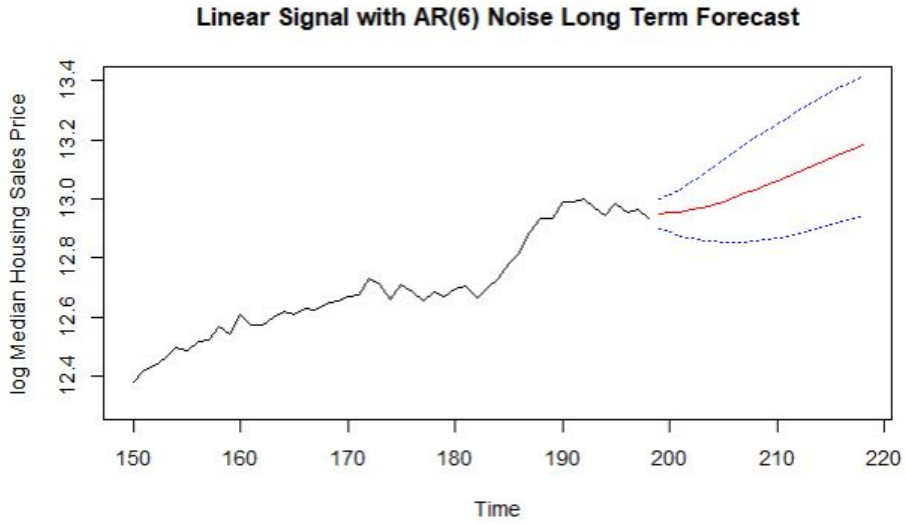




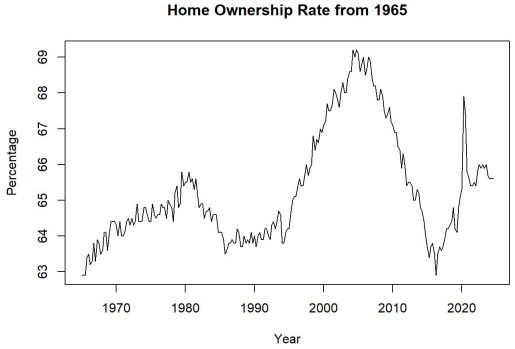


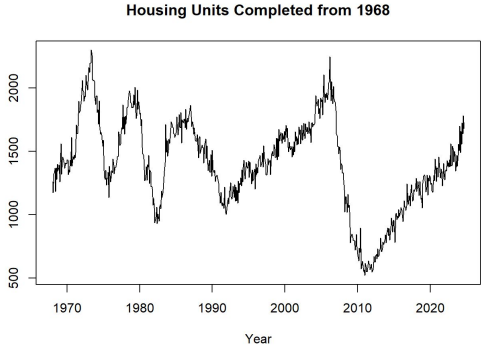
Signal Plus Noise Model Forecasts

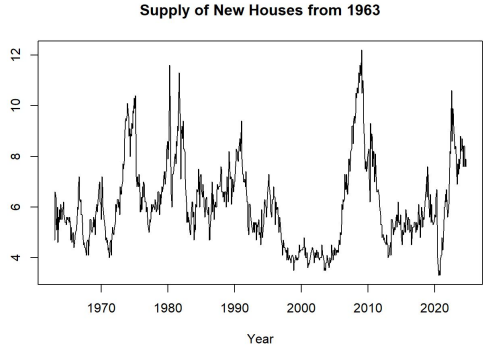


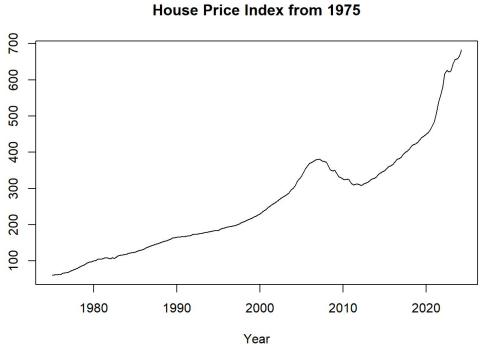


Exogenous Variables for Multivariate Model









MLR Model Selection

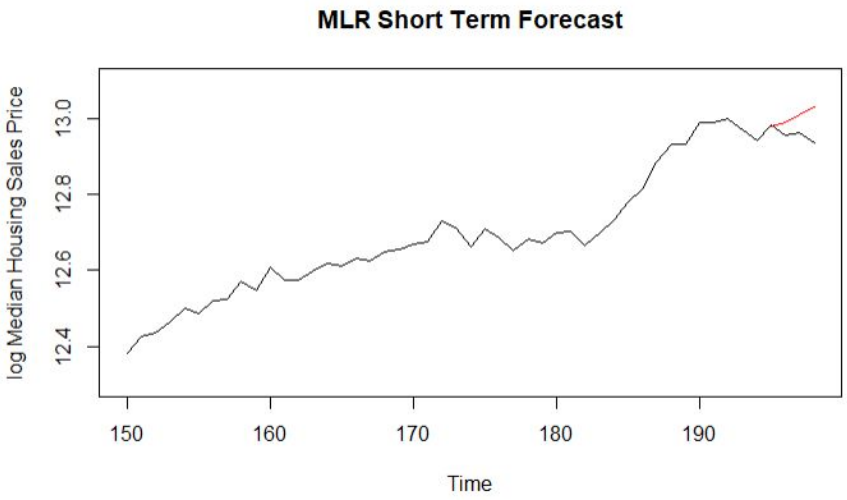
- Linear combination of Exogenous variables
- Used forecasted values for Exogenous variables
- Variables can be lagged
- AR(2) used to model residuals

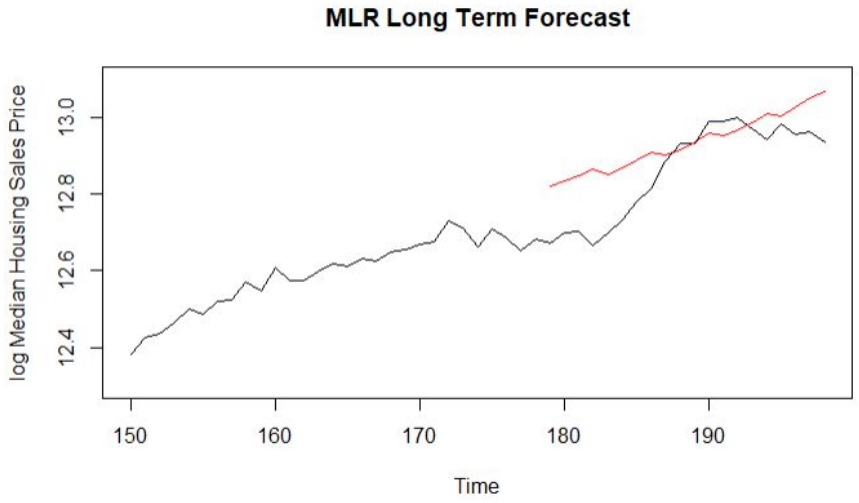

```
Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)              1.108e+01  3.039e-02 364.541  < 2e-16 ***
x$Housing_Units_Completed_l21 5.310e-05  1.679e-05   3.163  0.00185 **
x$Supply_New_Houses_l9   -2.821e-02  3.303e-03  -8.541 6.84e-15 ***
x$Housing_Price_Index     2.695e-04  1.055e-04   2.554  0.01151 *
t                         9.341e-03  2.925e-04  31.935  < 2e-16 ***
```

```
Five Smallest Values of  aic
   p  q        aic
   2  0 -6.610677
   1  1 -6.608485
   3  0 -6.599501
   2  1 -6.599466
   1  2 -6.595973
```

