

Visionary Final

Justin DePue, John Rooney, and Tony Jiang

2023-04-28

Contents

Use ETS to model – not finished – I don’t know how to do this, maybe we don’t need to include this	24
compare performance scores and generate tables for use	24
Use STL to model electricity and natural gas for the next 12 month	26

- 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.

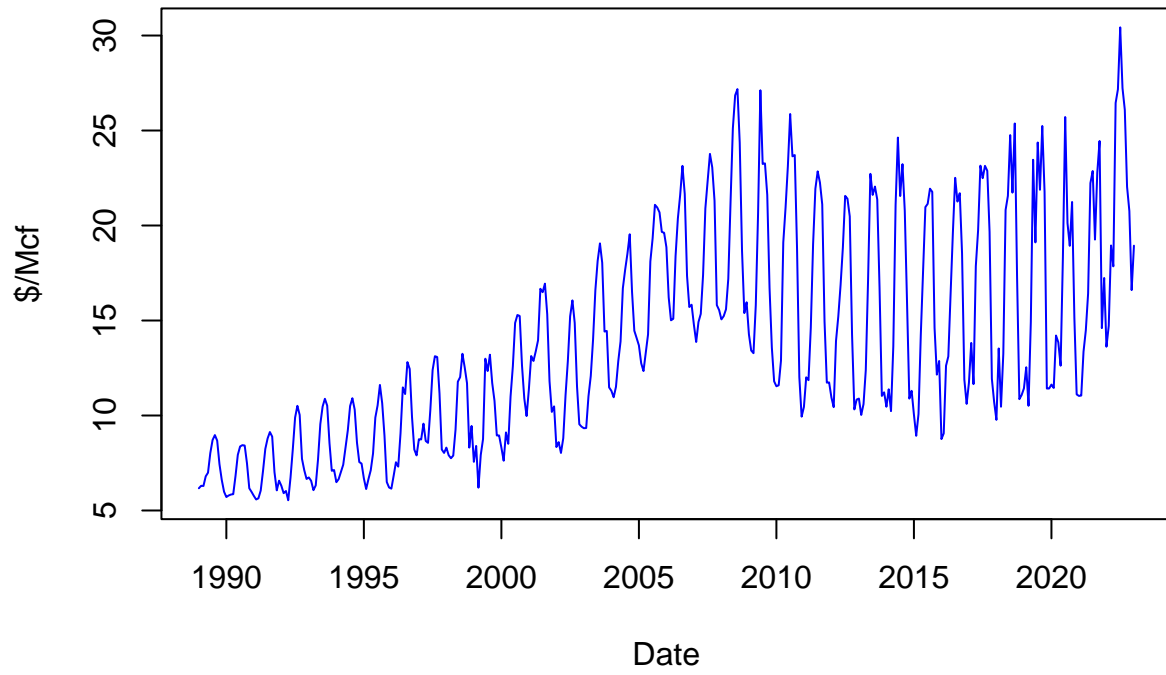
As students recently armed with the tools to conduct time series analysis and forecasting, we realized we could complete our final project while aiding our teammate with knowledge for the future. Our study question thus emerged: would it be more cost-effective to heat an apartment in North Carolina using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

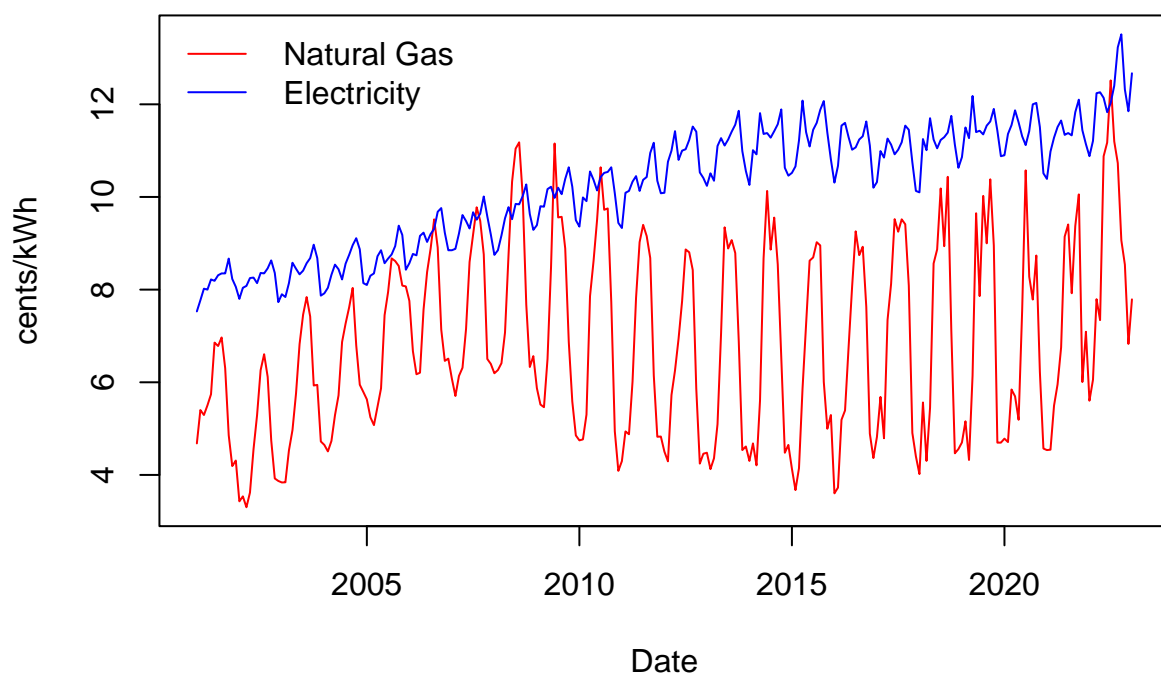
##Data

##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.

