Visionary Final

Justin DePue, John Rooney, and Tony Jiang

2023-04-28

Contents

- Knitting commands in code chunks:
  - `include = FALSE` - code is run, but neither code nor results appear in knitted file
  - `echo = FALSE` - code not included in knitted file, but results are
  - `eval = FALSE` - code is not run in the knitted file
  - `message = FALSE` - messages do not appear in knitted file
  - `warning = FALSE` - warnings do not appear…
  - `fig.cap = "..."` - adds a caption to graphical results

##Introduction This study was born out of both curiosity and necessity. One of our teammates faced higher than expected energy bills this winter and was suddenly faced with the question of whether to shiver to save money or spend it on heating bills and be forced to eat Spaghetti-O's as sustenance. While spring has sprung and summer is on the horizon, we know that winter is coming and that we should prepare now in order that history not repeat itself.

As students recently armed with the tools to conduct time series analysis and forecasting, we realized we could complete our final project while aiding our teammate with knowledge for the future. Our study question thus emerged: would it be more cost-effective to heat an apartment in North Carolina using electricity or a natural-gas powered heater?

Full of optimism that we could complete a class requirement while doing some good for the world (for as Marvel taught us, when you help someone, you help everyone), we set out to find the data that would lead us to the answer we sought. Our journey led us to that great repository of energy knowledge, the US Energy Information Administration. There we found two datasets we felt confident would help us help our teammate: "North Carolina Price of Natural Gas Delivered to Residential Customers (Dollars Per Thousand Cubic Feet)", which contained monthly data from January 1989 through January 2023, and "Average Retail Price of Electricity by State and Sector" which contained monthly data from January 2001 through January 2023.

##Data

Table 1: Table 1: Summary Statistics

Variable	Observations	Min	Max	Mean Price	Std. Dev.	Median Price
Natural Gas ($/Mcf)	410	5.54	30.43	13.78	5.64	12.54
Electricity(cents/kWh)	265	7.53	13.51	10.22	1.30	10.41

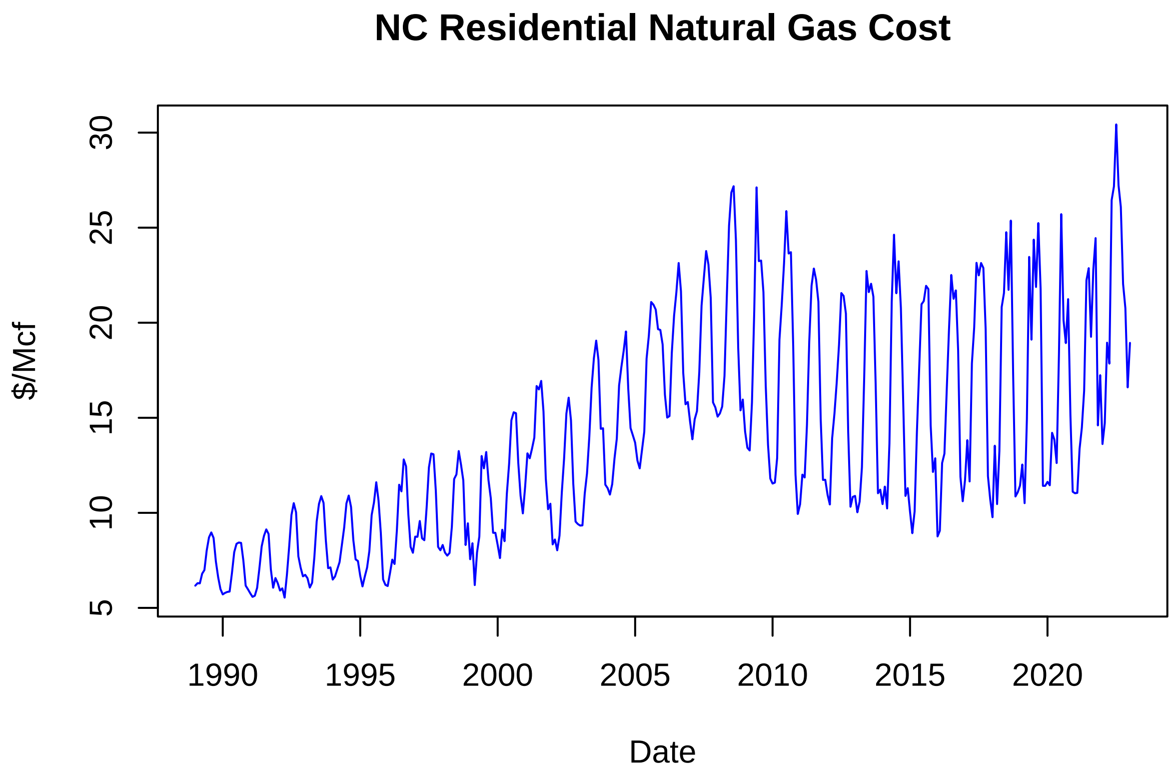
Table 2: Sample of Cleaned Natural Gas Data

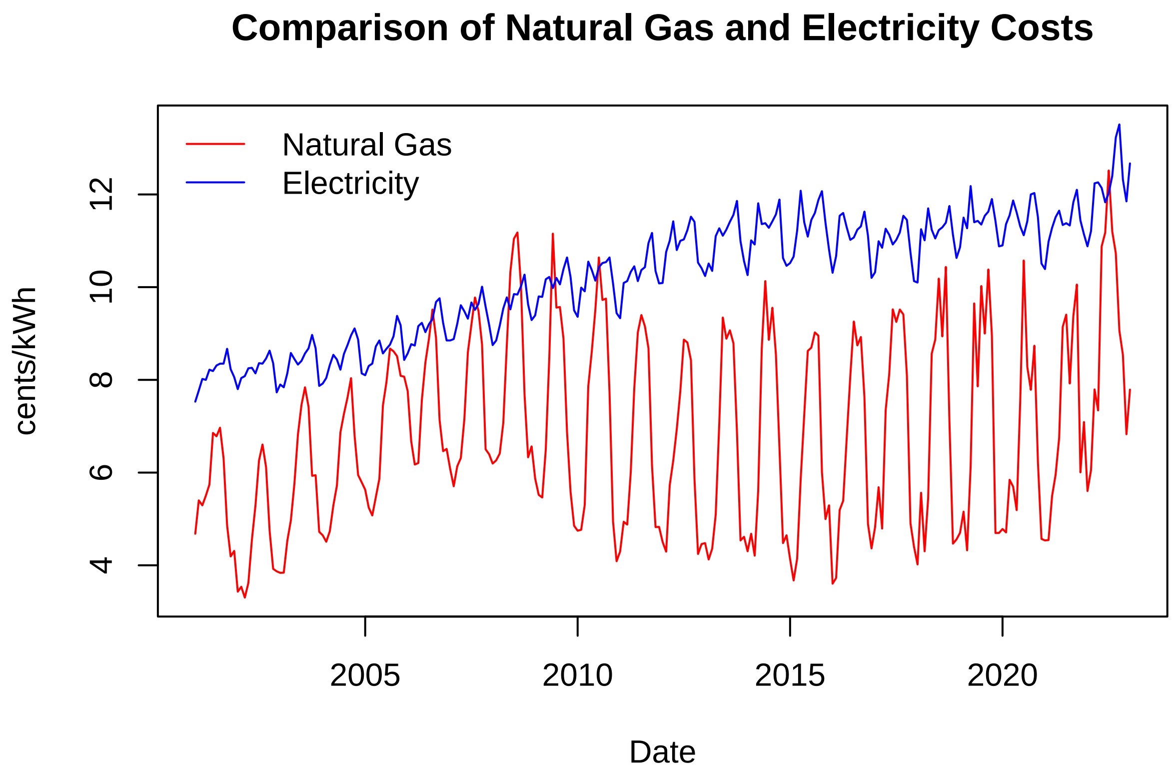
year	price
Jan-1989	6.17
Feb-1989	6.30
Mar-1989	6.29
Apr-1989	6.80
May-1989	6.99
Jun-1989	8.02
Jul-1989	8.71
Aug-1989	8.97
Sep-1989	8.68
Oct-1989	7.44

Table 3: Sample of Cleaned Electricity Data

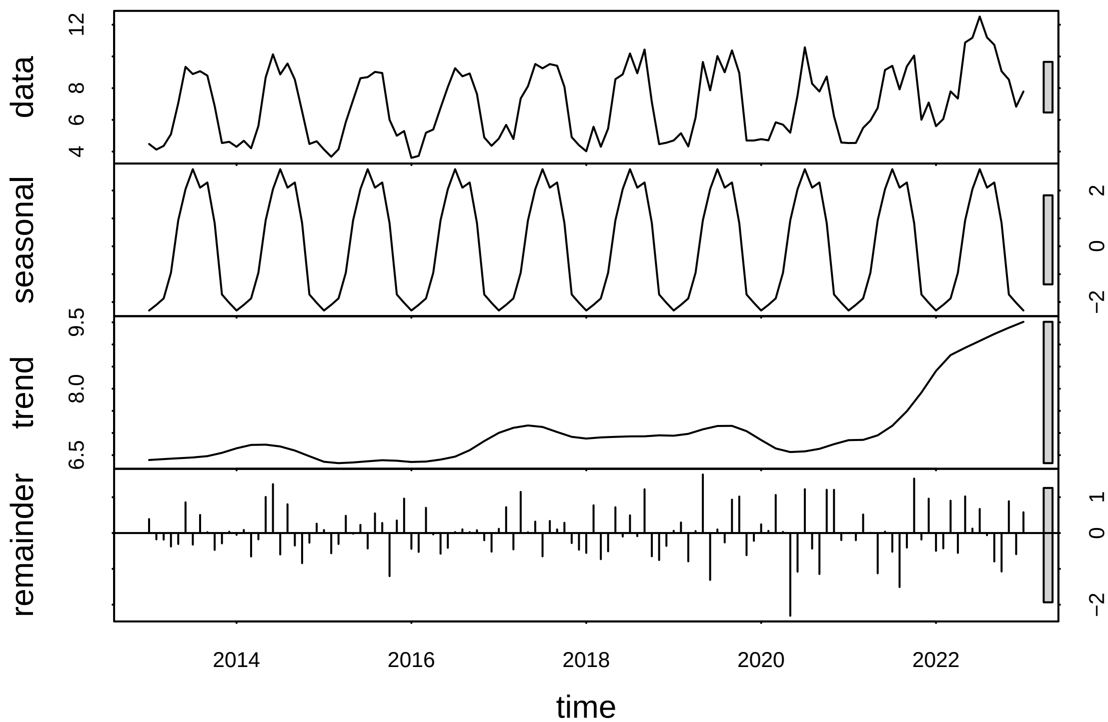
my_date	price_per_kWh
Jan.2001	7.53
Feb.2001	7.77
Mar.2001	8.02
Apr.2001	8.00
May.2001	8.22
Jun.2001	8.19
Jul.2001	8.31
Aug.2001	8.35
Sep.2001	8.35
Oct.2001	8.67