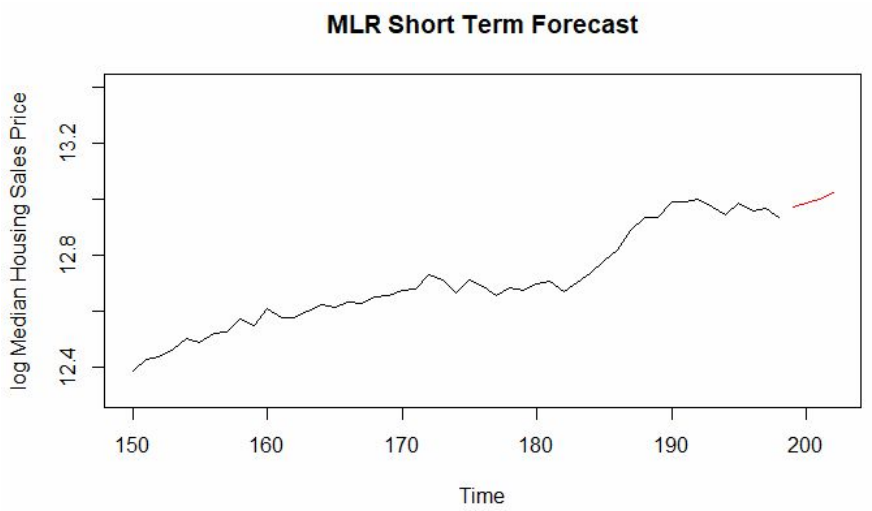
MLR Model Metrics

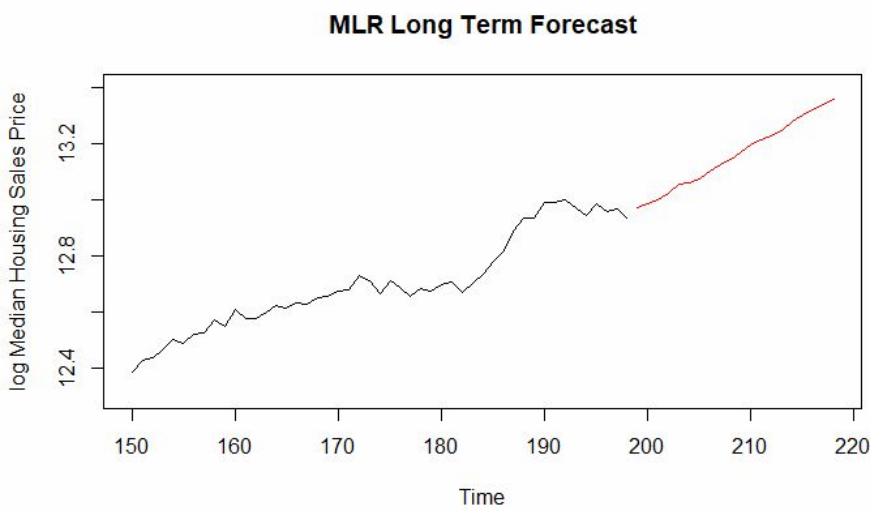
Short term ASE	558.7 M
Long term ASE	1.43 B