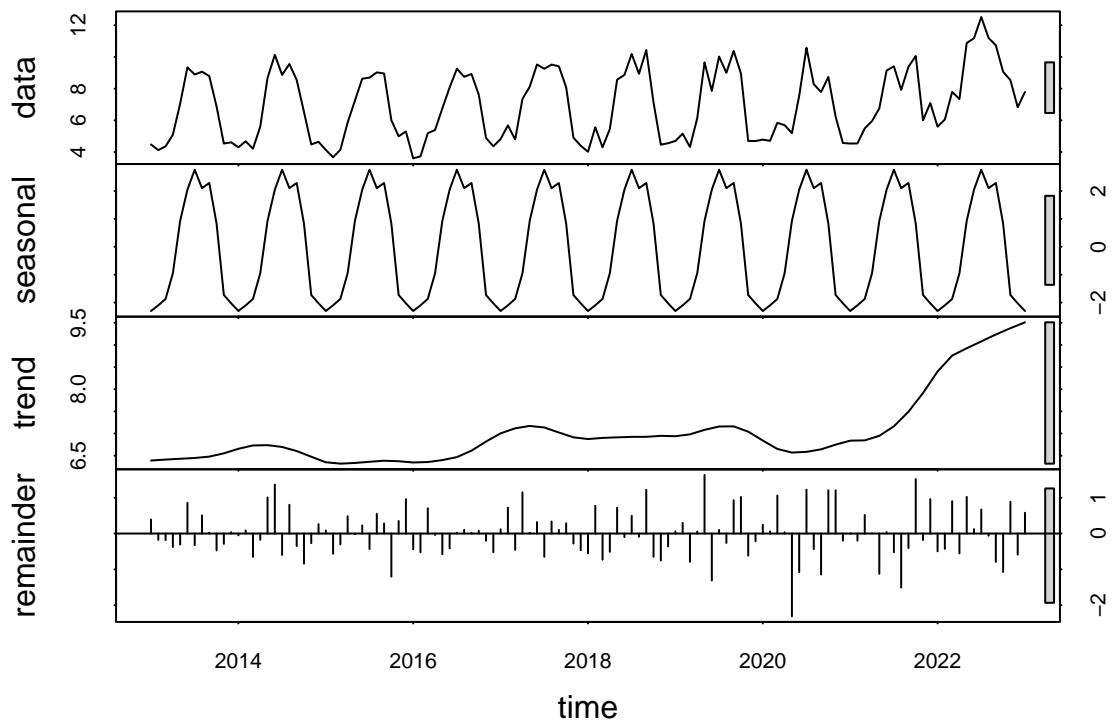
NC Residential Natural Gas Cost



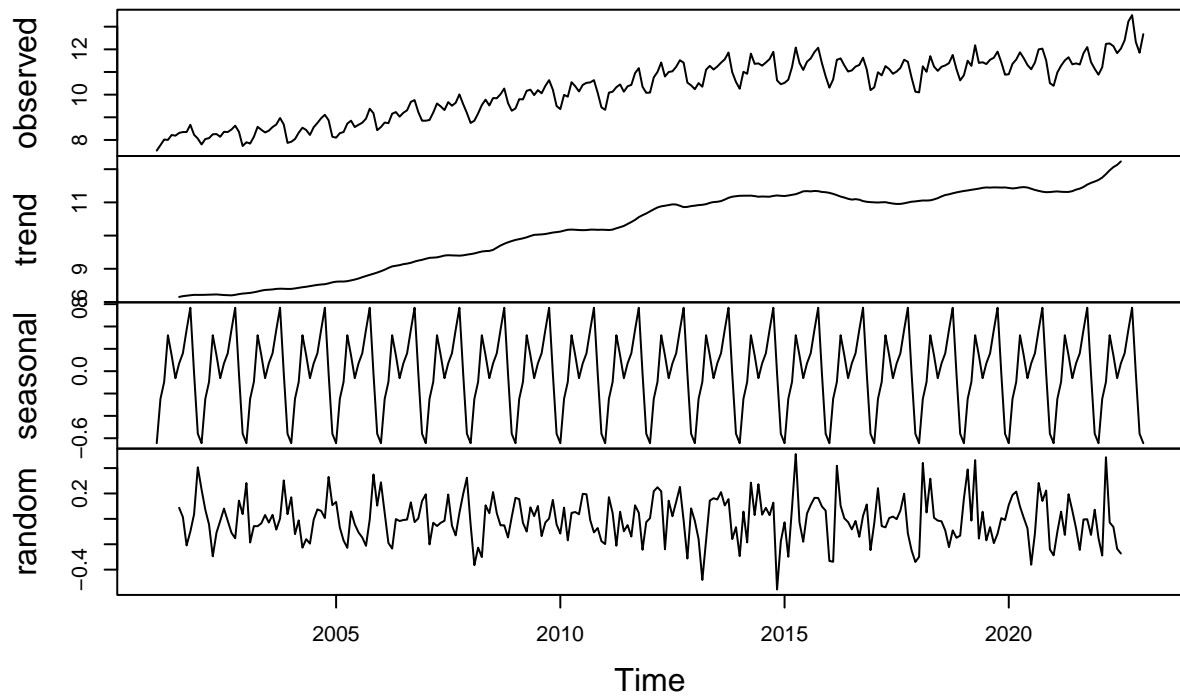
Comparison of Natural Gas and Electricity Costs



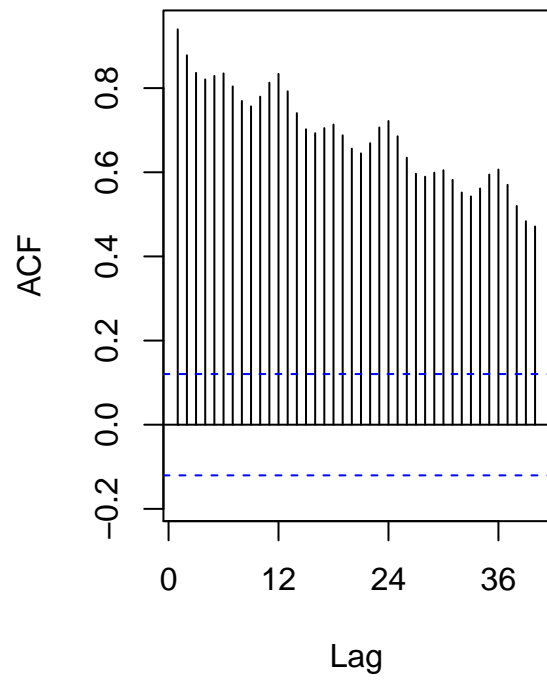
We then decomposed the time series objects for further analysis.