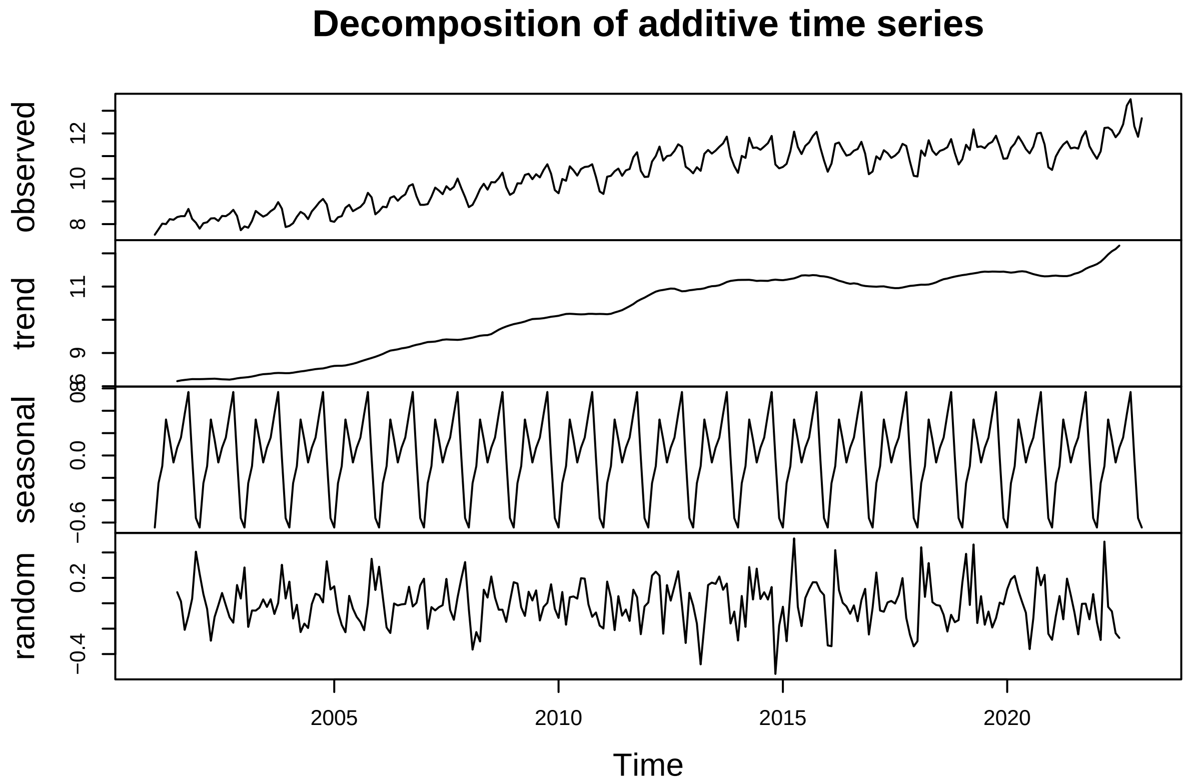
##Analysis We first created initial time series objects and plotted them, along with ACF and PACF plots to gain an initial sense of what the series looked like and what seasonality they may have.

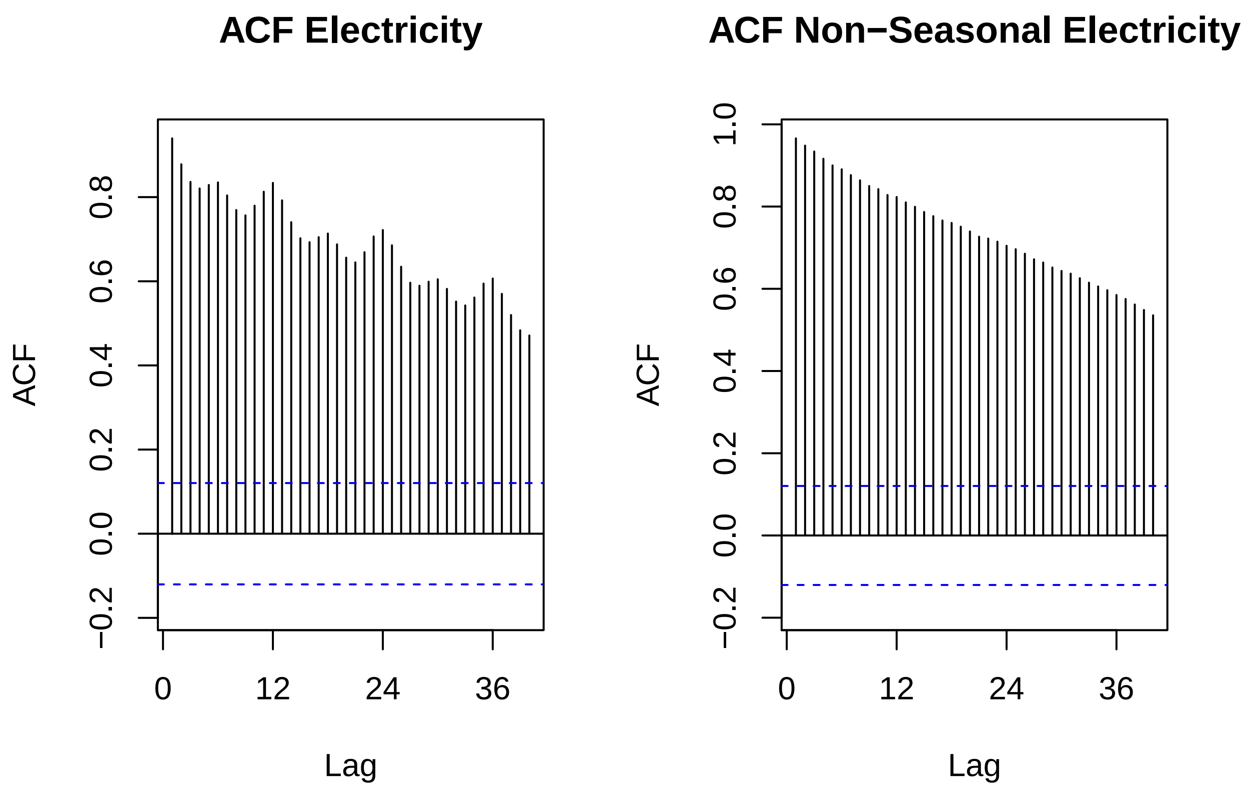

NC Residential Natural Gas Cost
$/Mcf
5
10
15
20
25
30
1990
1995
2000
2005
2010
2015
2020
Date



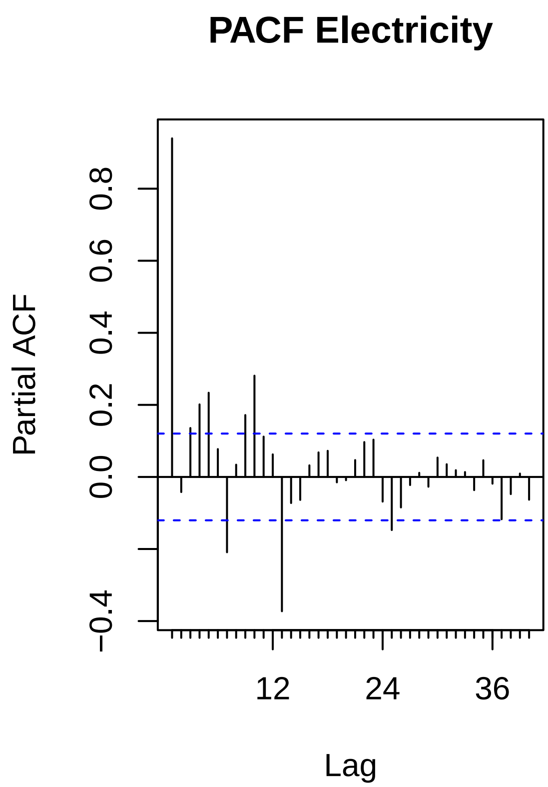
We then decomposed the time series objects for further analysis.

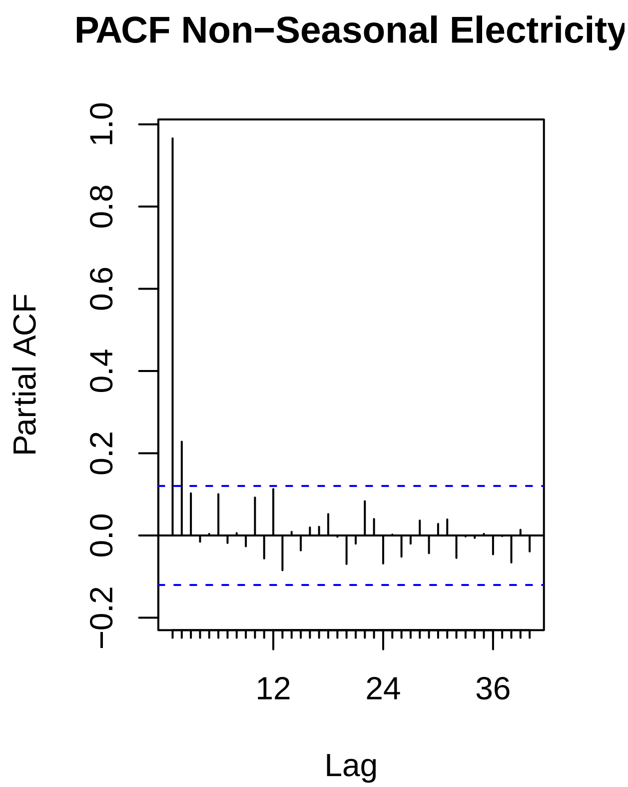

data
seasonal
trend
remainder
12
8
6
4
2
0
−2
9.5
8.0
6.5
1
0
−2
2014
2016
2018
2020
2022
time