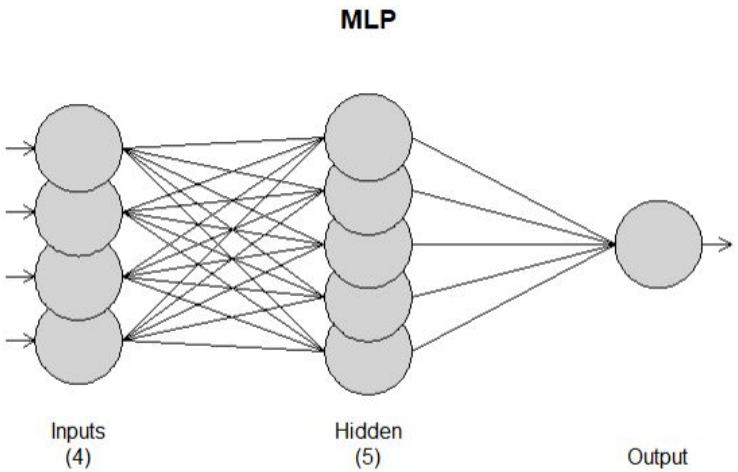
MLR Model Forecasts

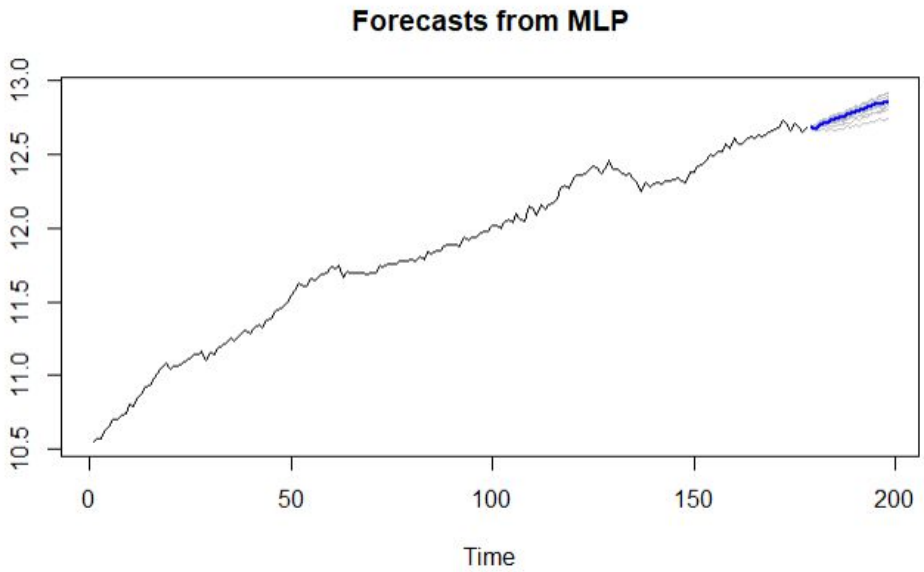




Univariate MLP Model Selection

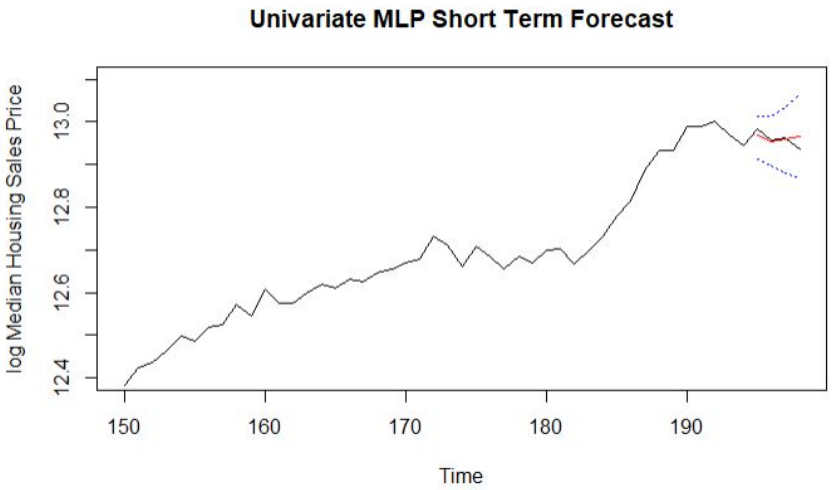
- MLP = Multilayer Perceptron
- 5 Hidden Nodes were used
- Median of 20 forecasts was used

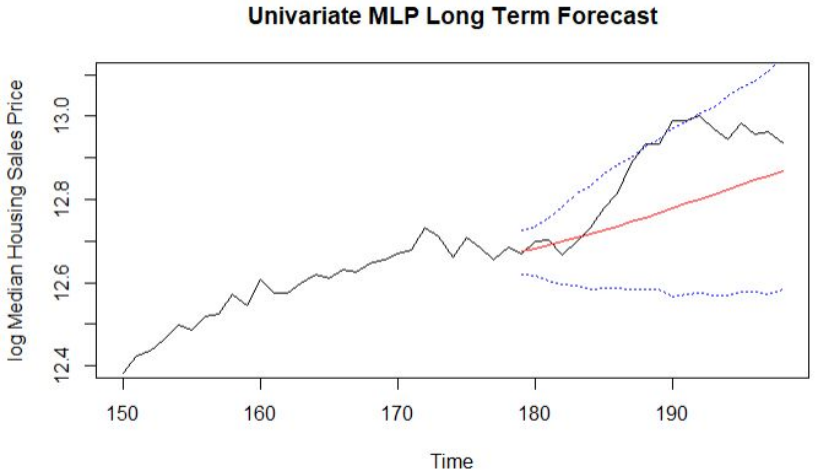




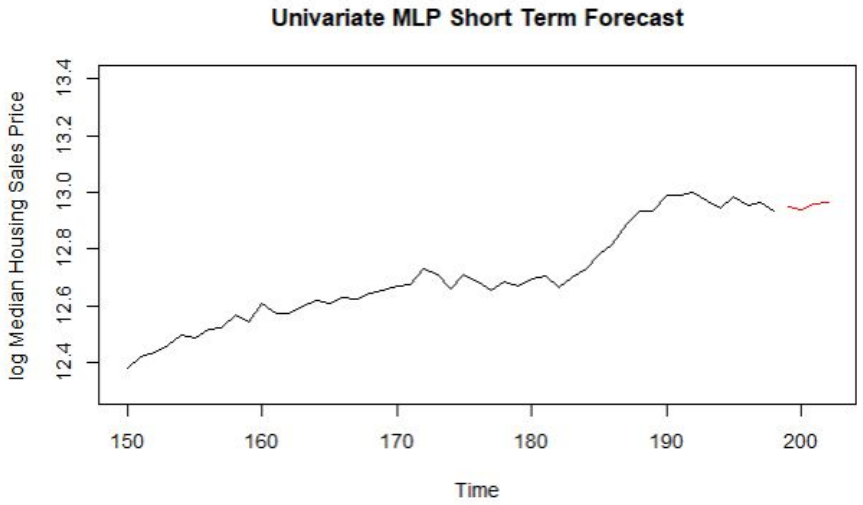
Univariate MLP Model Metrics

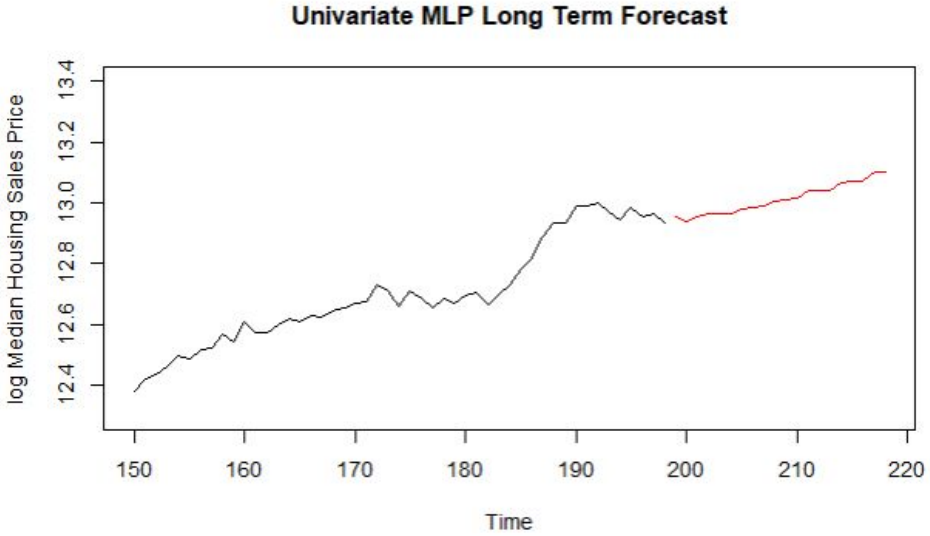
Short term rwRMSE	0.029
Long term rwRMSE	0.072
Short term ASE	61.0 M
Long term ASE	1.72 B