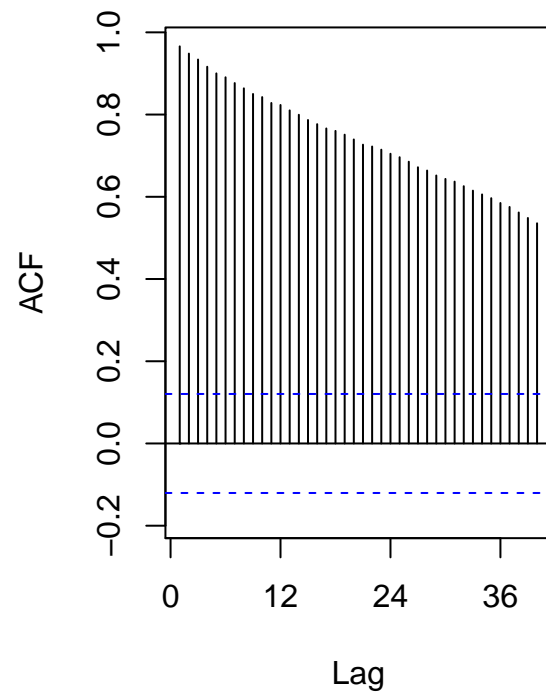
Decomposition of additive time series



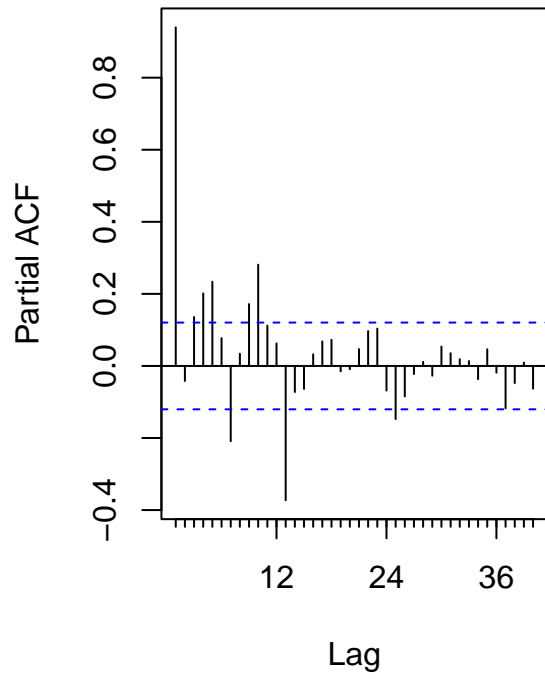
ACF Electricity



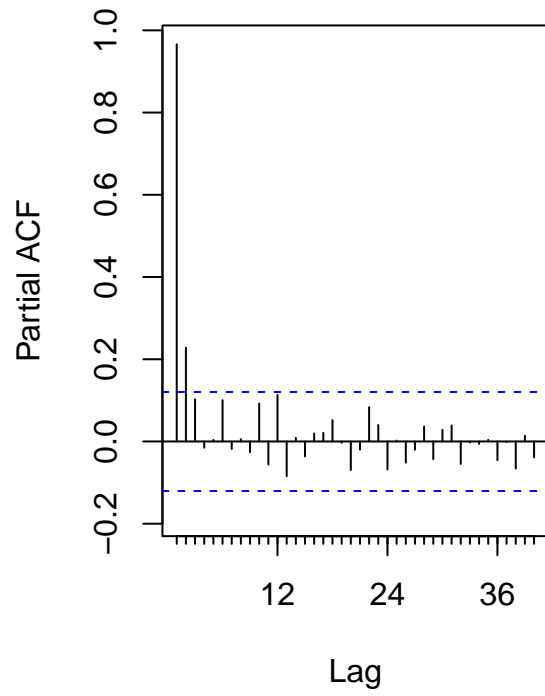
ACF Non-Seasonal Electricity



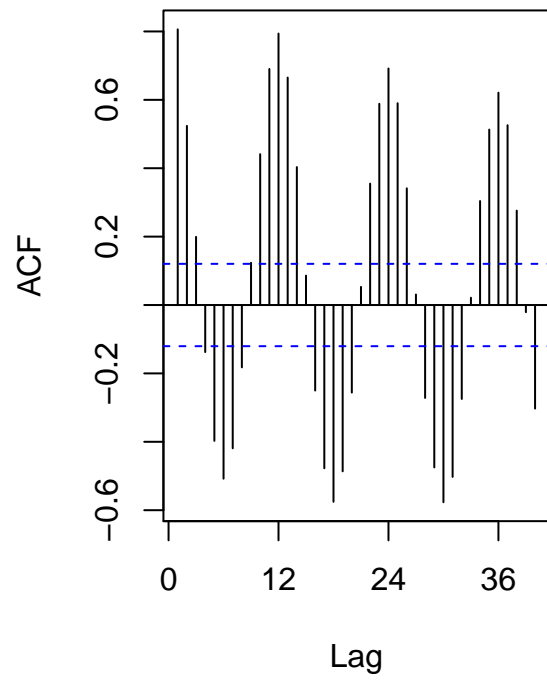
PACF Electricity



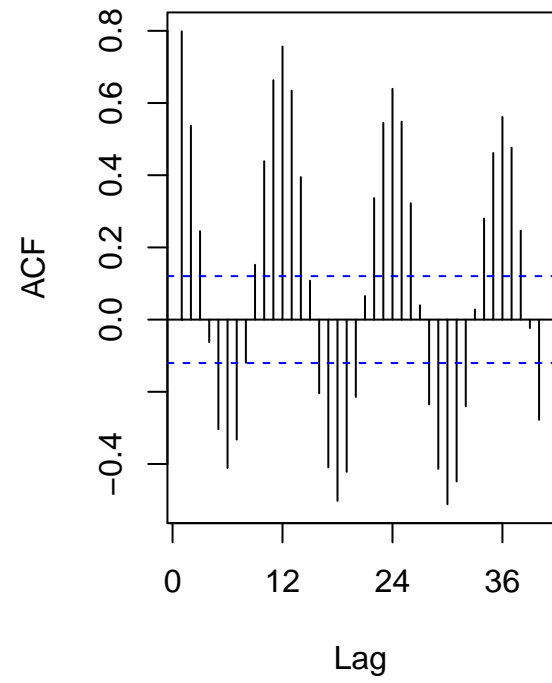
PACF Non-Seasonal Electricity

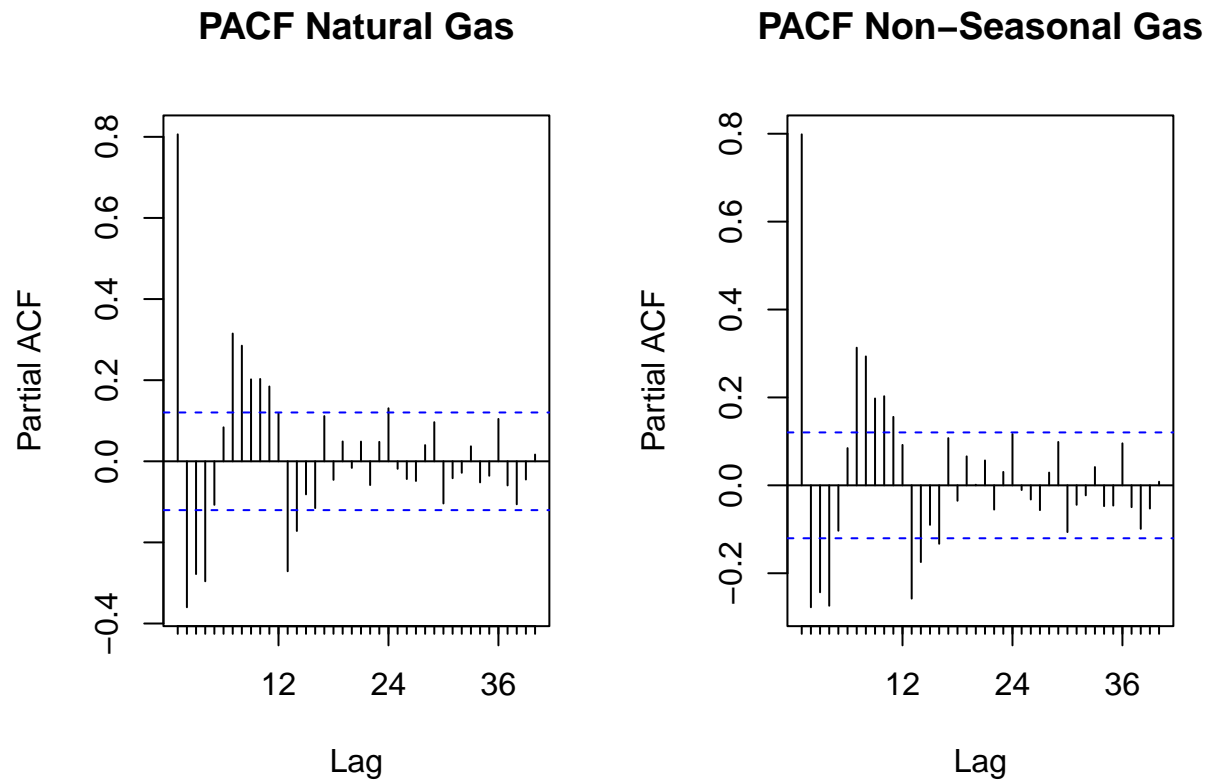


ACF Natural Gas



ACF Non-Seasonal Gas



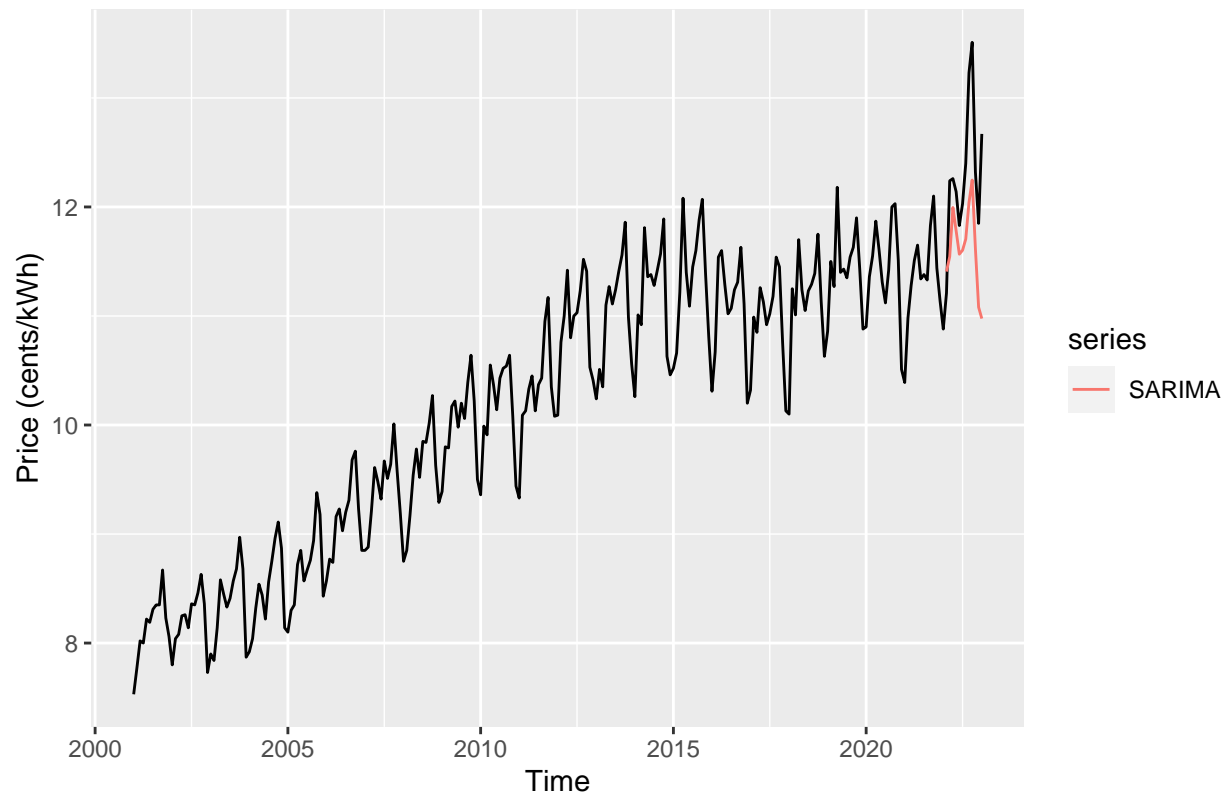


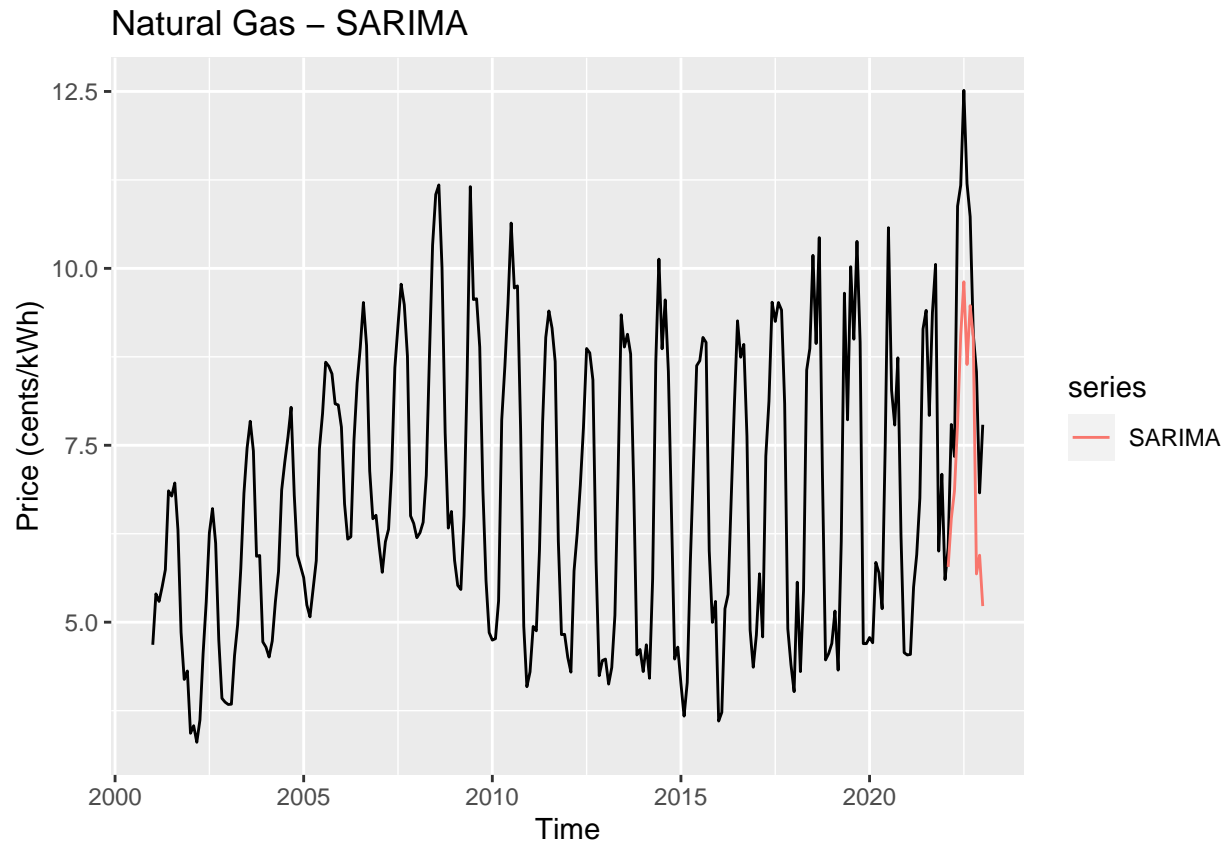
Several models were developed and tested to determine what would best fit the data we had and what may lead to the best forecast to determine whether heating via electricity or natural gas would be most cost-efficient in the upcoming winter. The five that were used were the Seasonal ARIMA, ARIMA with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