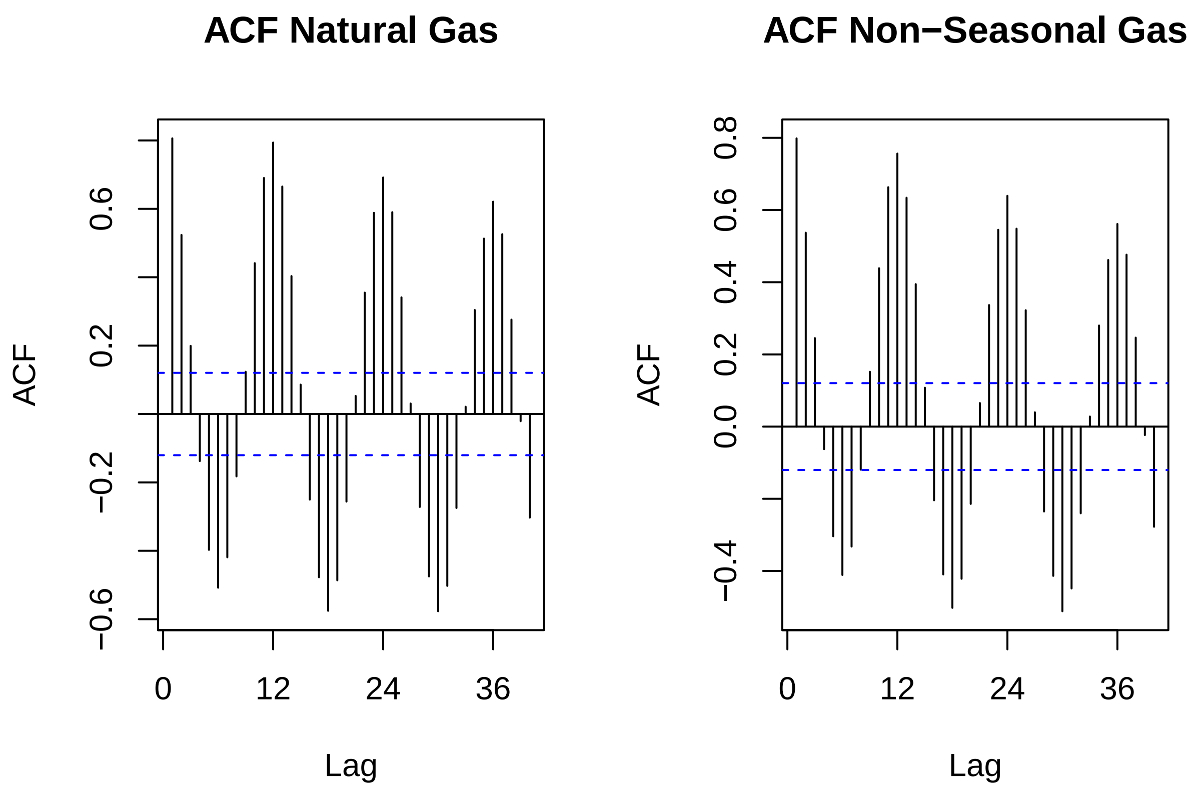

Decomposition of additive time series
observed
trend
seasonal
random
8
10
12
9
11
0.6
0.0
−0.6
0.2
−0.4
2005
2010
2015
2020
Time

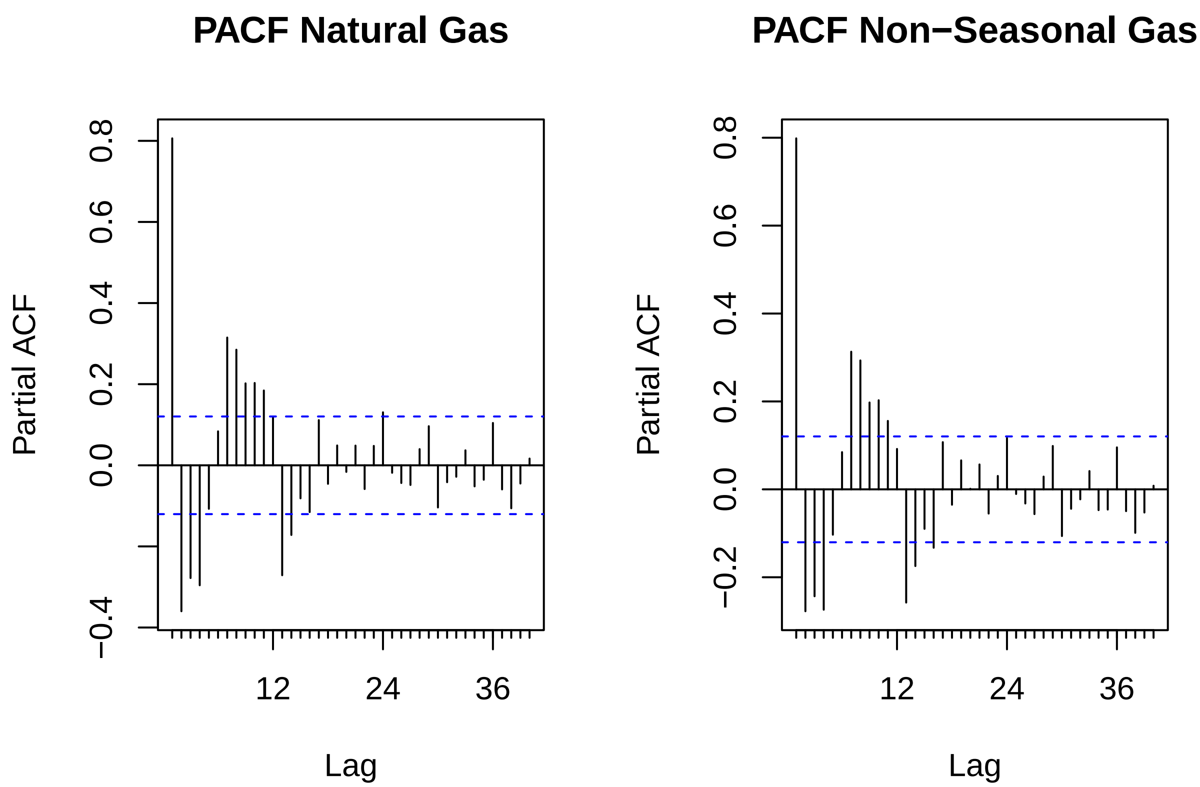


ACF Electricity
ACF
Lag
ACF Non−Seasonal Electricity
ACF
Lag


PACF Electricity
Partial ACF
−0.4
0.0
0.2
0.4
0.6
0.8
12
24
36
Lag


PACF Non−Seasonal Electricity
Partial ACF
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
12
24
36
Lag

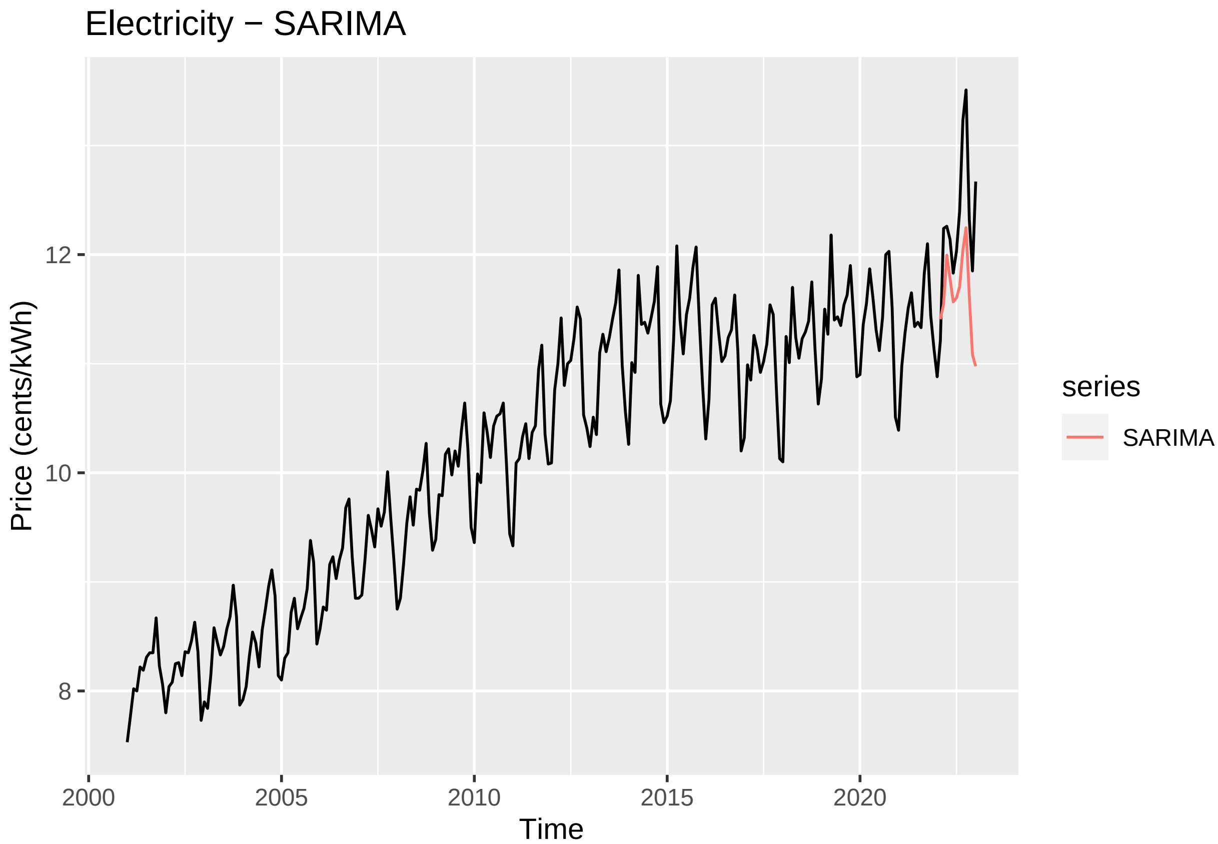

ACF Natural Gas
ACF
0.6
0.2
−0.2
−0.6
0
12
24
36
Lag
ACF Non−Seasonal Gas
ACF
0.8
0.6
0.4
0.2
0.0
−0.4
0
12
24
36
Lag

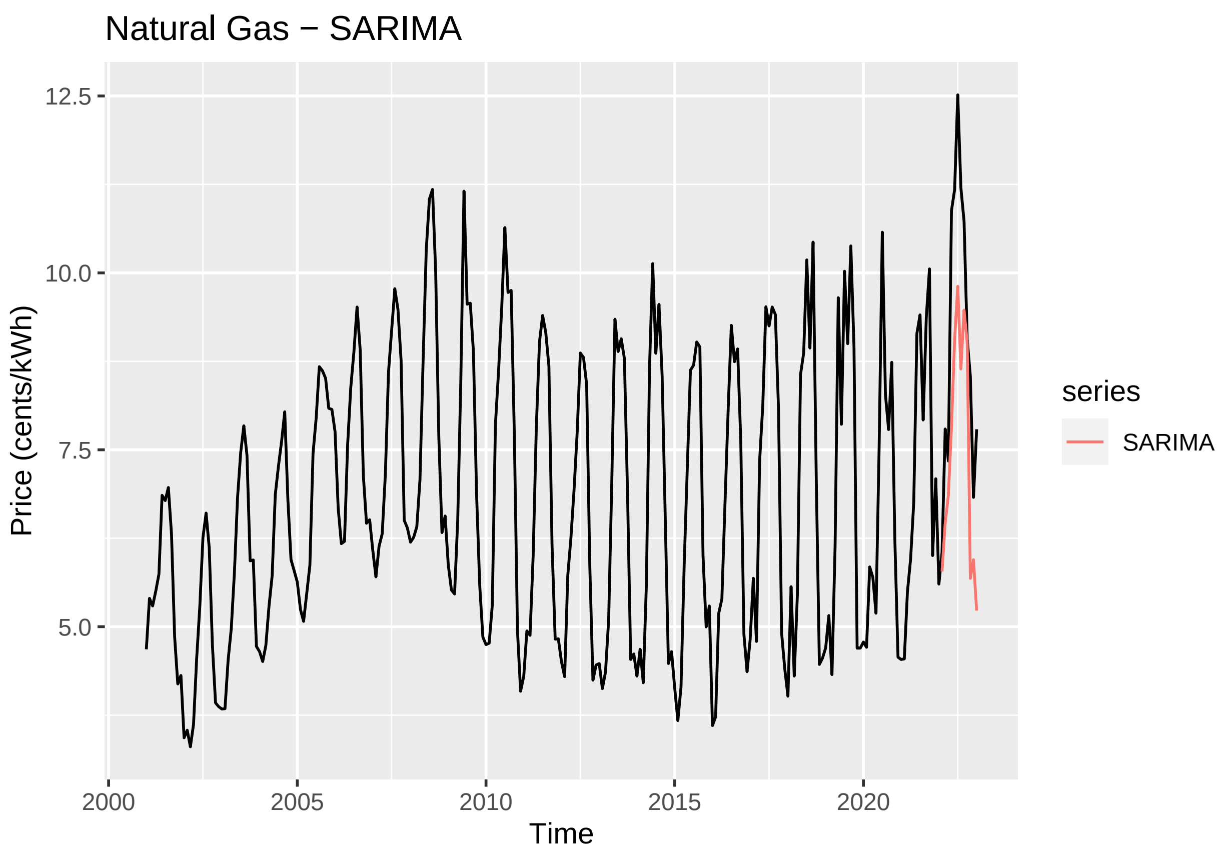


Several models were developed and tested to determine what would best fit the data we had and what may lead to the best forecast to determine whether heating via electricity or natural gas would be most cost-efficient in the upcoming winter. The five that were used were the Seasonal ARIMA, ARIMA with Fourier terms, Neural Networks, TBATS, and STL + EST.