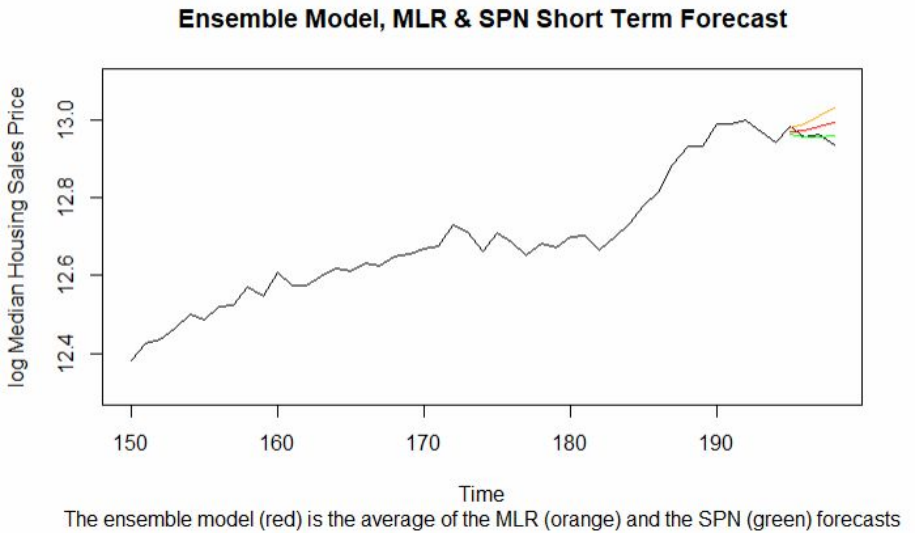
Univariate MLP Model Forecasts



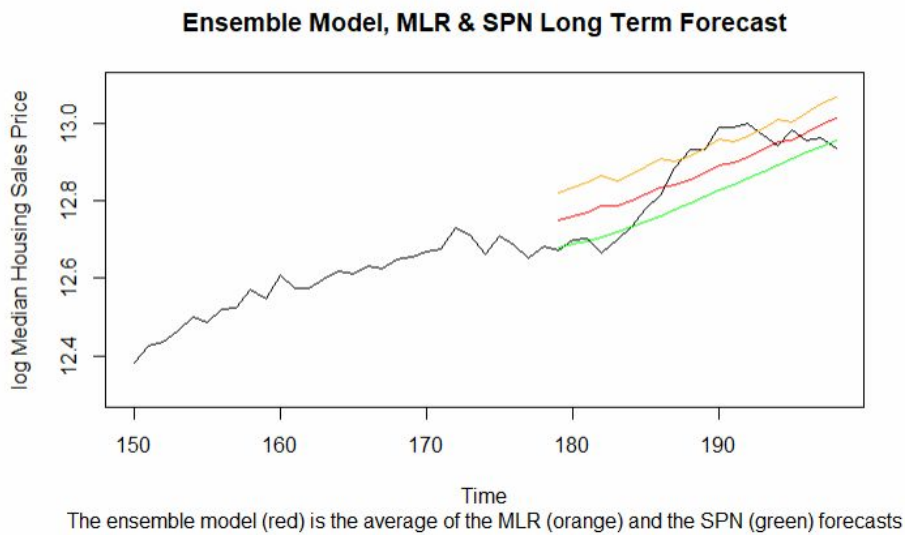


Ensemble Metrics

Short term ASE	207 M
Long term ASE	653 M

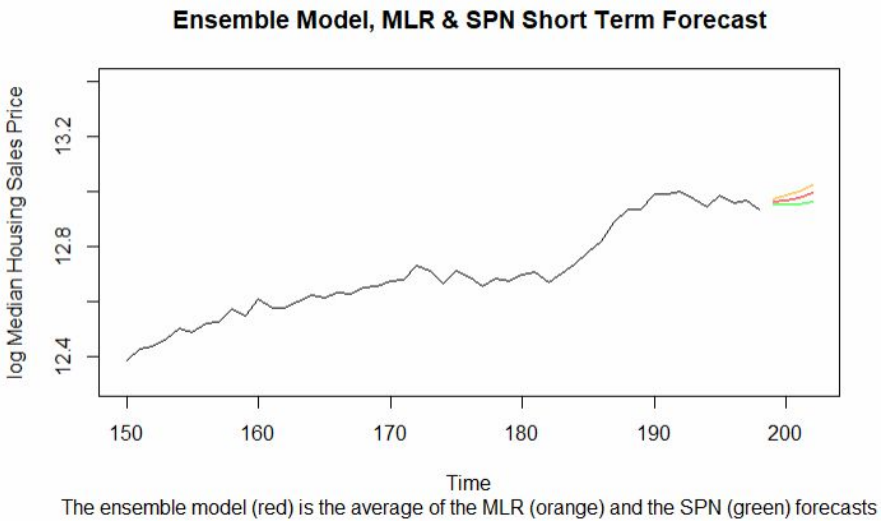


The ensemble model (red) is the average of the MLR (orange) and the SPN (green) forecasts

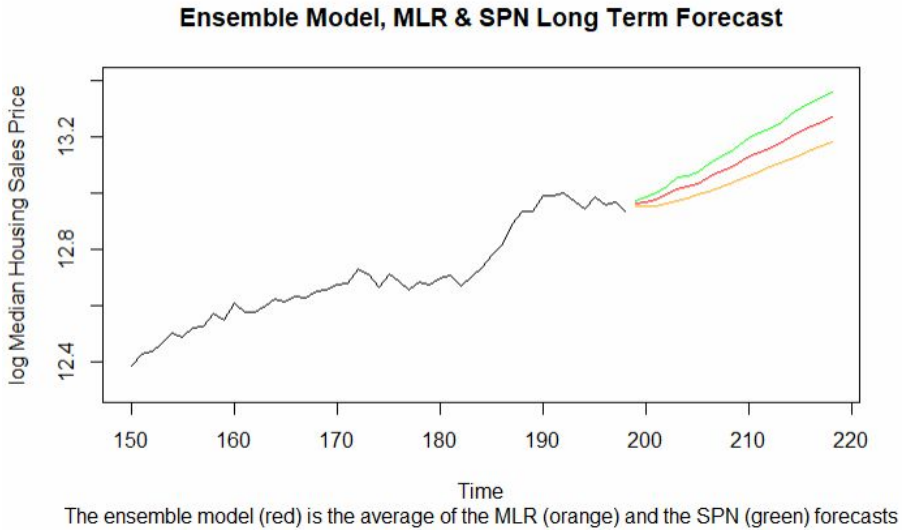


The ensemble model (red) is the average of the MLR (orange) and the SPN (green) forecasts

Ensemble Forecasts



The ensemble model (red) is the average of the MLR (orange) and the SPN (green) forecasts