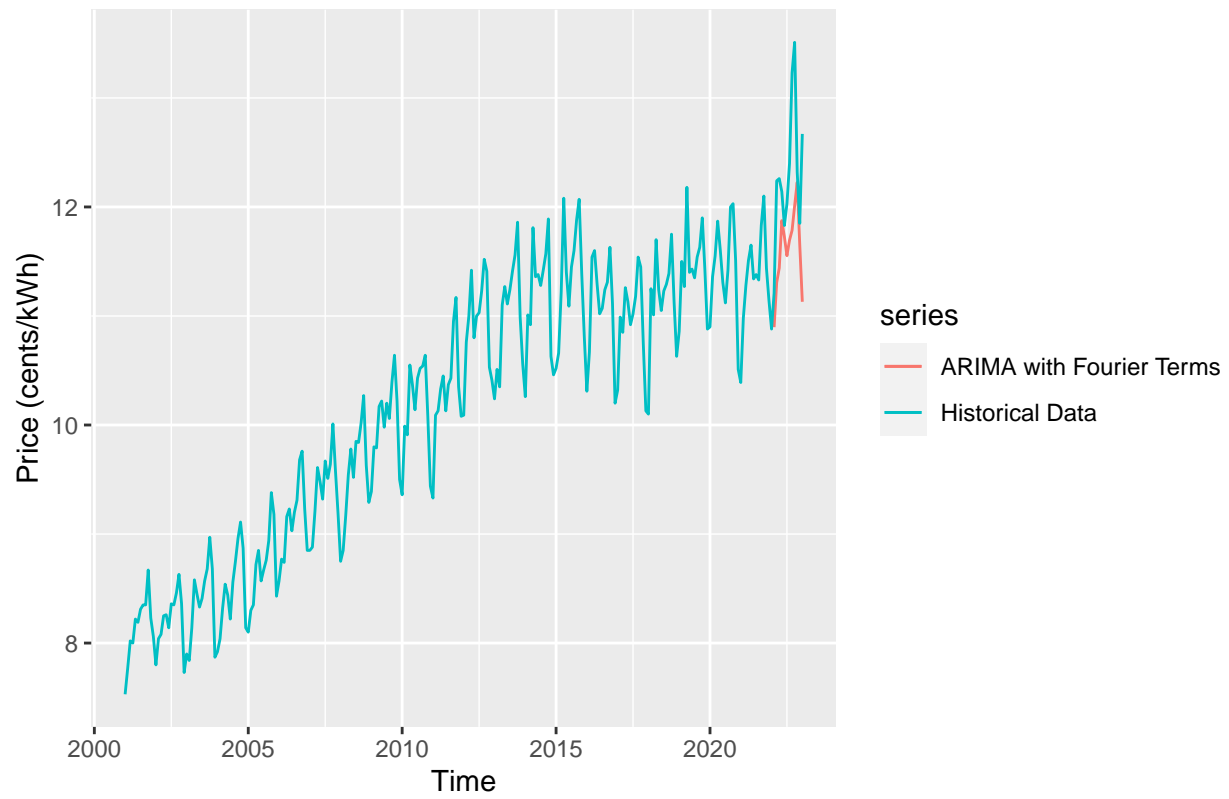
Electricity – SARIMA

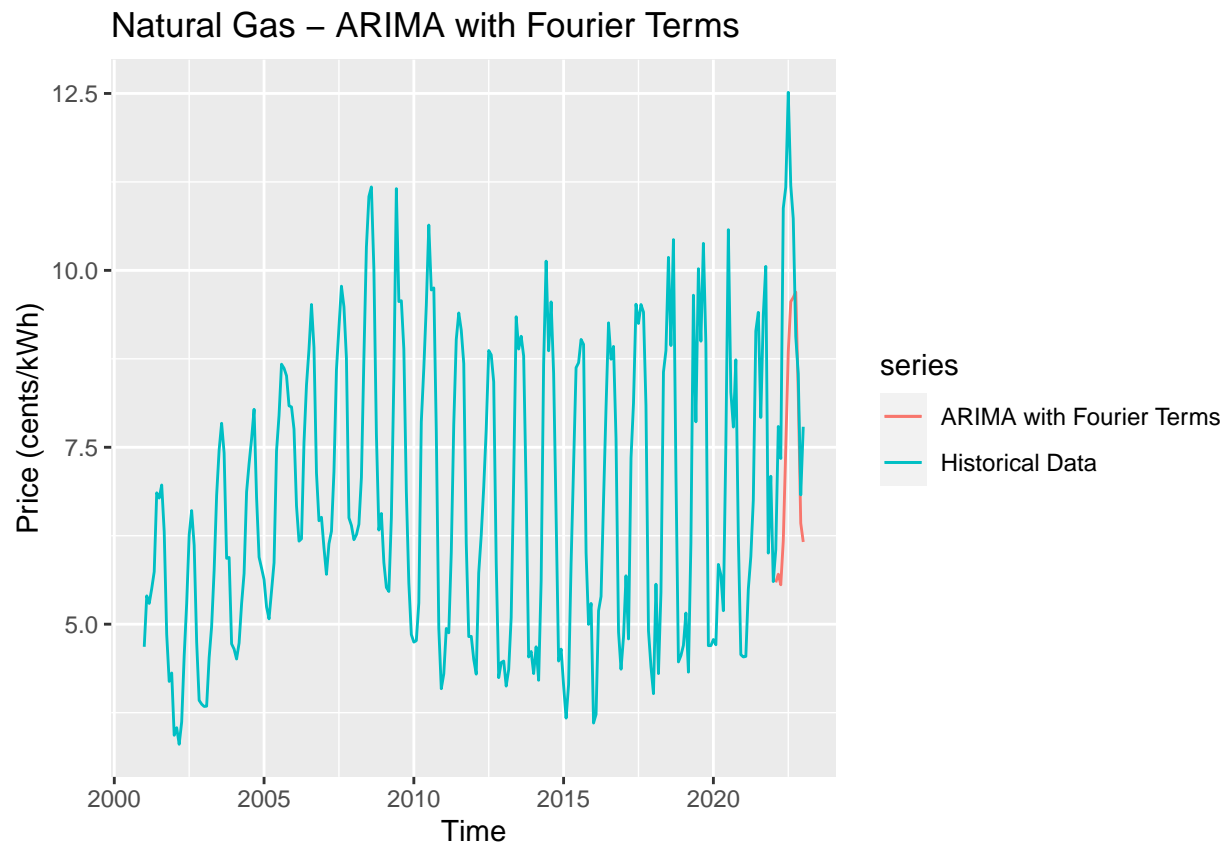




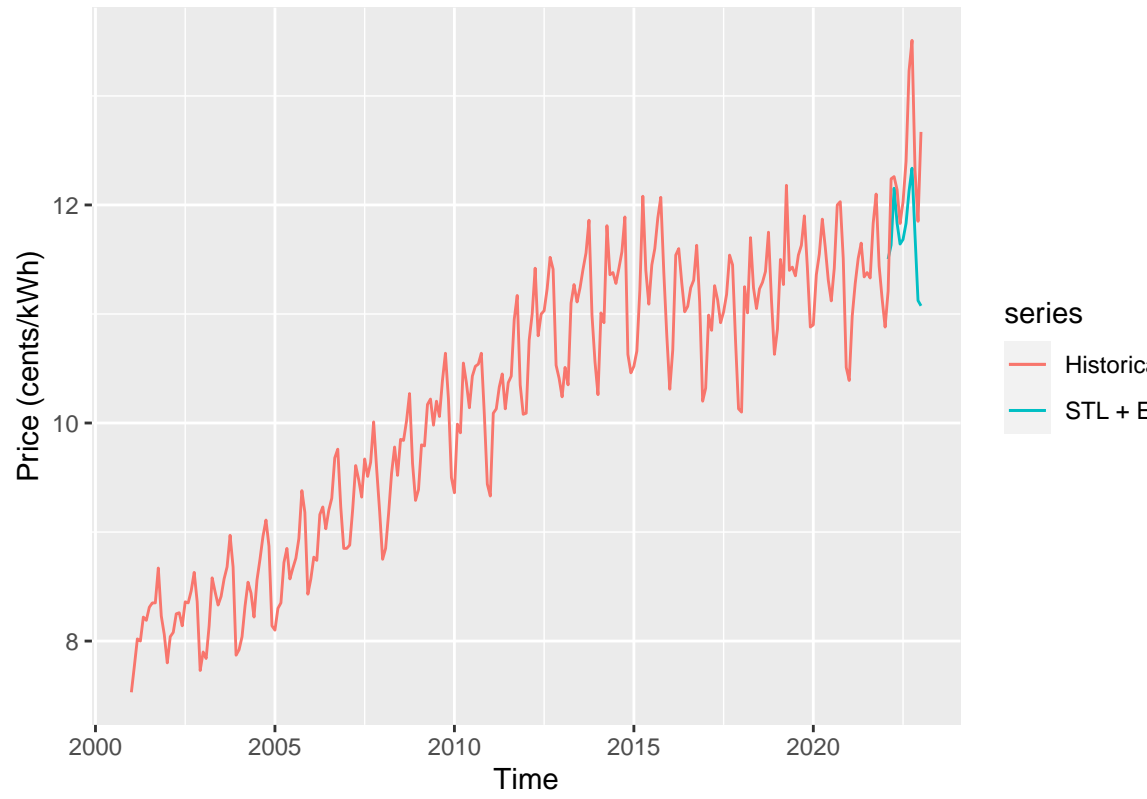
We decided to explore more advanced models that could handle the ARIMA with Fourier Terms model, Neural Networks, TBATS, and STL + EST. ARIMA with Fourier terms is known as a dynamic harmonic regression model with an ARMA error structure, using the “fourier” function from package “forecast” to find terms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