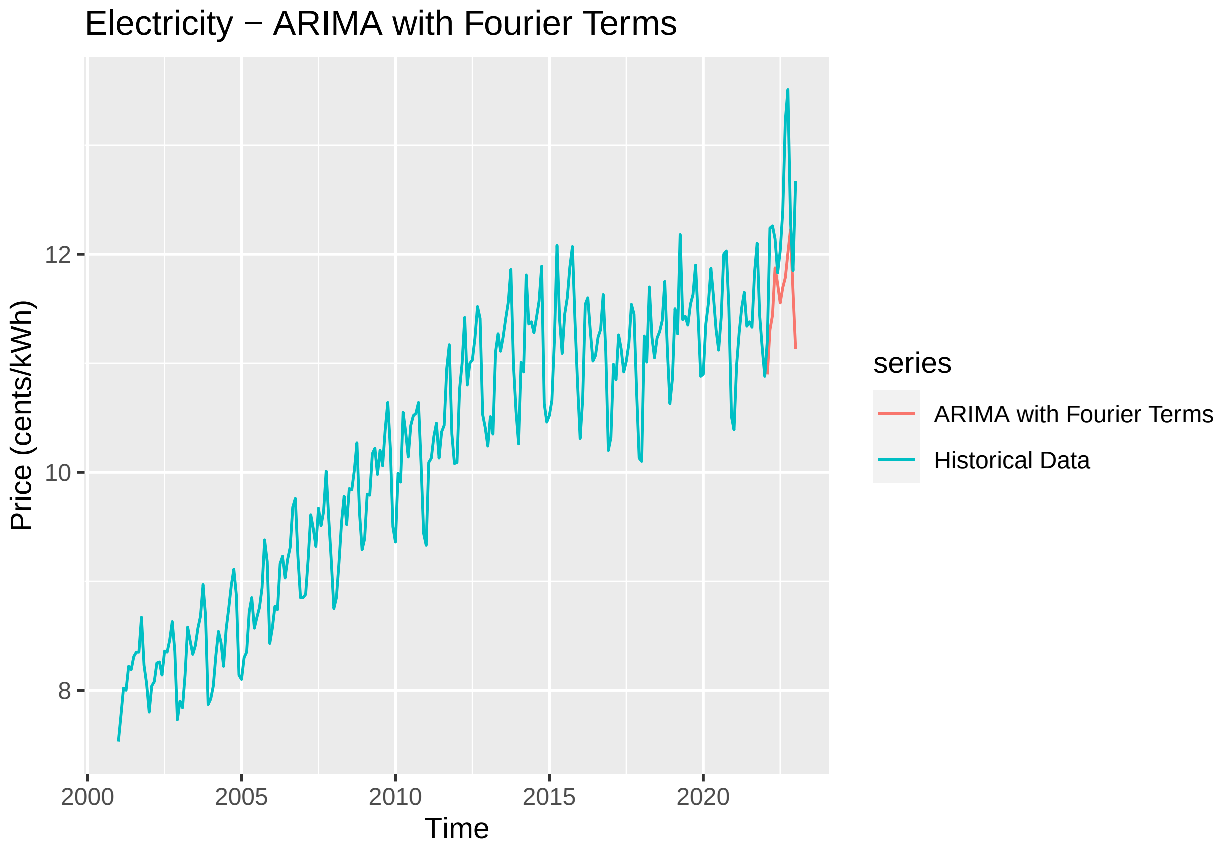
Each of these tests used functions from the "forecast" library and required the use of time series data which we created using the "ts" function from the "tseries" library.

While the Seasonal ARIMA was a logical first model to try, we quickly felt that the performance was not particularly strong.

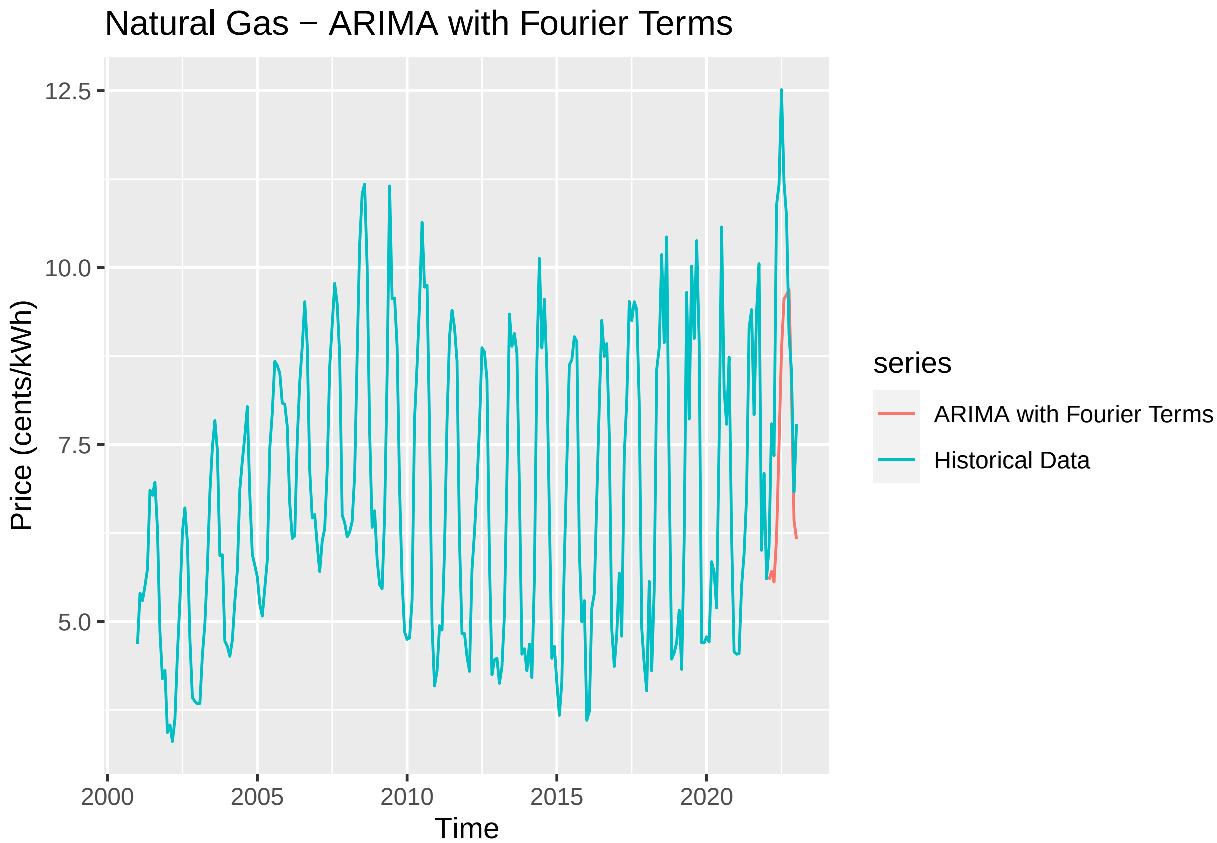

Electricity − SARIMA
Price (cents/kWh)
12
10
8
2000
2005
2010
2015
2020
Time
series
SARIMA



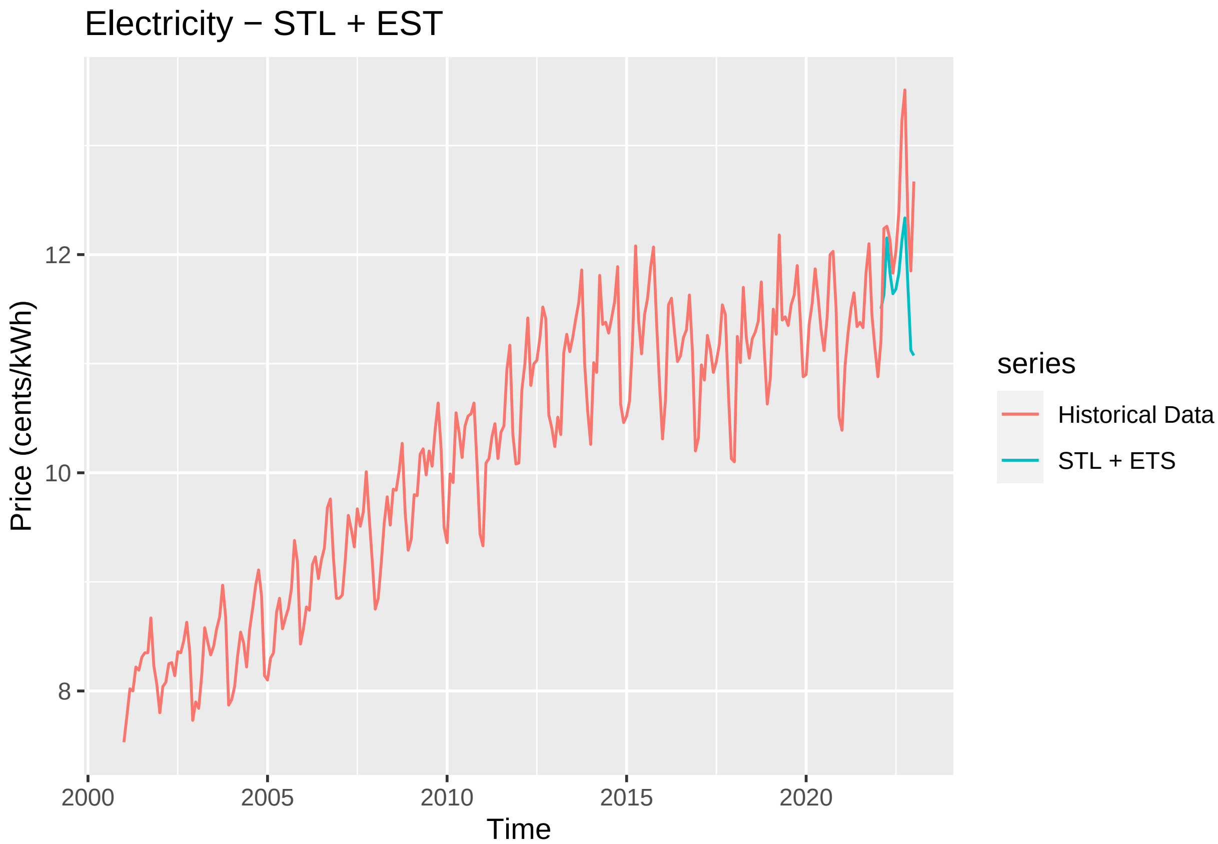
We decided to explore more advanced models that could handle the ARIMA with Fourier Terms model, Neural Networks, TBATS, and STL + EST. ARIMA with Fourier terms is known as a dynamic harmonic regression model with an ARMA error structure, using the "fourier" function from package "forecast" to find terms that model seasonal components.