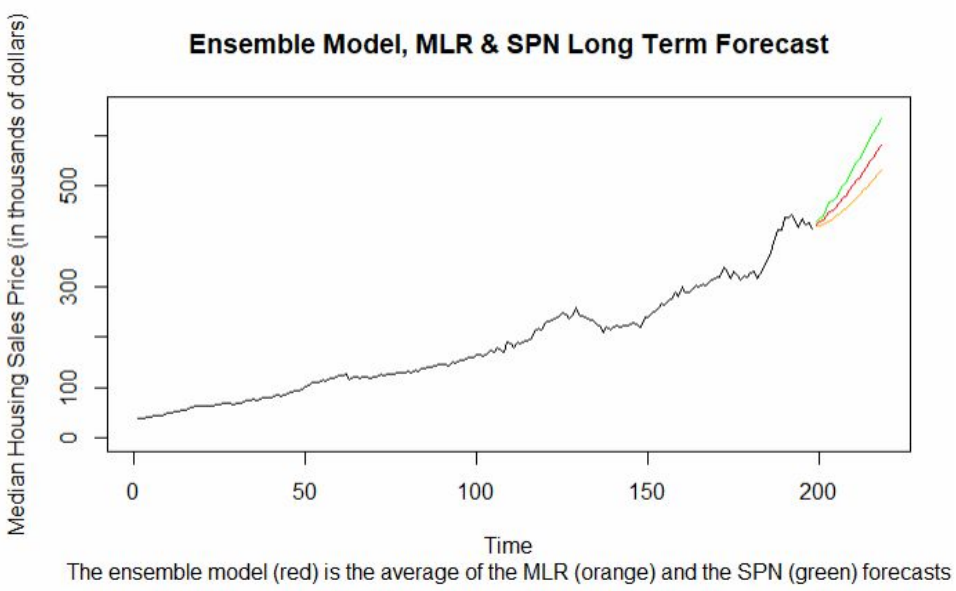
The ensemble model (red) is the average of the MLR (orange) and the SPN (green) forecasts

Overall Metric Comparisons

Model	Short term rwRMSE	Long term rwRMSE	Short term ASE	Long term ASE
ARIMA(1,1,2)	0.037	0.133	84.4 M	7.61 B
Signal Plus Noise	0.033	0.076	50.9 M	1.10 B
MLR			558 M	1.43 B
Univariate MLP	0.029	0.072	60.9 M	1.72 B
Ensemble MLR & SPN			207 M	653 M

Final Forecasts: Ensemble Model

- Ensemble Model does a good job of continuing the visible trend
- 2030 median prices forecast to be ~600k



The ensemble model (red) is the average of the MLR (orange) and the SPN (green) forecasts

THANK YOU

Aaron Abromowitz
aabromowitz@smu.edu

Alexandra Thibeaux
athibeaux@smu.edu

GitHub Repository for this project:
https://github.com/aabromowitz/TimeSeriersProject

Backup Slides

VAR Model Selection

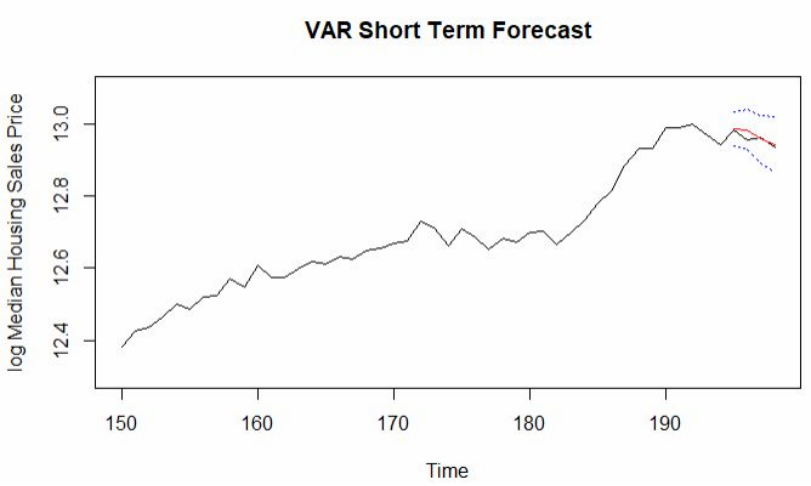
- VAR = Vector Autoregression
- Forecasts multiple variables simultaneously
- Lag 4 had lowest AIC

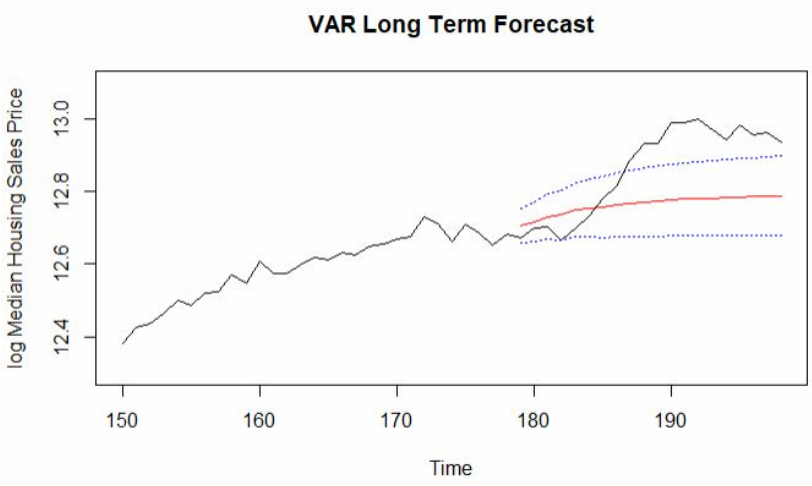
```
$selection
AIC(n)  HQ(n)  SC(n) FPE(n)
     4      2      1      4

$criteria
                1         2         3         4         5
AIC(n)  2.330754 1.933777 1.831970 1.690855 1.754028
HQ(n)   2.580534 2.361972 2.438579 2.475879 2.717466
SC(n)   2.946909 2.990043 3.328346 3.627343 4.130626
FPE(n) 10.287646 6.922207 6.263258 5.456431 5.841904
```