Electricity – ARIMA with Fourier Terms

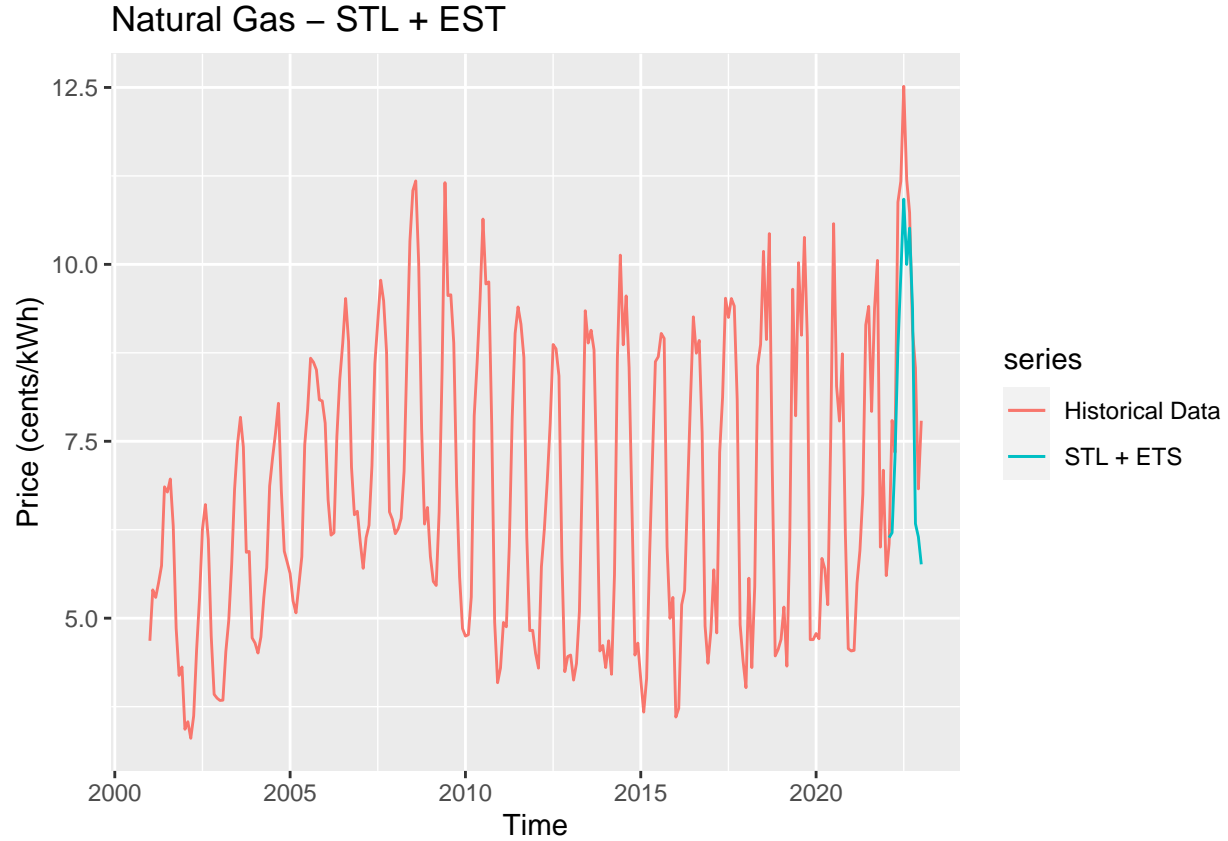




Electricity – STL + EST



Then we tried STL model



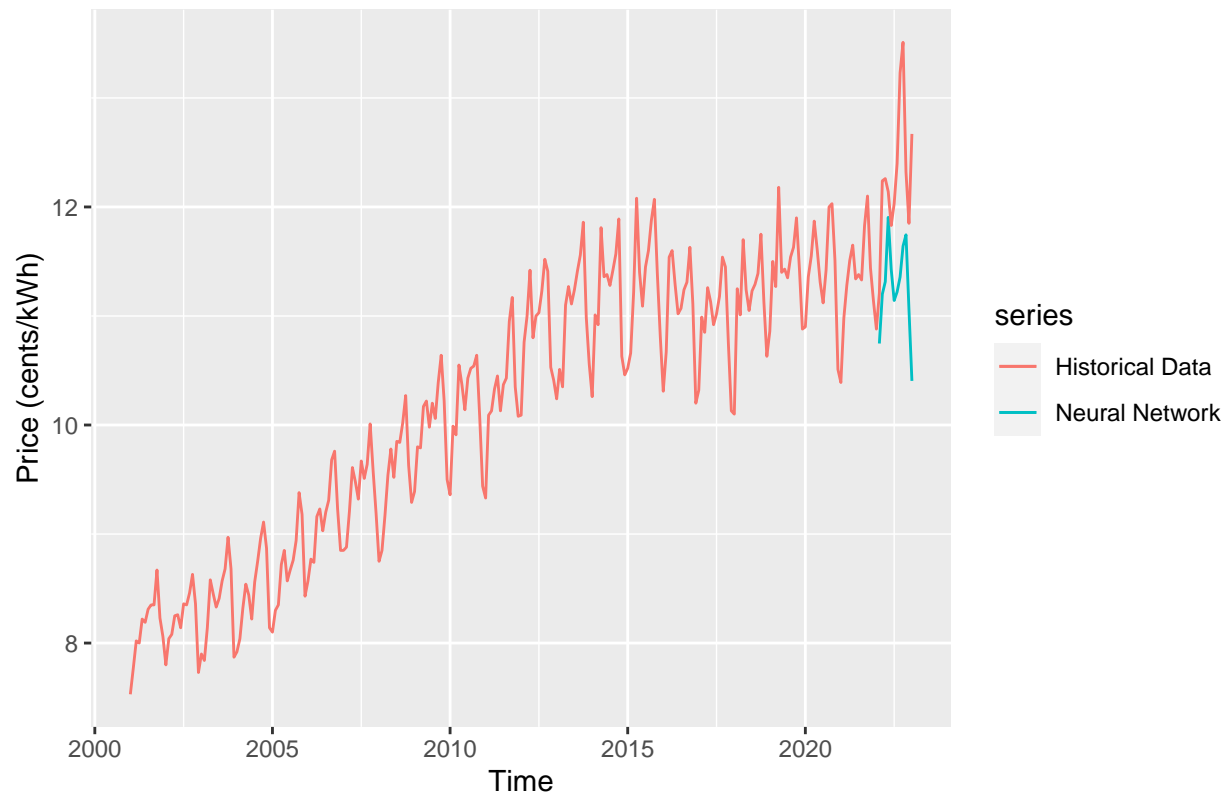
Then we tried to use neural network model. We used `nnetar()` in forecast package. We figured out `p` and `P` arguments in `nnetar()` have significant impact on model performance. Therefore, we first tried to identify the optimal `p` and `P` combination by trying different combinations.