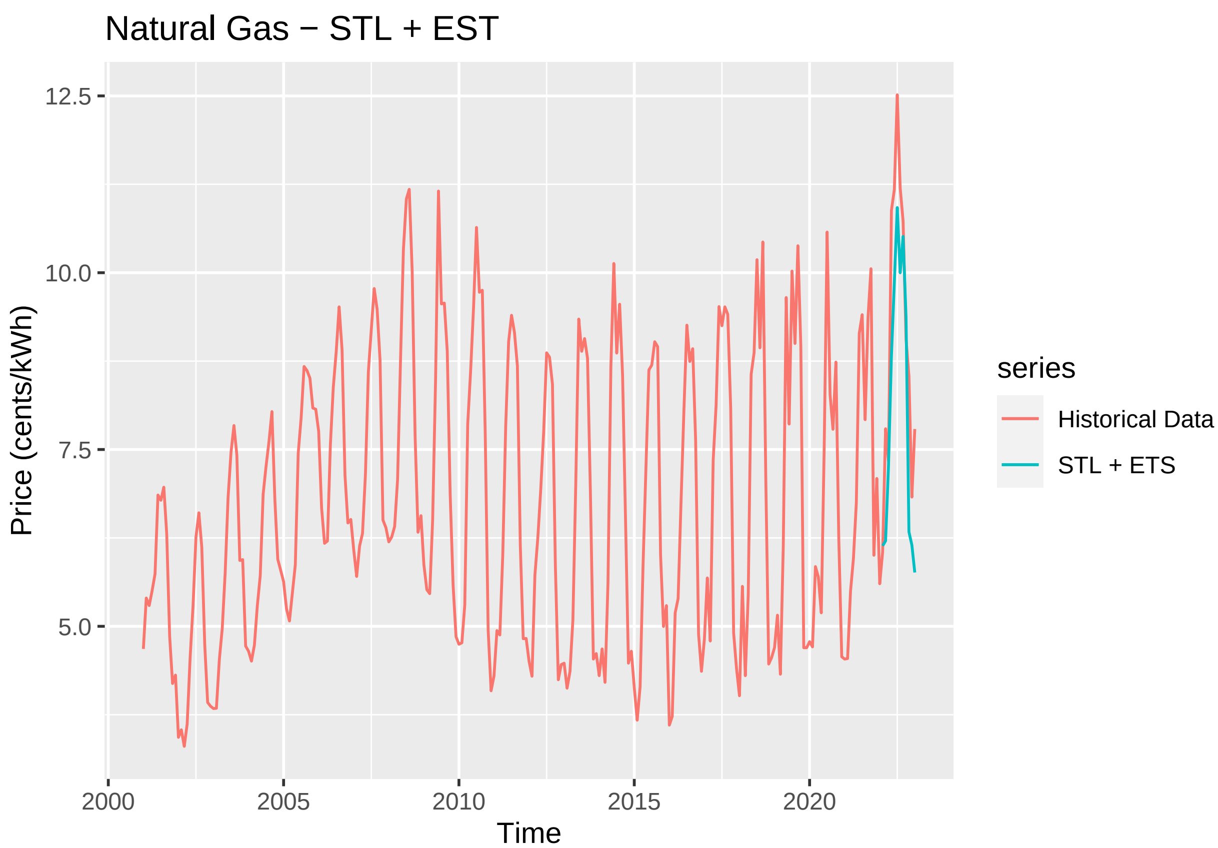


Electricity − ARIMA with Fourier Terms
Price (cents/kWh)
12
10
8
2000
2005
2010
2015
2020
Time
series
ARIMA with Fourier Terms
Historical Data



Then we tried STL model


Electricity − STL + EST
Price (cents/kWh)
12
10
8
2000
2005
2010
2015
2020
Time
series
Historical Data
STL + ETS



Then we tried to use neural network model. We used nnetar() in forecast package. We figured out p and P arguments in nnetar() have significant impact on model performance. Therefore, we first tried to identify the optimal p and P conbination by trying different combinations.

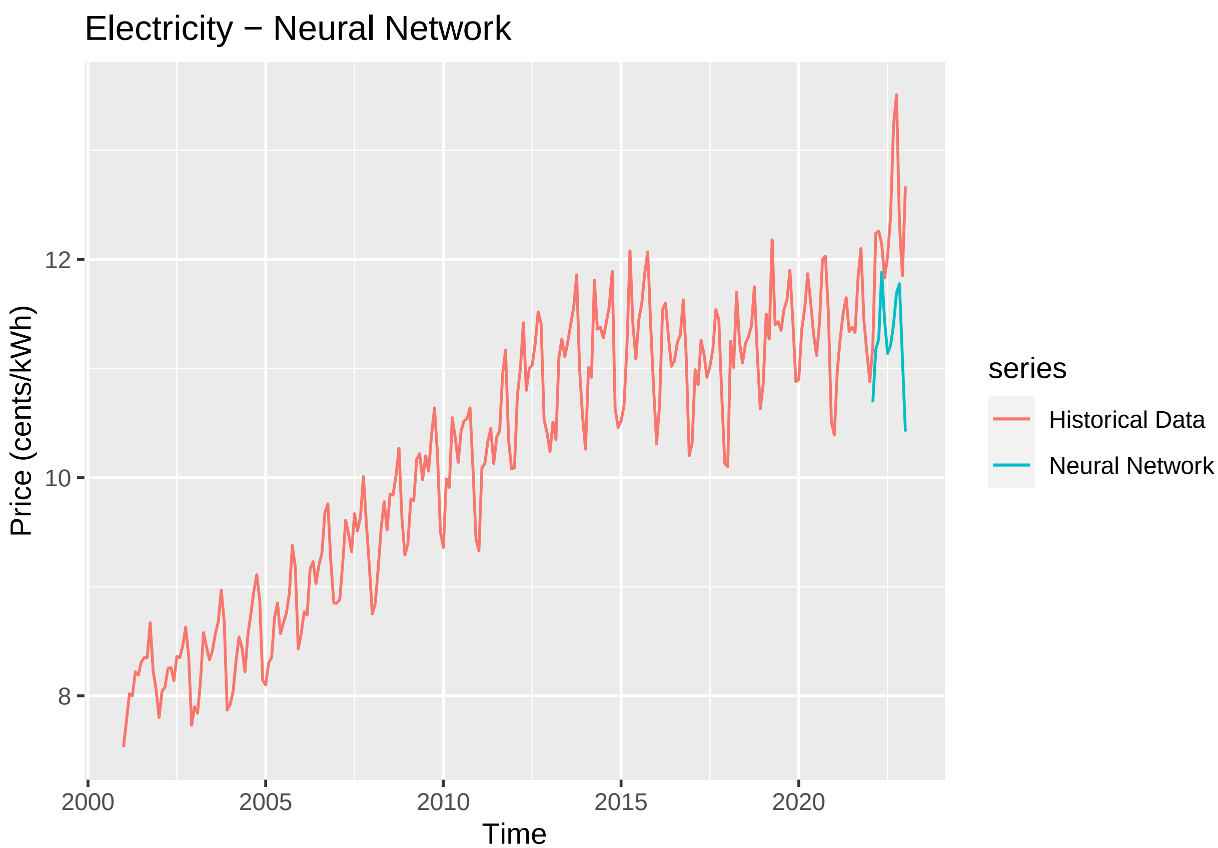
Table 4: Natural Gas Neural Net Forecast Accuracy using Various Seasonal Lag Inputs (p/P)

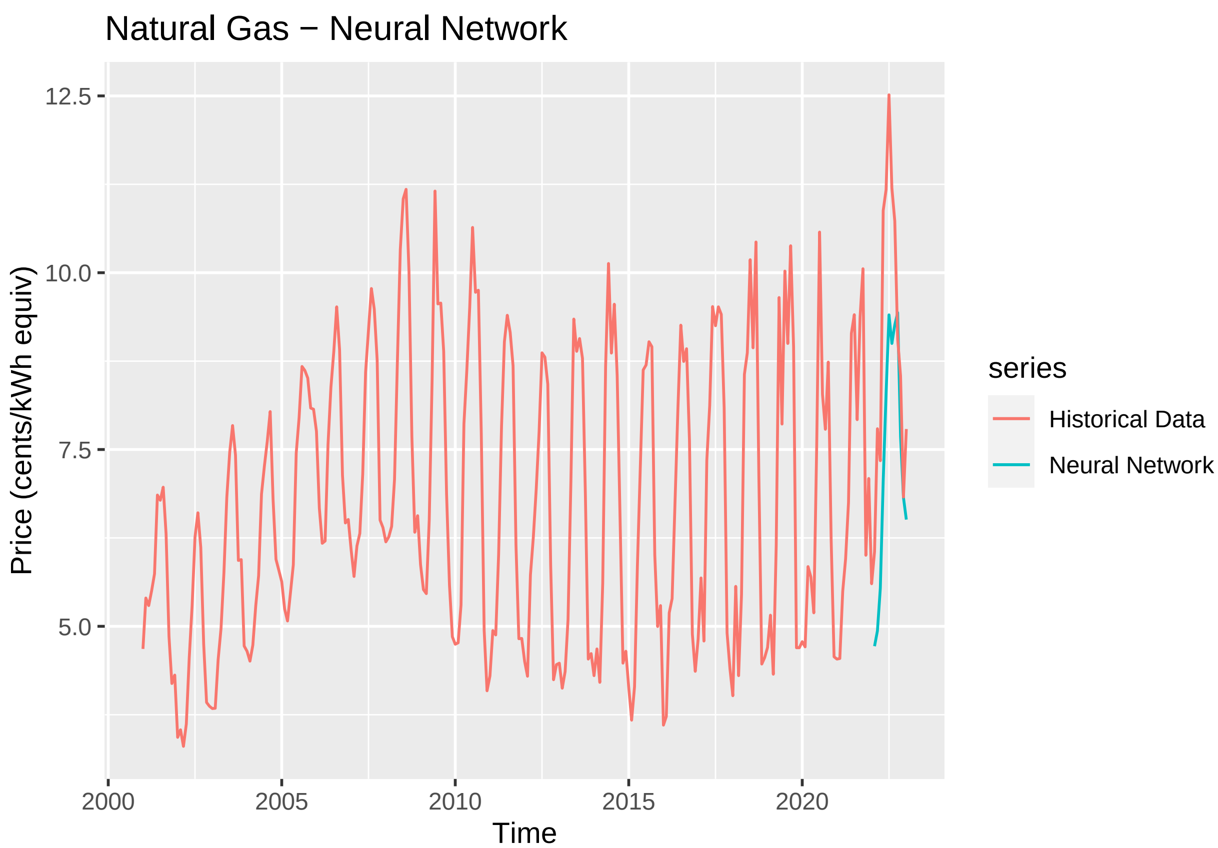
	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
1/0	-1.17116	1.29506	1.17116	-10.50129	10.50129	0.28456	45.32644
1/1	-0.87636	0.98886	0.87636	-7.63924	7.63924	0.18596	4.74519
2/0	-1.00476	1.13652	1.00476	-8.86944	8.86944	0.29360	17.40876
2/1	-0.88838	1.00382	0.88838	-7.75659	7.75659	0.21024	4.74951
2/2	-0.87296	0.99692	0.87296	-7.61877	7.61877	0.22012	4.53454
1/2	-0.87369	0.99258	0.87369	-7.62059	7.62059	0.19242	4.55570
3/1	-0.88305	1.00140	0.88305	-7.70706	7.70706	0.22418	4.79807