	Estimate	Std. Error	t value	Pr(>|t|)	
Ownership_Rate.l1	-6.003e-03	5.113e-03	-1.174	0.241960	
Housing_Units_Completed.l1	1.114e-05	2.294e-05	0.486	0.627737	
Supply_New_Houses.l1	-7.269e-03	2.119e-03	-3.431	0.000753	***
Housing_Price_Index.l1	1.695e-03	5.347e-04	3.170	0.001806	**
Median_Sales_Price.l1	4.816e-01	7.655e-02	6.291	2.52e-09	***
Ownership_Rate.l2	5.273e-03	6.663e-03	0.791	0.429799	
Housing_Units_Completed.l2	-2.239e-05	2.682e-05	-0.835	0.404910	
Supply_New_Houses.l2	4.012e-03	2.584e-03	1.553	0.122329	
Housing_Price_Index.l2	-3.055e-04	1.013e-03	-0.301	0.763406	
Median_Sales_Price.l2	3.289e-01	8.227e-02	3.998	9.46e-05	***
Ownership_Rate.l3	-3.138e-04	6.568e-03	-0.048	0.961956	
Housing_Units_Completed.l3	4.574e-05	2.641e-05	1.732	0.085056	.
Supply_New_Houses.l3	-1.031e-03	2.645e-03	-0.390	0.697110	
Housing_Price_Index.l3	-1.645e-03	1.046e-03	-1.573	0.117607	
Median_Sales_Price.l3	1.478e-01	8.207e-02	1.801	0.073454	.
Ownership_Rate.l4	2.436e-03	5.053e-03	0.482	0.630377	
Housing_Units_Completed.l4	-2.301e-05	2.101e-05	-1.095	0.275009	
Supply_New_Houses.l4	3.184e-03	2.376e-03	1.340	0.181943	
Housing_Price_Index.l4	9.110e-05	6.411e-04	0.142	0.887170	
Median_Sales_Price.l4	-3.937e-02	7.274e-02	-0.541	0.589017	
const	8.113e-01	2.505e-01	3.239	0.001441	**
trend	1.077e-03	3.625e-04	2.972	0.003381	**

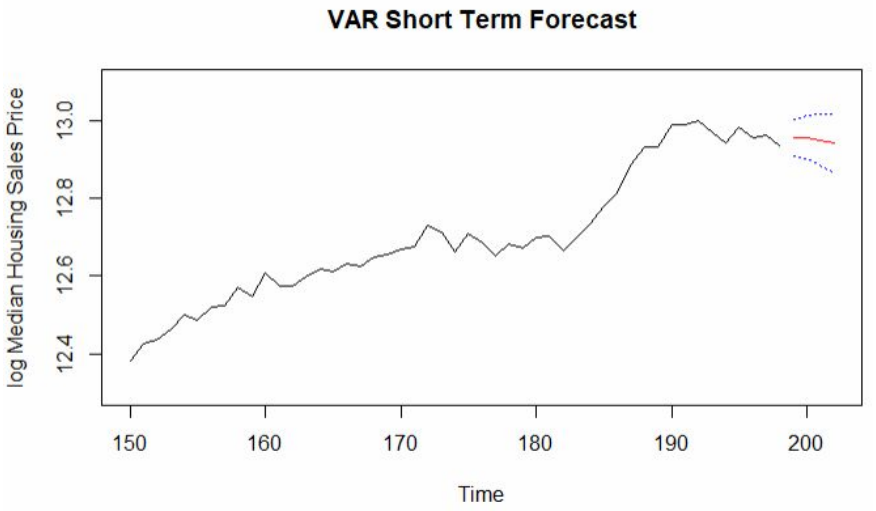
VAR Model Metrics

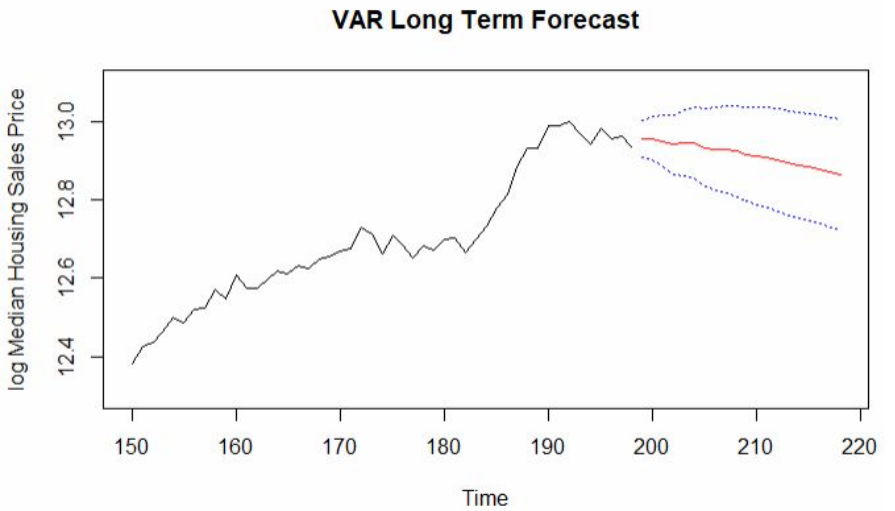
Short term ASE	43.6 M
Long term ASE	3.01 B