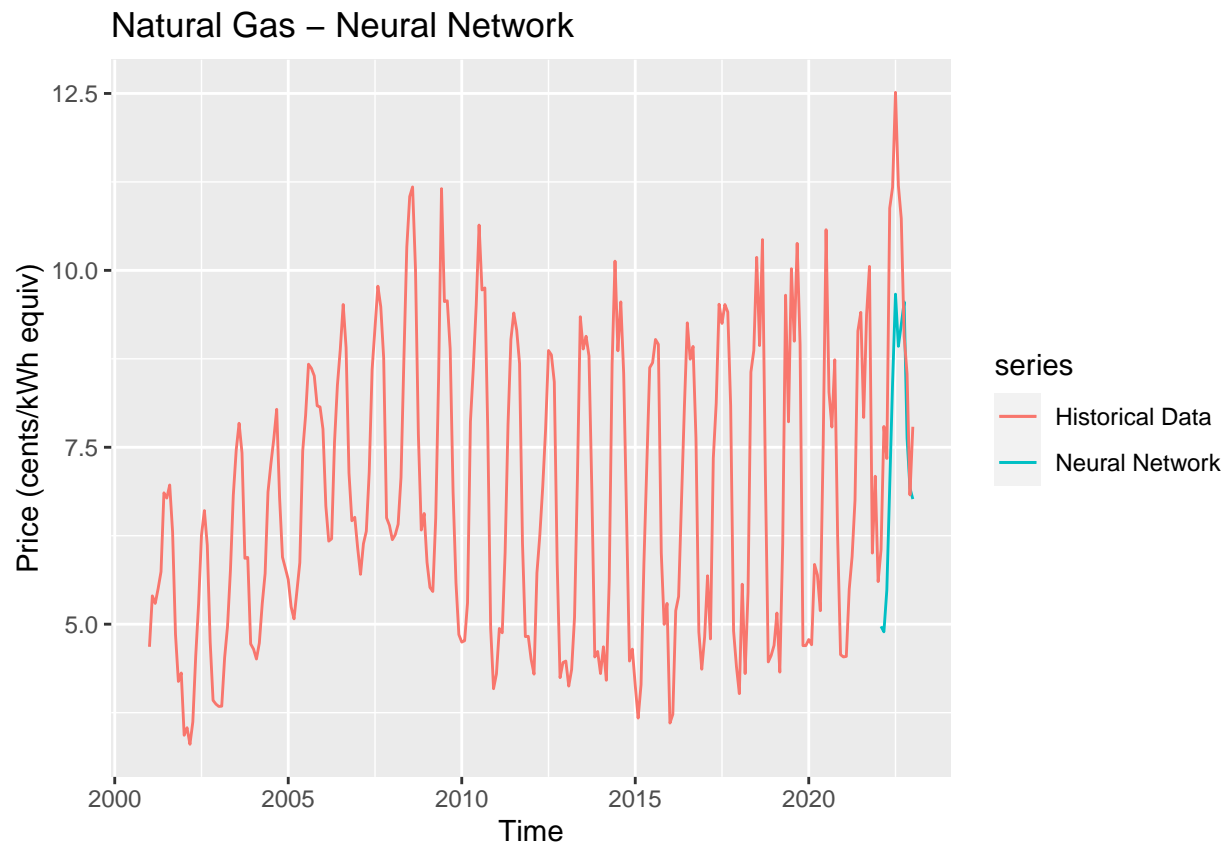
##	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
## 10	-1.1711603	1.2950610	1.1711603	-10.501248	10.501248	0.2845596	45.325630
## 11	-0.8765233	0.9885803	0.8765233	-7.640658	7.640658	0.1849960	4.715329
## 20	-1.0243211	1.1573350	1.0243211	-9.059969	9.059969	0.2983285	19.562196
## 21	-0.8844563	1.0002431	0.8844563	-7.718279	7.718279	0.2105635	4.838872
## 22	-0.8693153	0.9960095	0.8693153	-7.586973	7.586973	0.2289934	4.511358
## 12	-0.8740881	0.9953082	0.8740881	-7.625628	7.625628	0.2009892	4.692610
## 31	-0.8961983	1.0137286	0.8961983	-7.833745	7.833745	0.2274143	4.668464

##	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
## 10	-2.290261	3.002885	2.395262	-33.52706	34.93504	0.6286144	18.473851
## 11	-1.797149	2.262919	1.987532	-25.74734	28.03077	0.3115636	2.500644
## 20	-2.482079	3.264903	2.668922	-39.55985	41.93018	0.6738091	12.108573
## 21	-1.907981	2.238994	1.995563	-27.83246	28.74568	0.2309223	2.506992
## 22	-2.013561	2.344736	2.053495	-29.89788	30.32707	0.2528332	2.690566
## 12	-1.917998	2.335665	1.988542	-28.00271	28.74619	0.2842787	2.602672
## 31	-1.949496	2.284456	2.005087	-28.92141	29.51289	0.2961621	2.632223
## 32	-1.921584	2.252297	1.972852	-28.29839	28.84538	0.2631475	2.584061
## 33	-2.036527	2.372362	2.101734	-30.45576	31.14532	0.2511010	2.601636

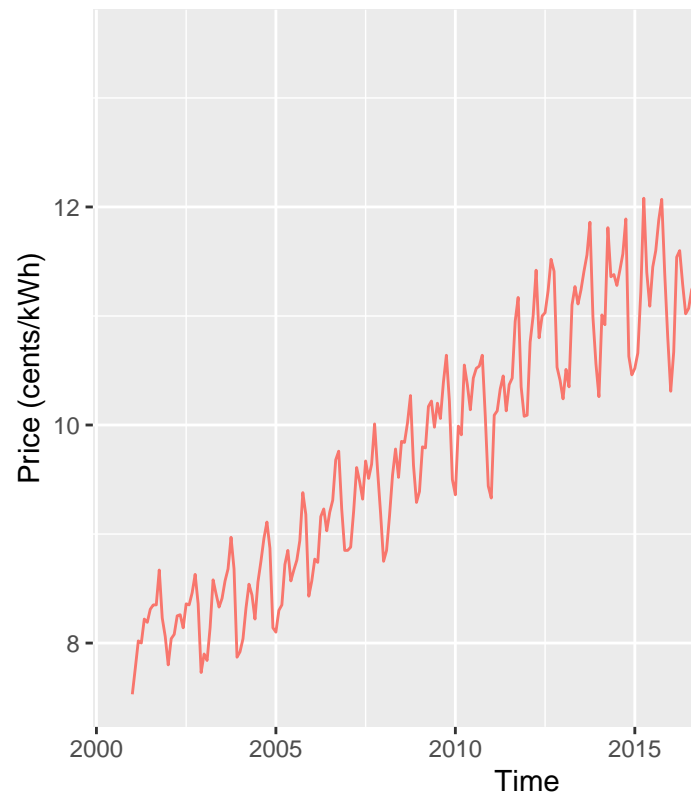
We found that the combination 11 ($p = 1$ and $P = 1$) has the best modeling performance for electricity series, and the combination 32 ($p = 3$ and $P = 2$) has the best modeling performance for natural gas series. Then we use this combination to run the neural network model with fourier terms for electricity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