Table 5: Natural Gas Neural Net Forecast Accuracy using Various Seasonal Lag Inputs (p/P)

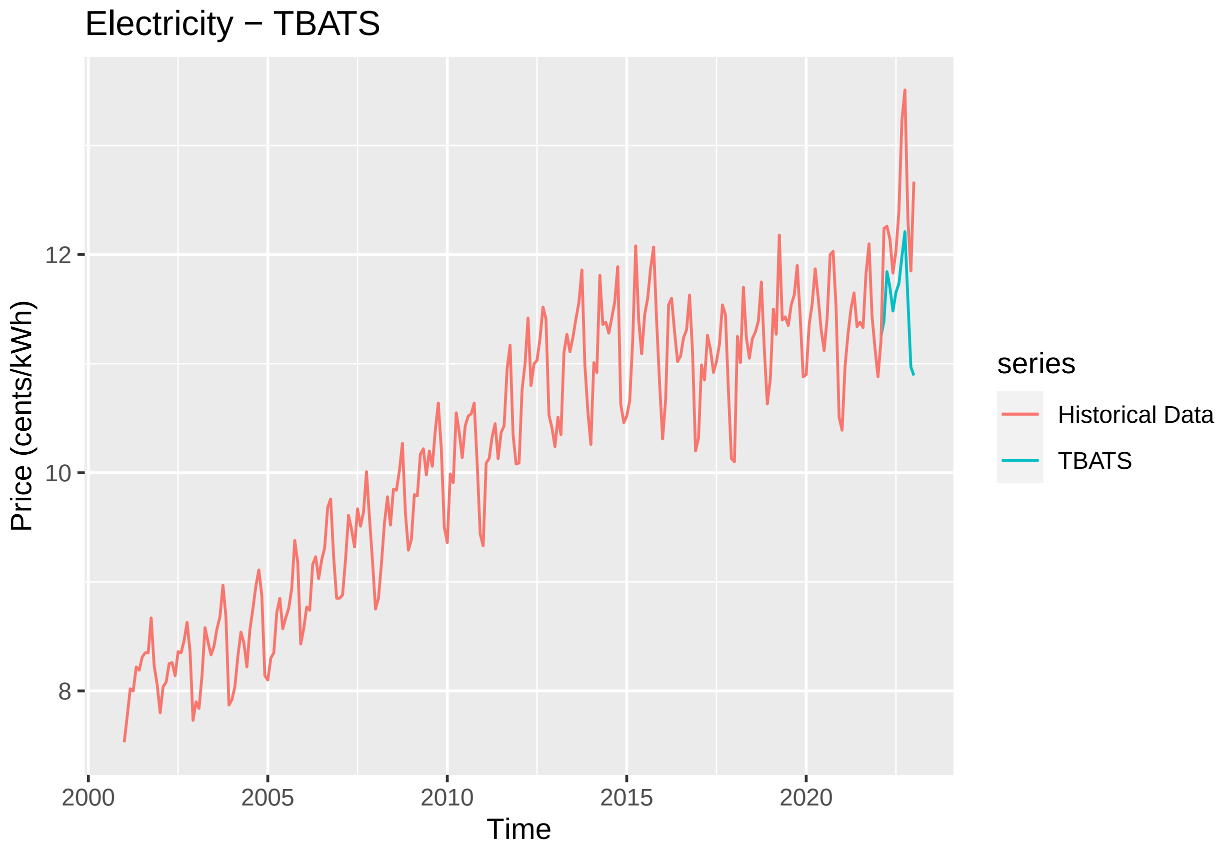
	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
1/0	-2.29006	3.00261	2.39503	-33.52210	34.92963	0.62858	18.47118
1/1	-1.79327	2.25683	1.98281	-25.67527	27.94876	0.30213	2.47245
2/0	-2.46216	3.21259	2.62514	-38.98366	41.07177	0.67079	11.16199
2/1	-1.88880	2.22005	1.96516	-27.38375	28.18552	0.22939	2.51696
2/2	-2.03080	2.36929	2.08233	-30.31031	30.85998	0.28152	2.70754
1/2	-1.92003	2.32631	1.98496	-27.92407	28.61083	0.27618	2.60885
3/1	-1.90816	2.24818	1.96377	-28.05590	28.64755	0.27111	2.57534
3/2	-1.93187	2.26983	1.99013	-28.59361	29.21249	0.28097	2.61144
33	-2.06286	2.38445	2.15201	-31.18643	32.11504	0.23250	2.60344

We found that the combination 11 (p = 1 and P = 1) has the best modeling performance for electricity series, and the combination 32 (p = 3 and P = 2) has the best modeling performance for natural gas series. Then we use this combination to run the neural network model with fourier terms for elecricity series.