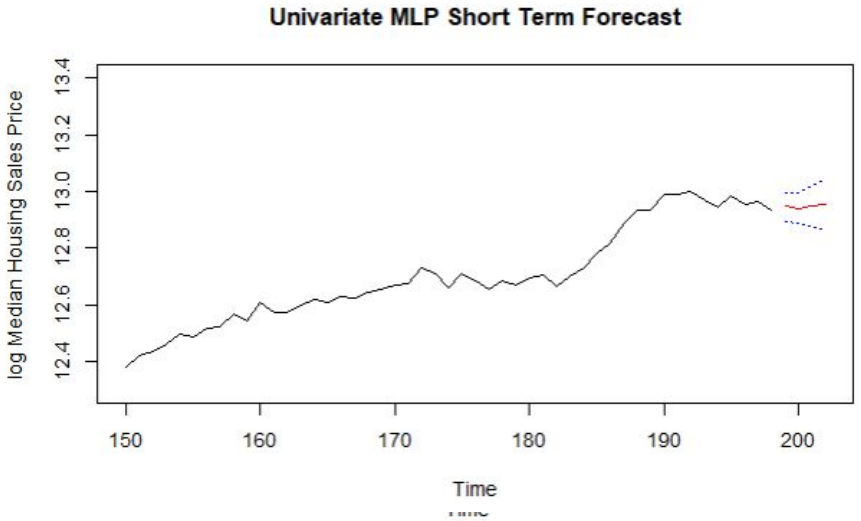


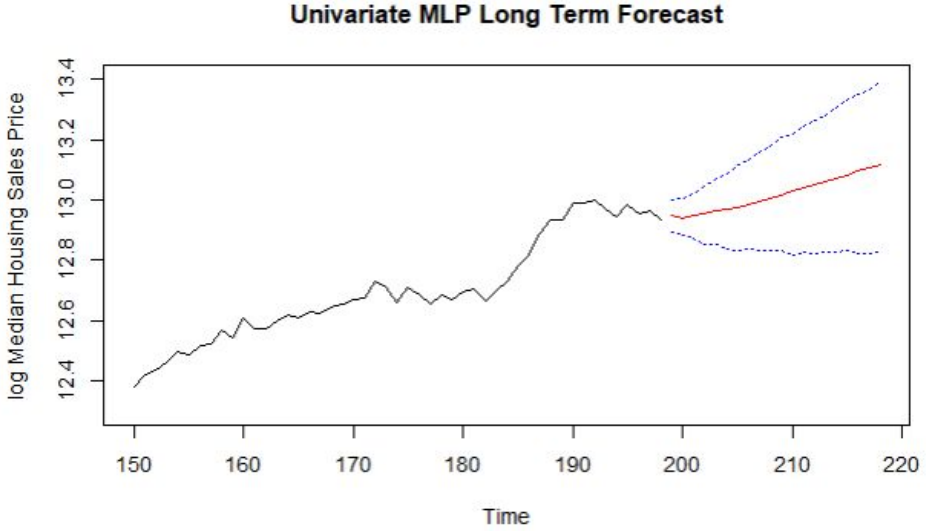
VAR Model Forecasts





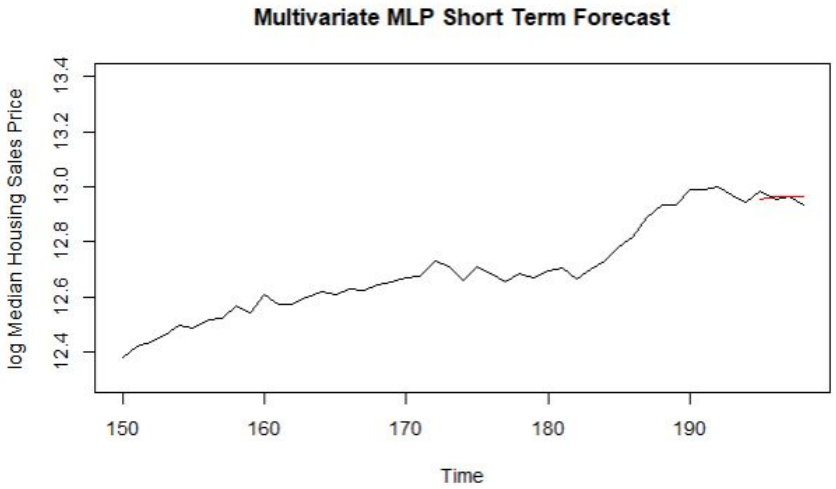
Univariate MLP Model Bootstrapped Forecasts with Confidence Intervals

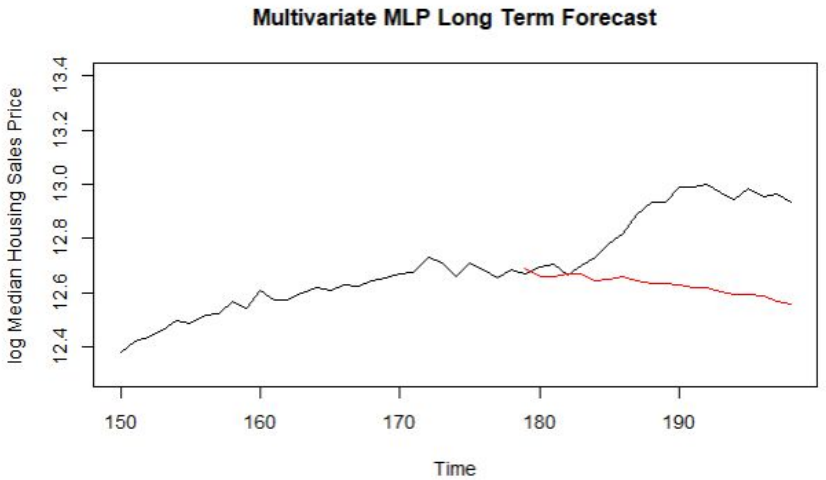




Multivariate MLP Model Metrics

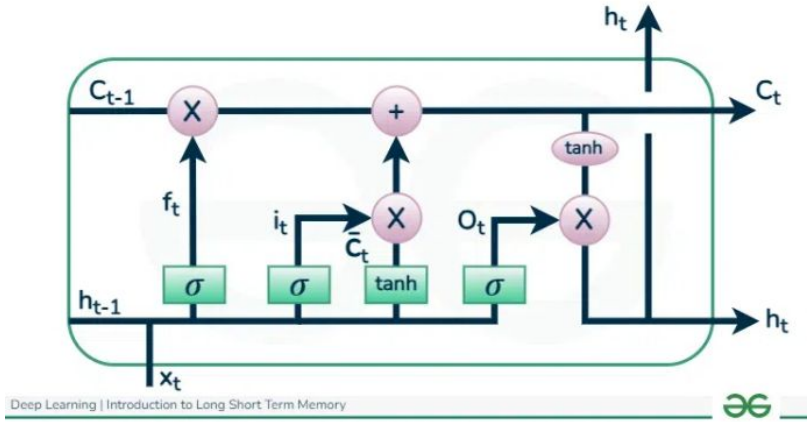
Short term ASE	254 M
Long term ASE	8.44 B





LSTM Model Information

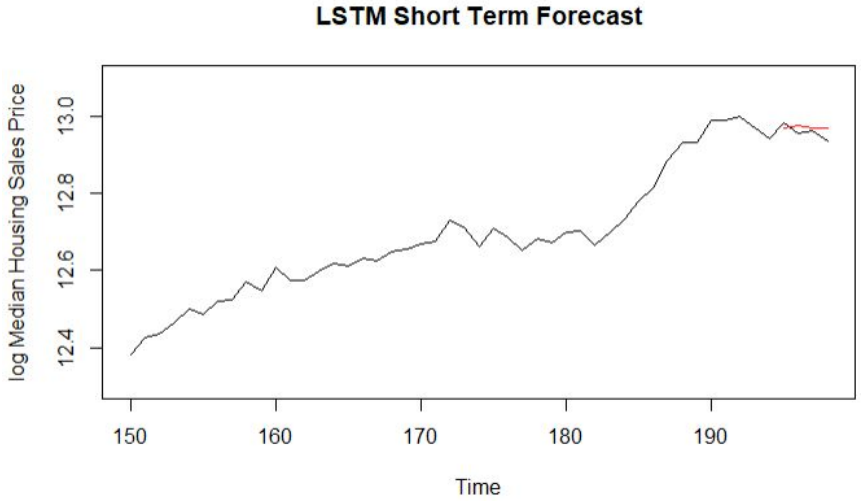
- LSTM = Long Short-Term Memory
- Neural Net that can hold information for an extended period
- Forecasts well on Time Series data