Electricity – Neural Network

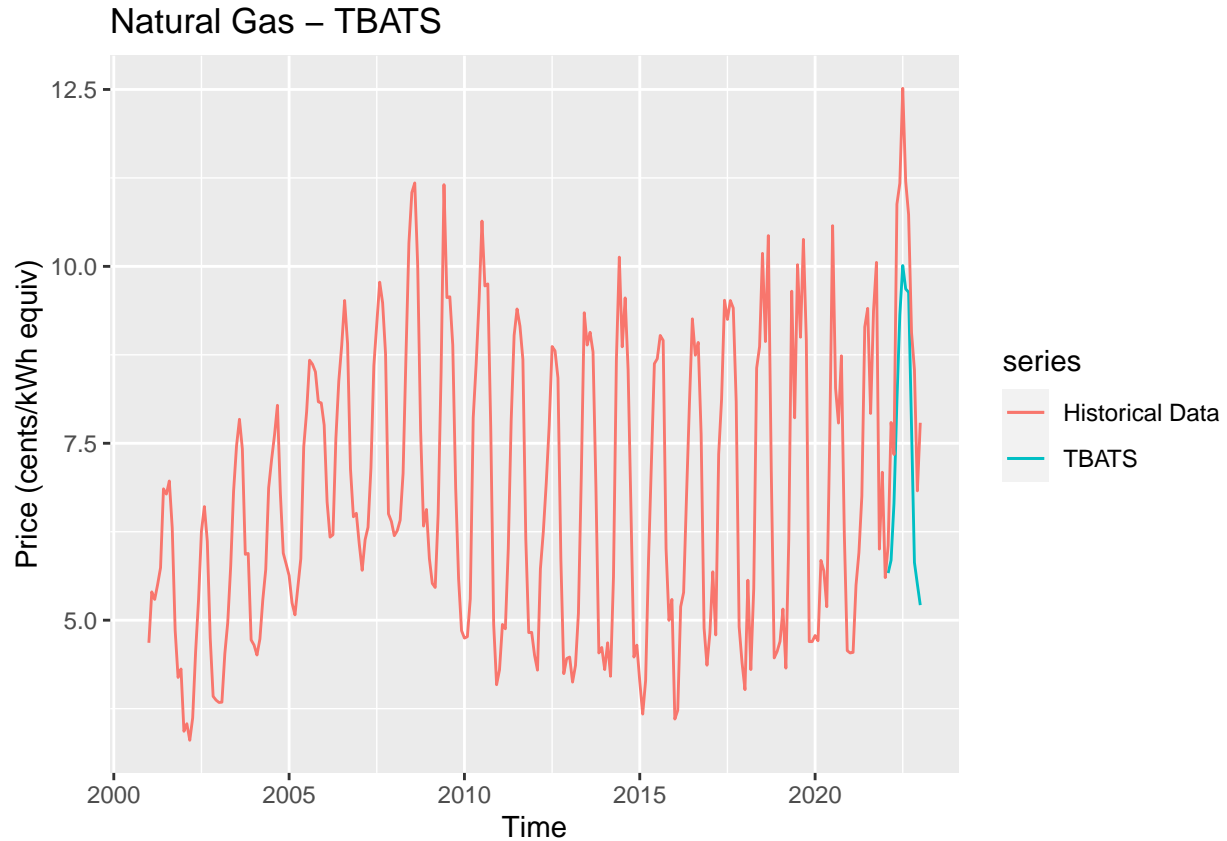




Electricity – TBATS



Then we tried TBATS models for electricity and natural gas series

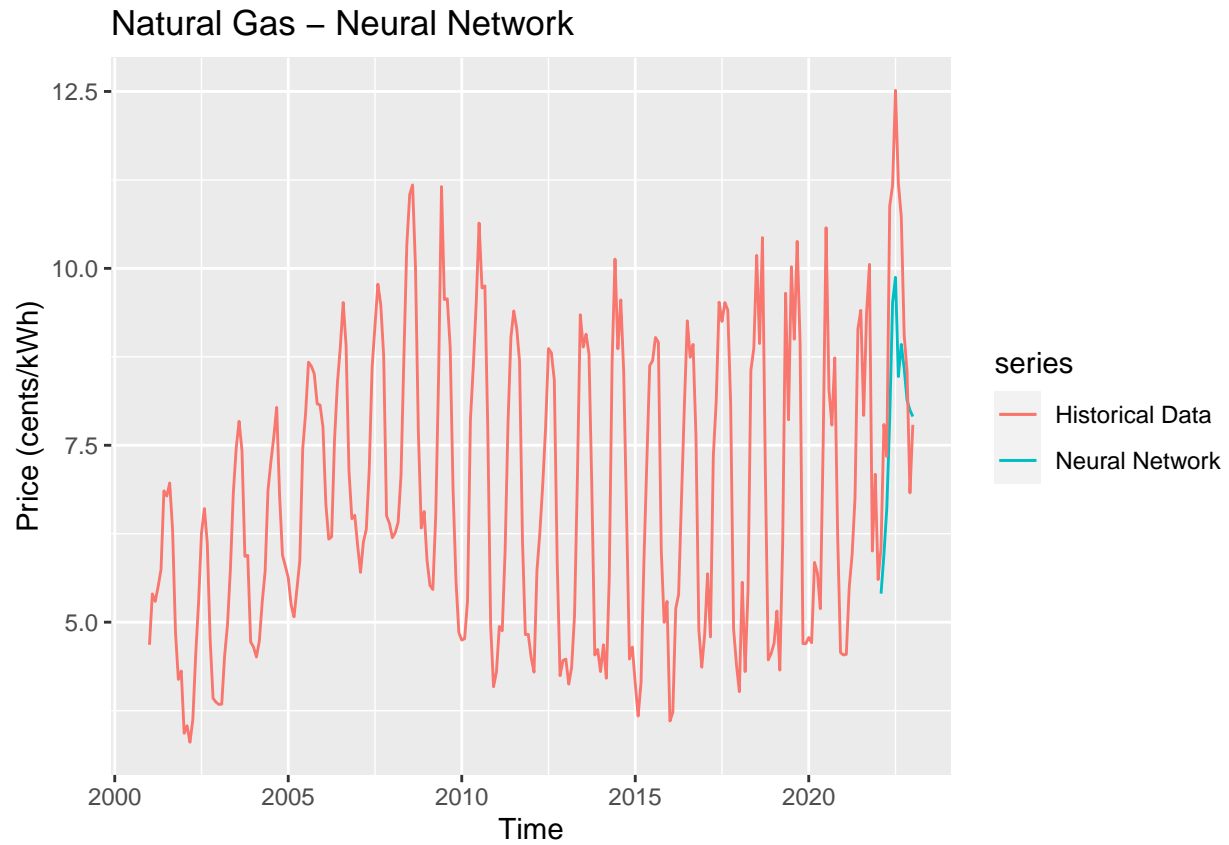


Then we think Ukraine War should have a significant impact on natural gas price and probably electricity gas too. Also, temperature should be a good regressor to include since utility bills normally fluctuate in the same direction with temperature. Therefore, we created two covariates: *UKRWAR* and temperature. *UKRWAR* is an indicator 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 2022 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.