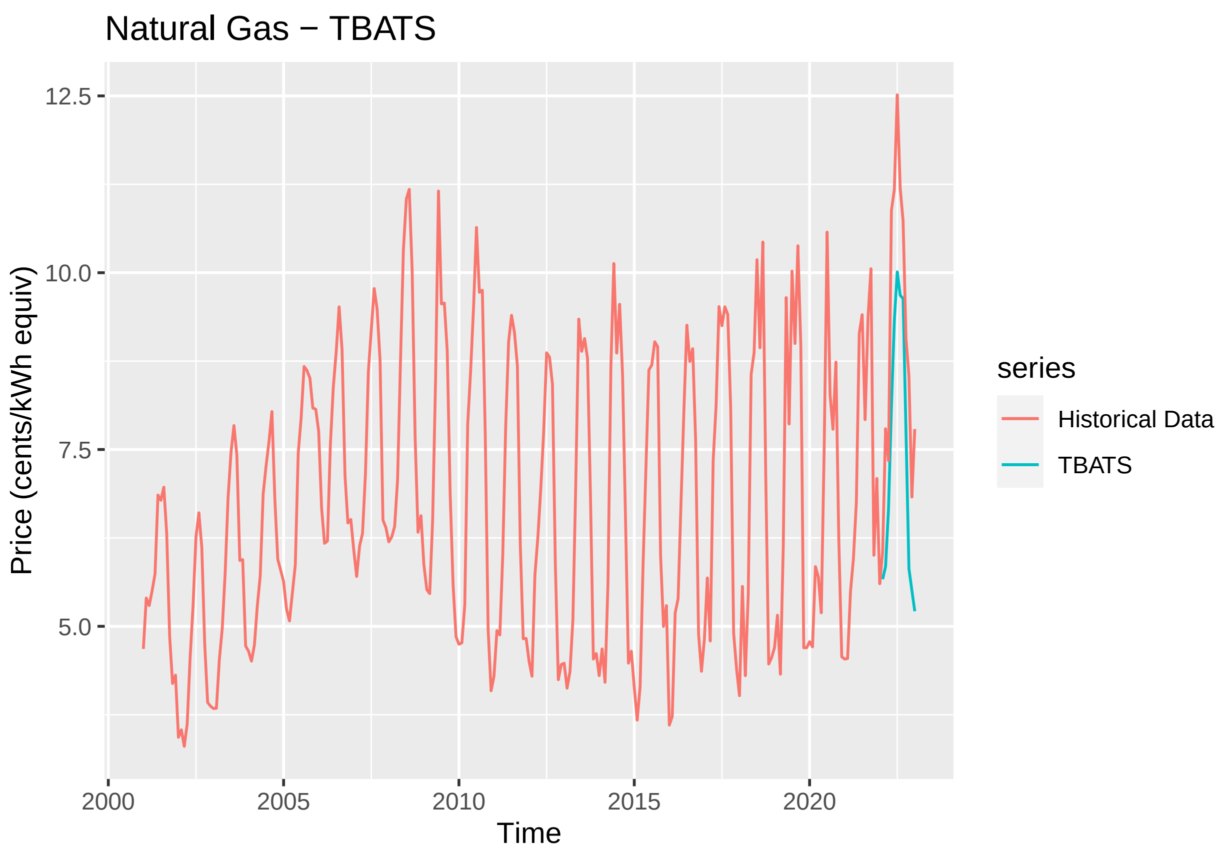




Then we tried TBATS models for electricity and natural gas series

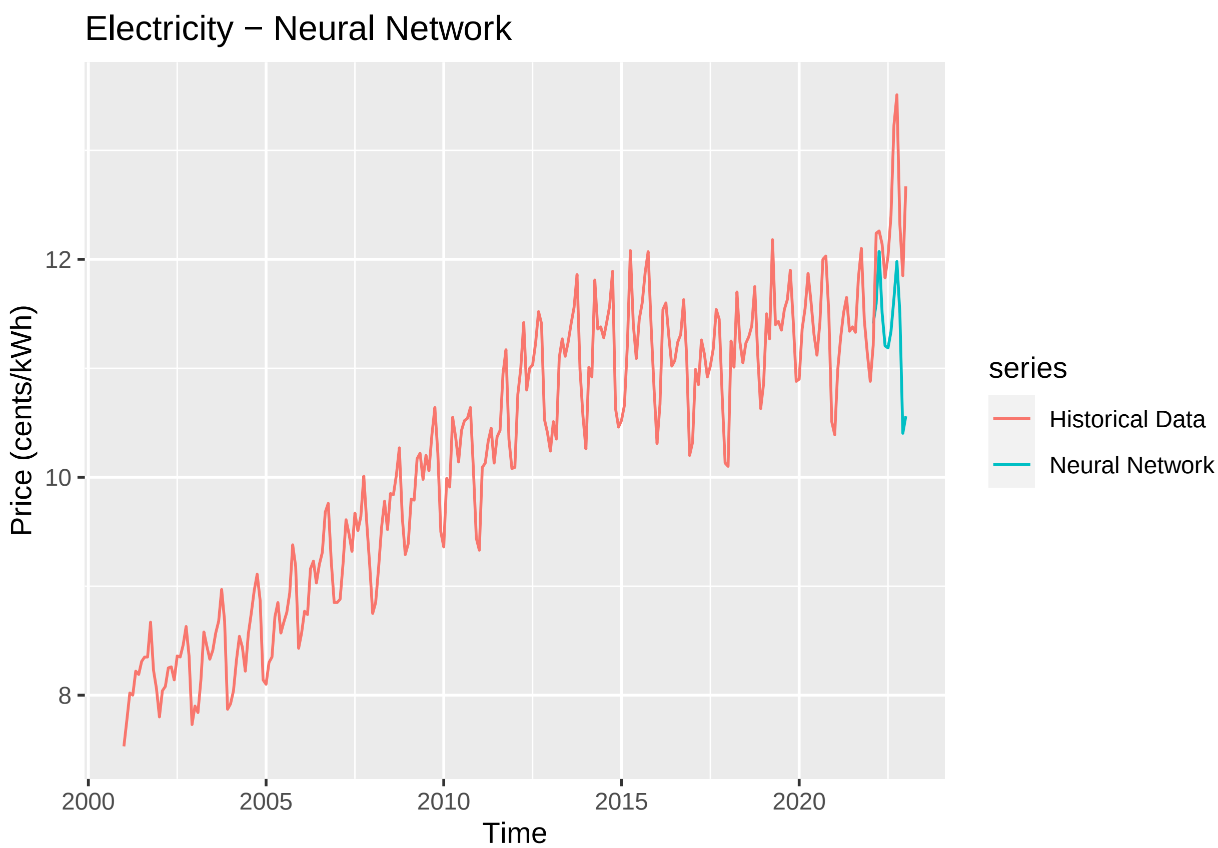


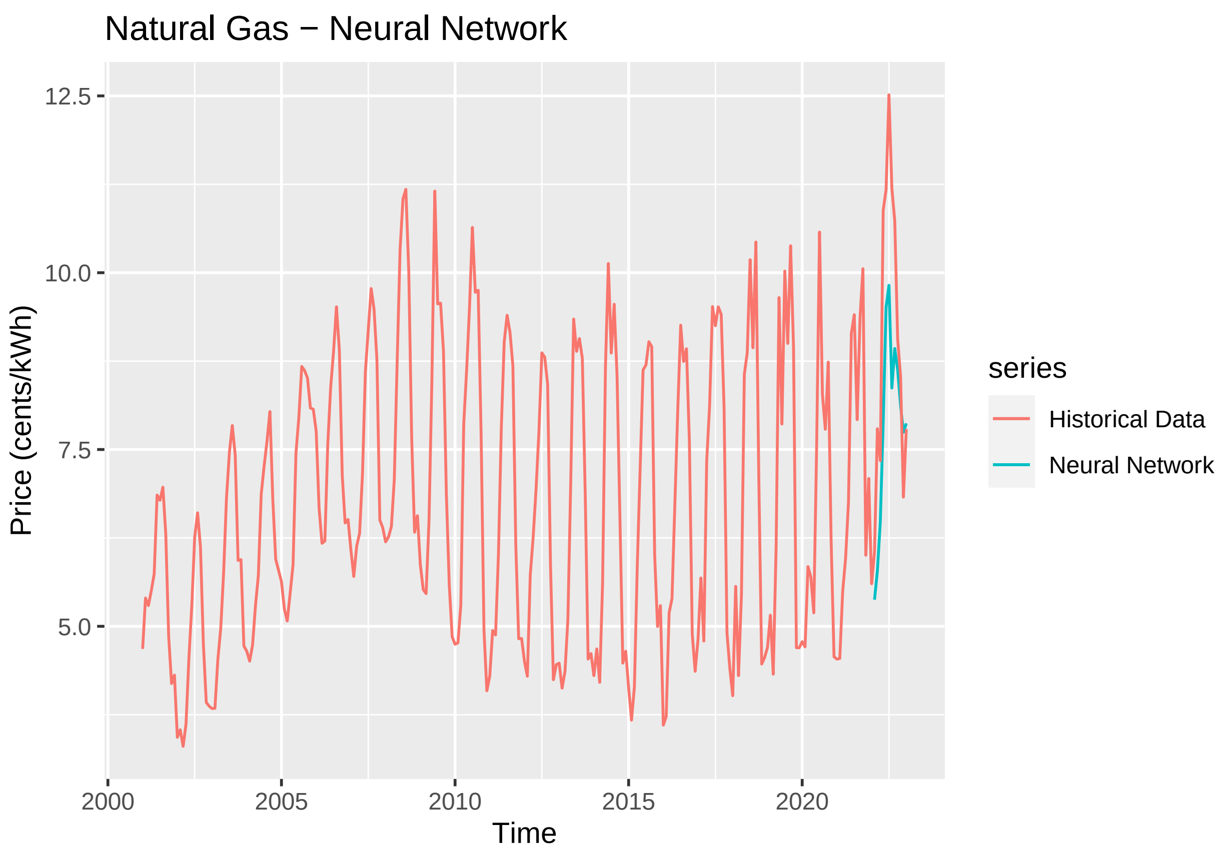
Electricity − TBATS
Price (cents/kWh)
12
10
8
2000
2005
2010
2015
2020
Time
series
Historical Data
TBATS



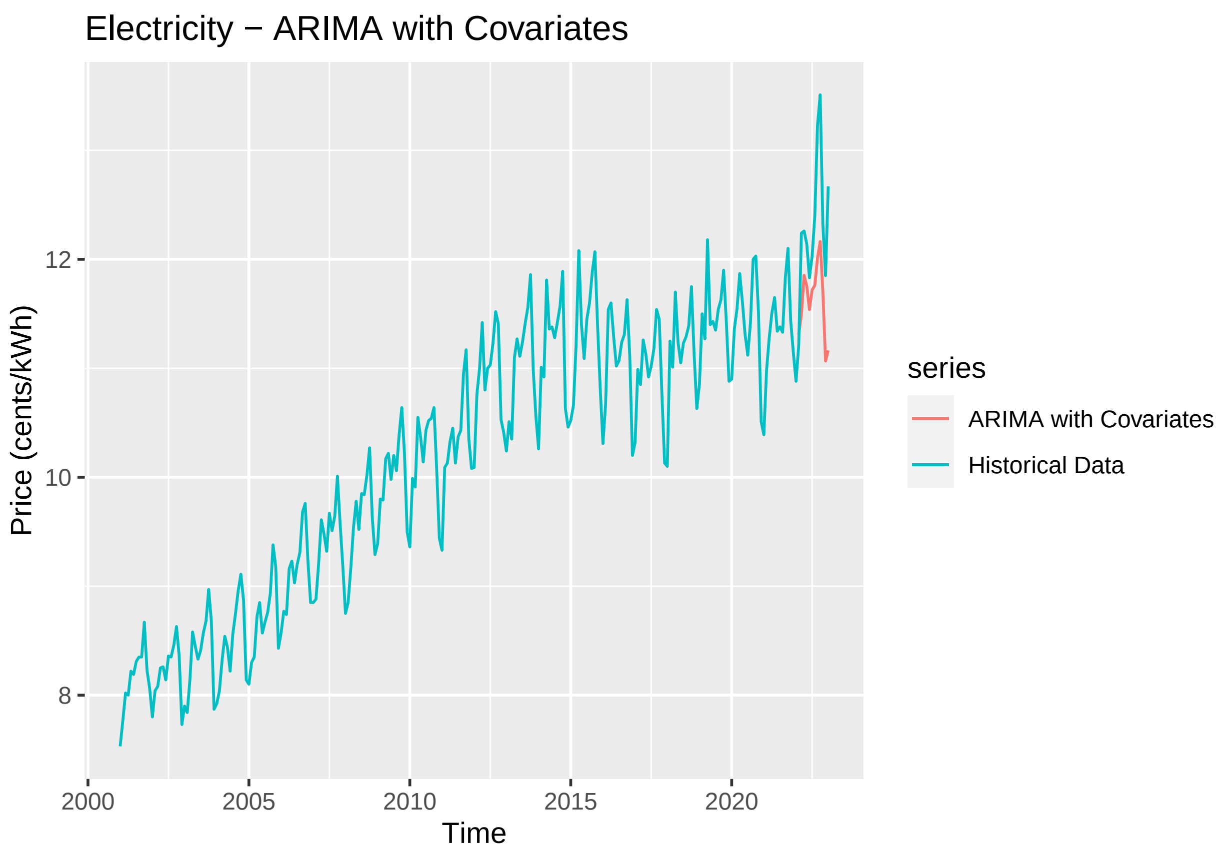
Then we think Ukraine War should have a significant impact on natural gas price and probably eletricity gas too. Also, temperature should be a good regressor to include since utility bills normally fluctuate in the same direction with temperature. Therefore, we created two covariates: UKRWAR and temperature. $UKRWAR$ is an indiciator variable with values of 0 and 1. Months before March 2022 have a value of 0, while months after and including March 2022 have a value of 1. The reason why we set the cutoff month at March 2002 despite the war started from last February is because the impact of the war on monthly natural gas price in February 2022 should be limited since the war started in late February. The temperature series is the monthly average temperature of Raleigh area. This is largest geographic level of historical temperature data.