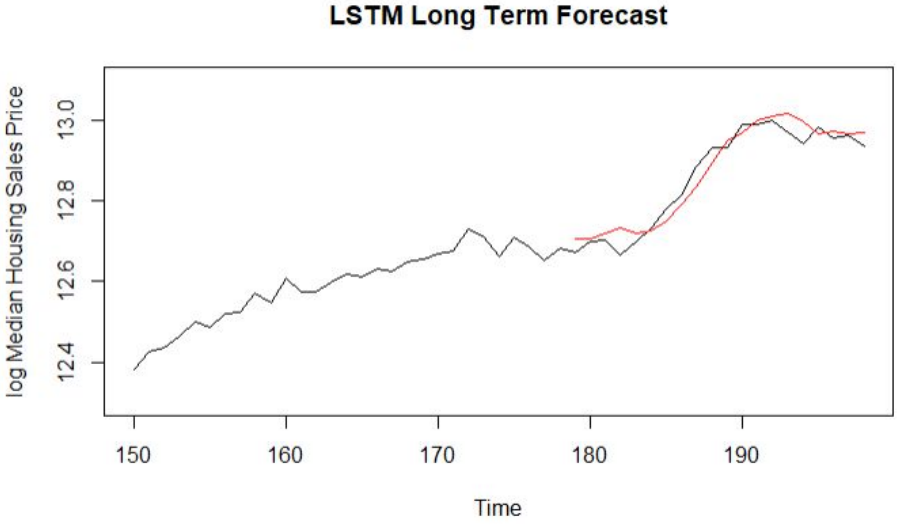


https://www.geeksforgeeks.org/deep-learning-introduction-to-long-short-term-memory/

LSTM Metrics

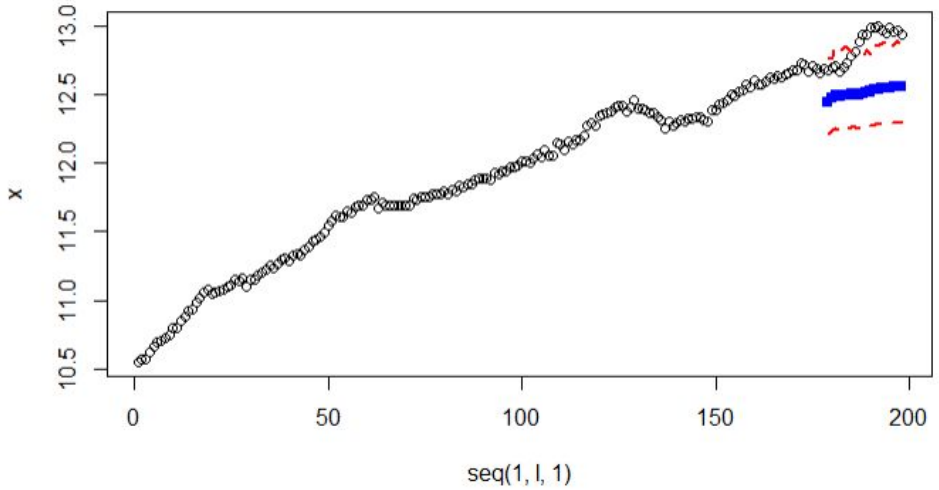
Short term ASE	81.8 M
Long term ASE	419 M





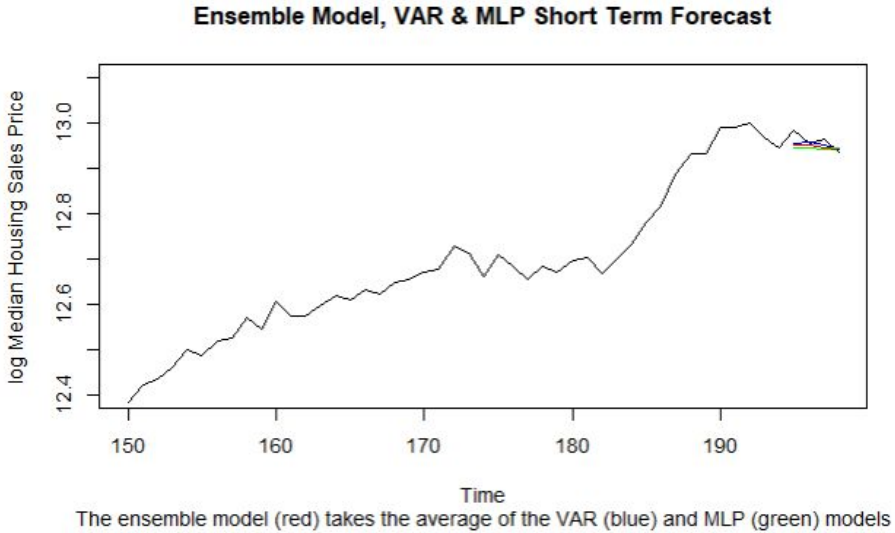
TFT Forecasts

Short term ASE	11.5 B
Long term ASE	15.0 B

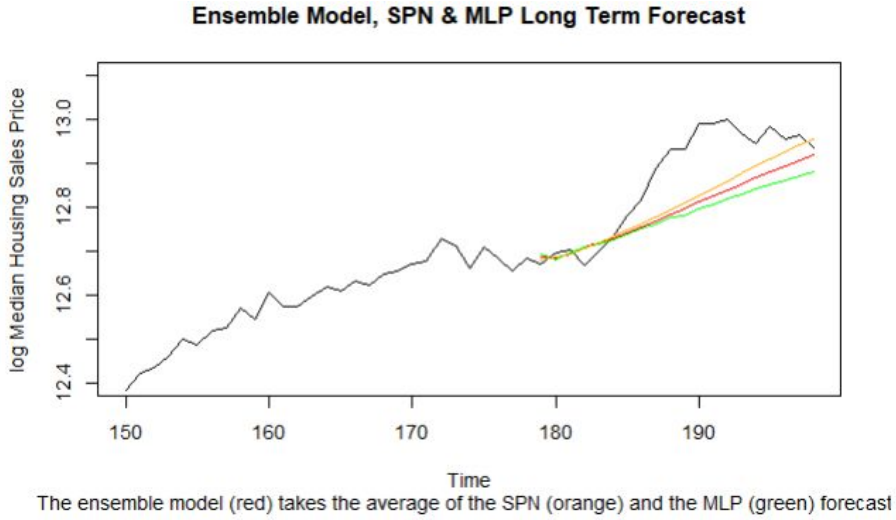


Ensemble Models

Short term ASE	67.0 M
Long term ASE	1.45 B



The ensemble model (red) takes the average of the VAR (blue) and MLP (green) models



The ensemble model (red) takes the average of the SPN (orange) and the MLP (green) forecast

Overall Metric Comparisons

Model	Short term rwRMSE	Long term rwRMSE	Short term ASE	Long term ASE
ARIMA(1,1,2)	0.037	0.133	84.4 M	7.61 B
Signal Plus Noise	0.033	0.076	50.9 M	1.10 B
MLR			558 M	1.43 B
VAR			46.3 M	3.50 B
Univariate MLP	0.029	0.072	60.9 M	1.72 B
Multivariate MLP			254 M	8.44 B
LSTM			81.8 M	419 M
Ensemble VAR & MLP			34.9 M	
Ensemble SPN & MLP				1.66 B