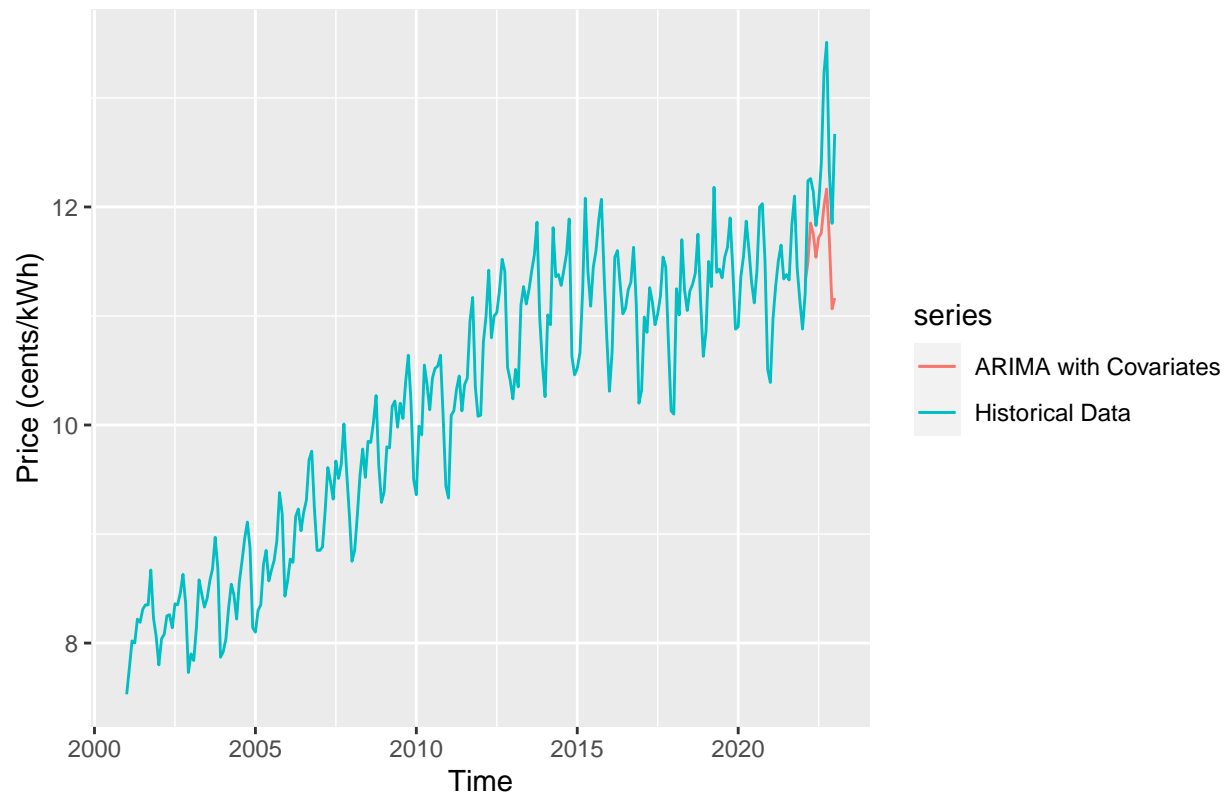
Electricity – Neural Network



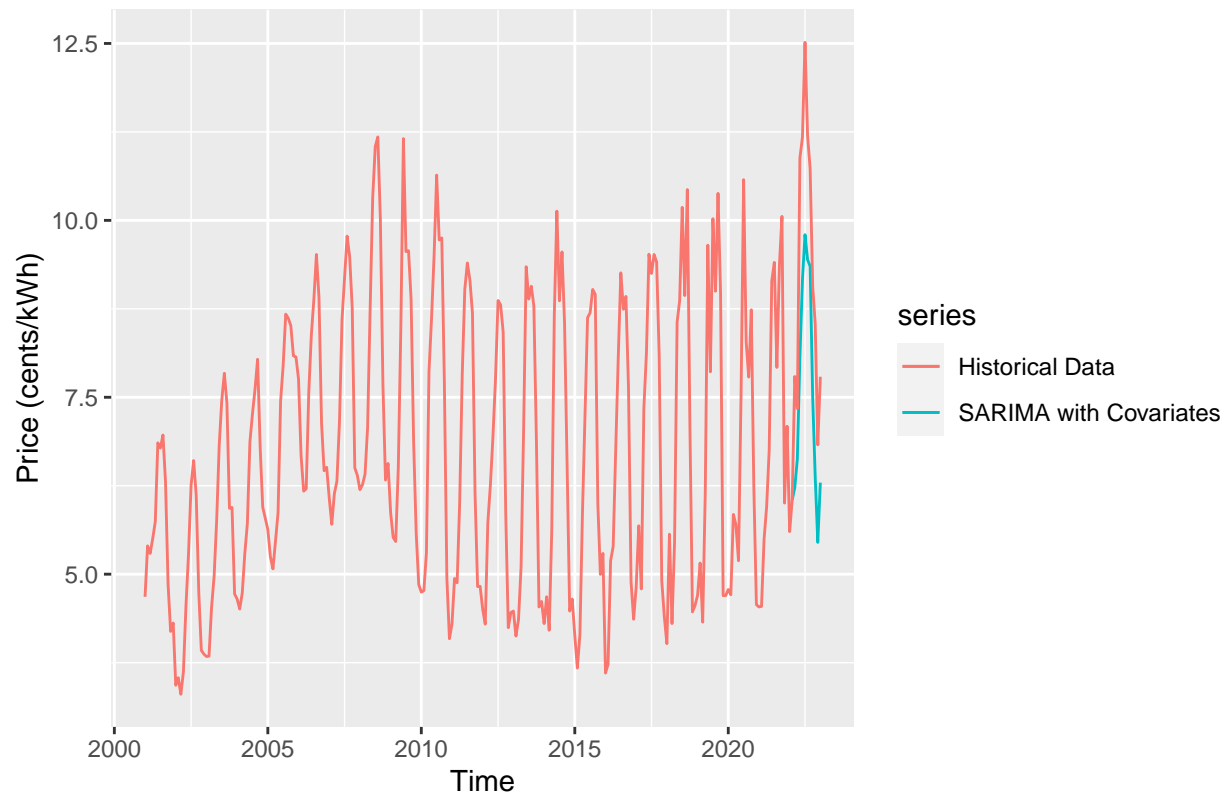


Then we use seasonal arima model with temperature and fourier terms to model two series. UKRWAR is excluded because R reports no suitable ARIMA model when UKRWAR is included. Function used: `auto.arima(xreg)`

Electricity – ARIMA with Covariates



Natural Gas – SARIMA with Covariates

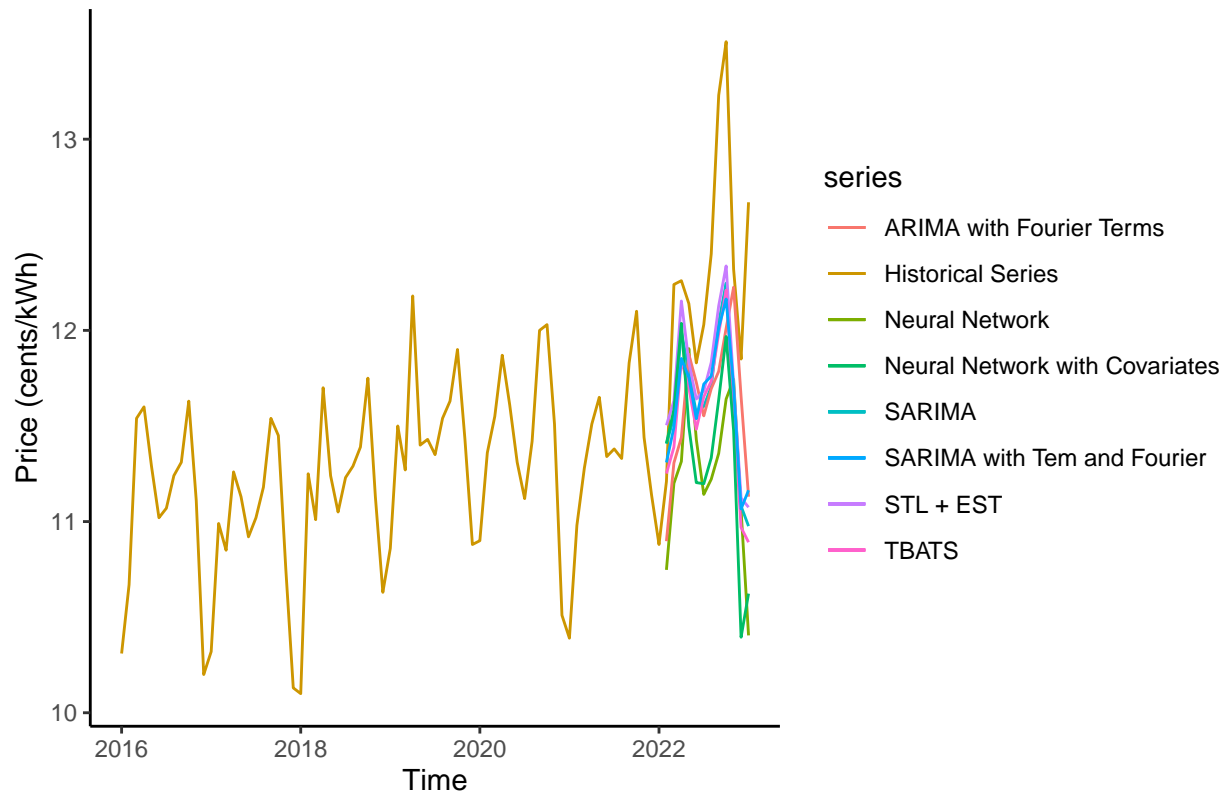


##Summary and Conclusions

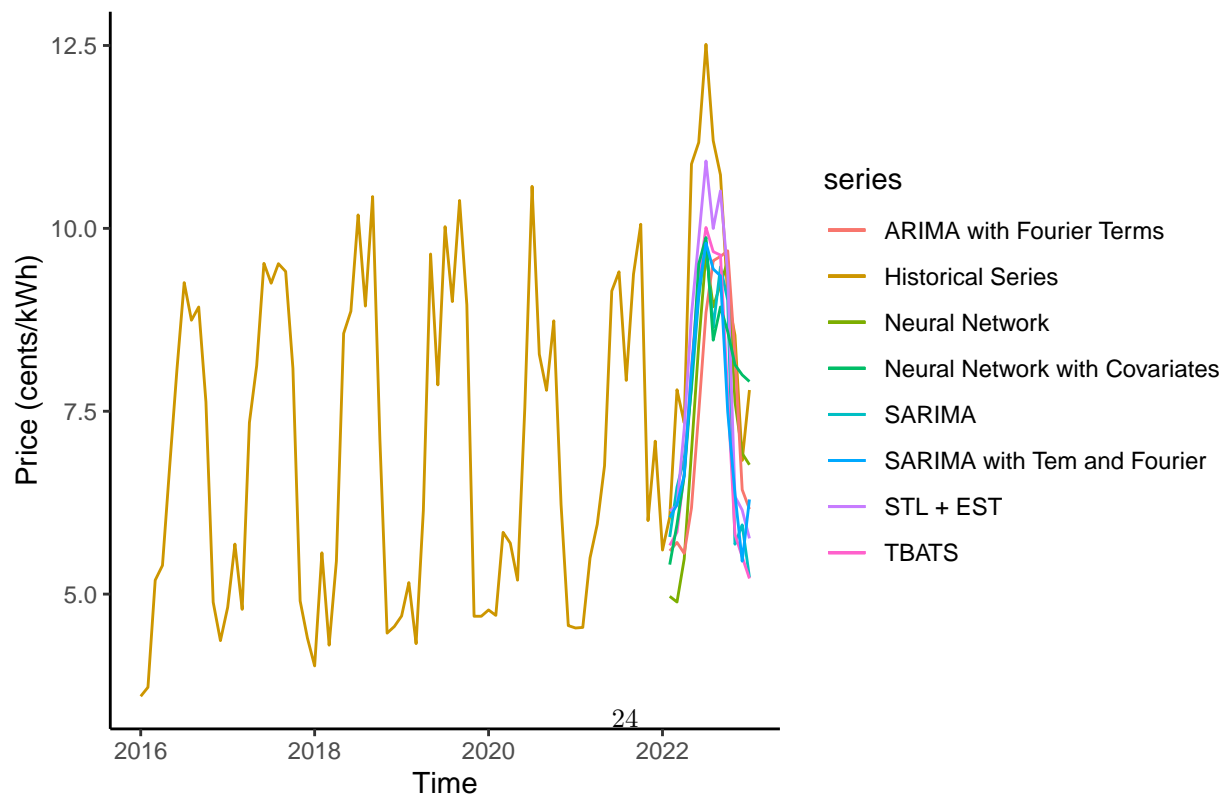
Use ETS to model – not finished – I don't know how to do this, maybe we don't need to include this

compare performance scores and generate tables for use

Comparison of Four Modeling Methods – Electricity



Comparison of Four Modeling Methods – Natural Gas



The best model for electricity by RMSE is: STL

Table 1: 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	-1.04357	1.21351	1.04357	-9.35483	9.35483	0.19522	3.08365
TBATS	-0.74971	0.89063	0.75658	-6.50522	6.56626	0.23462	2.86546
Neural Network with Covariates	-0.94383	1.12115	0.97698	-8.41344	8.70401	0.40842	2.65980
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 2: 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	-1.70904	2.11775	1.80678	-24.68441	25.77080	0.46652	2.56911
TBATS	-1.72858	1.89245	1.72858	-24.53539	24.53539	-0.32231	2.08610
Neural Network with Covariates	-1.22105	1.72181	1.43538	-15.26163	17.94442	0.46435	2.14884
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