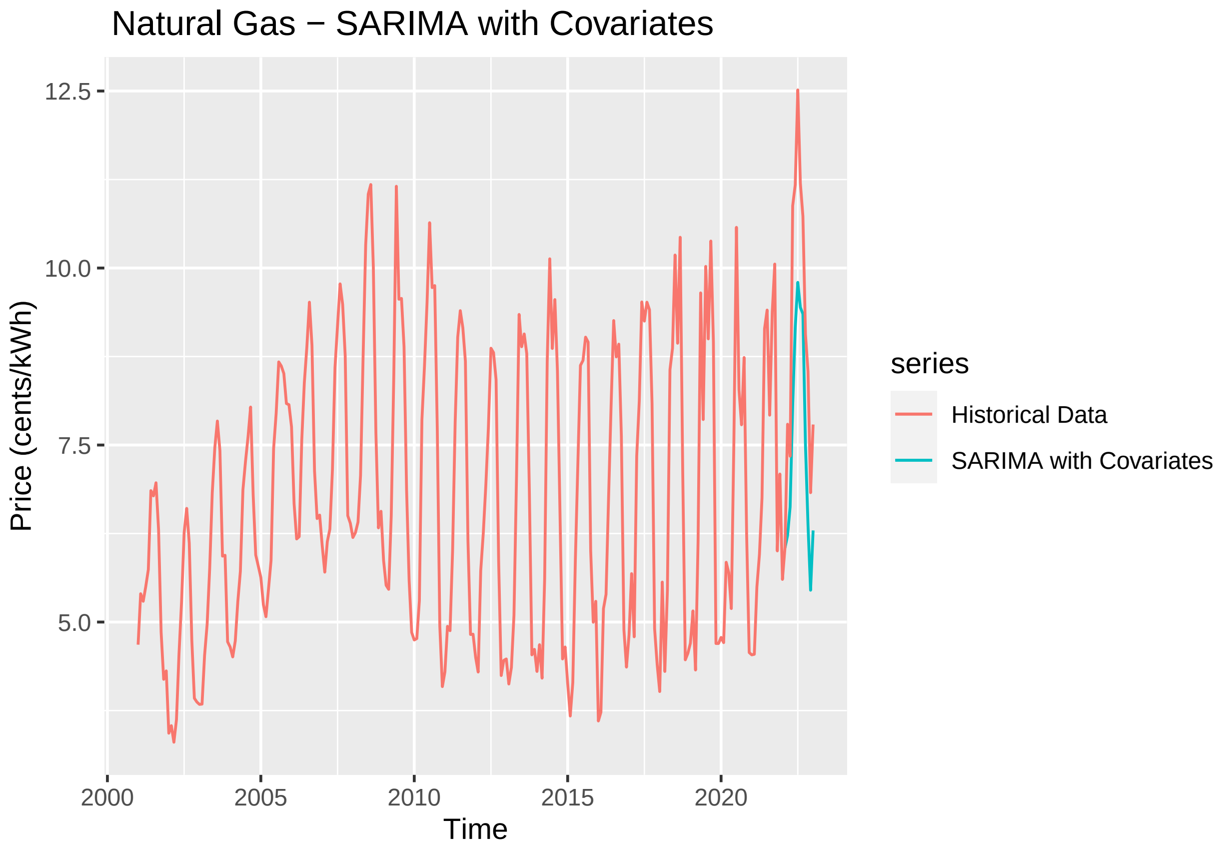
After creating all the covariates, we repeated our modeling but with covariates to improve the accuracy of the models. First we incorporated covariates to neural network model.


Electricity − Neural Network
Price (cents/kWh)
12
10
8
2000
2005
2010
2015
2020
Time
series
Historical Data
Neural Network



Then we use seasonal arima model with temperature and fourier terms to model two series. UKRWAR is excluded because R reports no suitable ARIMA model when UKRWAR is included. Function used: auto.arima(xreg)

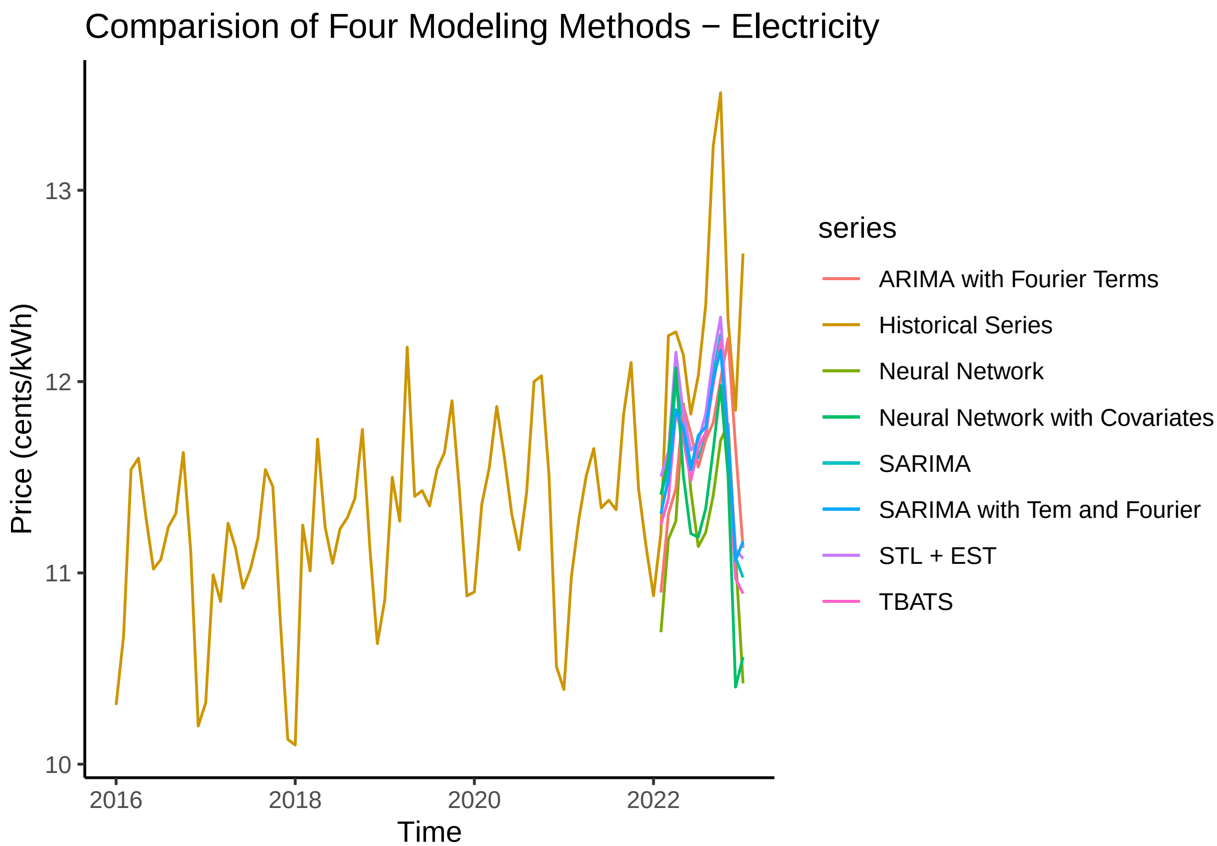

Electricity − ARIMA with Covariates
Price (cents/kWh)
12
10
8
2000
2005
2010
2015
2020
Time
series
ARIMA with Covariates
Historical Data

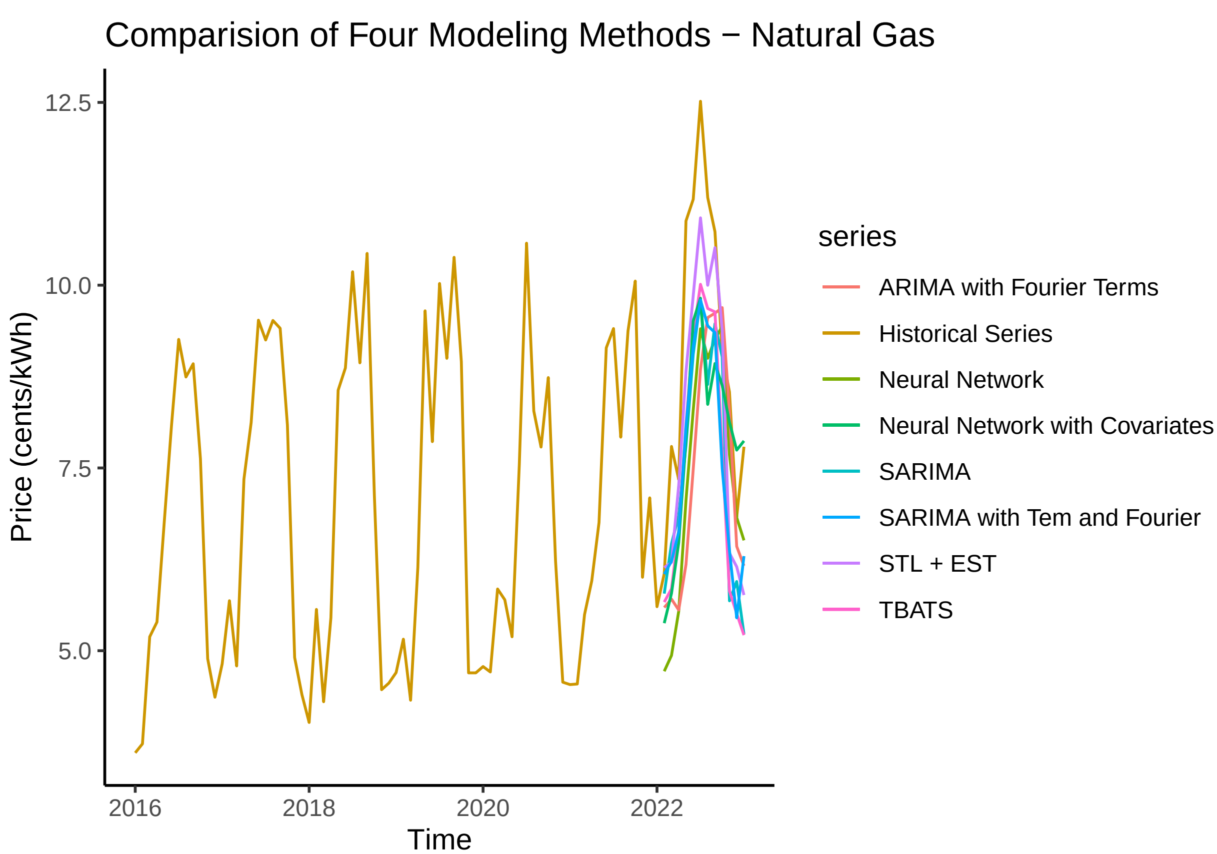


##Summary and Conclusions

Use ETS to model – not finished – I don't know how to do this, maybe we don't need to include this

compare performance scores and generate tables for use





```
## The best model for electricity by RMSE is:  STL
```

Table 6: Forecast Accuracy for NC Residential Electricity Price

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
SARIMA	-0.67702	0.83764	0.70998	-5.83726	6.12612	0.25988	2.73983
ARIMA with Fourier	-0.69901	0.87542	0.69901	-6.04750	6.04750	0.08222	2.82856
STL	-0.58551	0.76708	0.63447	-5.01390	5.43949	0.25835	2.36056
Neural Network	-1.04198	1.20332	1.04198	-9.33410	9.33410	0.17119	2.98545
TBATS	-0.74971	0.89063	0.75658	-6.50522	6.56626	0.23462	2.86546
Neural Network with Covariates	-0.94051	1.12537	0.97386	-8.39474	8.68702	0.39768	2.63993
SARIMA with Tem and Fourier	-0.67930	0.81795	0.69585	-5.83257	5.97890	0.26220	2.92127

```
## The best model for natural gas by RMSE is:  STL
```

Table 7: Forecast Accuracy for NC Residential Natural Gas Price

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
SARIMA	-1.67902	1.97784	1.67902	-23.35174	23.35174	-0.15535	1.81675
ARIMA with Fourier	-1.73517	2.31199	1.83980	-25.17030	26.24979	0.57850	3.12540
STL	-1.04788	1.35599	1.11314	-14.03428	14.80904	-0.33348	1.28849
Neural Network	-1.77145	2.14734	1.83338	-25.67612	26.33247	0.50464	2.70401
TBATS	-1.72858	1.89245	1.72858	-24.53539	24.53539	-0.32231	2.08610
Neural Network with Covariates	-1.28253	1.74881	1.44855	-16.24018	18.38130	0.46723	2.11881
SARIMA with Tem and Fourier	-1.63275	1.79495	1.63275	-21.52013	21.52013	-0.01071	1.95239

Use STL to model electricity and natural gas for the next 12 month