Group Project Report

Members: Minxuan Chen, Jie Min, Guangyu Xing, Xuan Yang

Tutor: Refik Soyer

Time Series Forecasting

The George Washington University

May 13th, 2018

1. Introduction

As the growing volumes of energy traded nowadays, power exchanges have taken a key role. As a result, forecasting for power exchanges has become increasingly important for market participants both for short term bidding and for long term budgeting. Electricity is one source of energy and an essential part of our daily life as well. The main goal of our analysis is to predict future values of electricity.

We have 335 observations in our dataset starts from January 1990 to November 2017. There are three variables that we used in this analysis: monthly electricity production, CPI and natural gas imports. The cost of production affects production, and then the prices of electricity. Electricity prices, as one of the fastest growing sub groups of the consumer price index (CPI) recently, is largely increased and is expected to increase continually in the next few years. Also, natural gas is one of the important part of the source of electricity. It accounted for 21% of the source of generation in 2008. We set 36 samples as our hold-out samples.

Because of deregulation reforms, electricity prices forecasting becomes very important to participants in electricity markets. On the other hand, Modeling the price of electricity is a challenging task, considering its features: electricity cannot be stored, production needs to be balanced regularly against demand, and we have to consider its seasonality, volatility and inelasticity.

2. Univariate Time-series Models

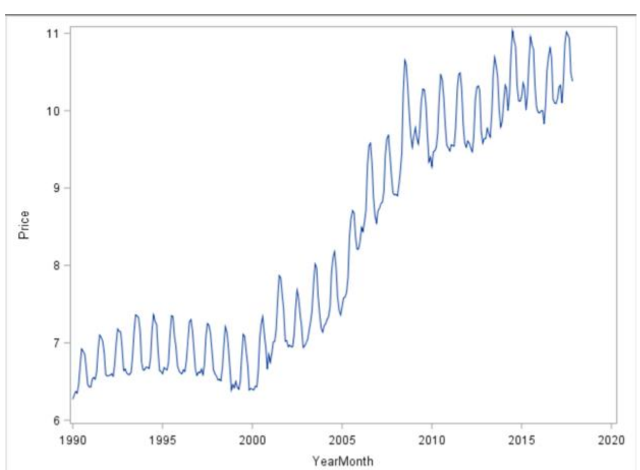

Figure 2.1

According to time series data view function of SAS, we graph the trend of average retail price of electricity by month (Figure 2.1). As is shown in the graph, the whole year trend of retail price at first remain stable before 2000. After 2000, the whole year trend of average retail price shoots up and keeps increasing until now. Therefore, we considered a dummy variable to represent the changes, that is, before or after 2000.

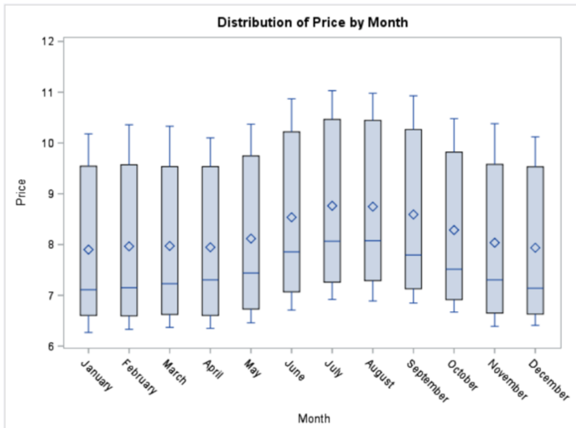

Figure 2.2

According to seasonality analysis, there is a remarkable seasonal trend run through whole year (Figure 2.2). We can see that there is a peak appearing in summer between June and September. Average price gently decreases after September till December and increases slowly from January to May.

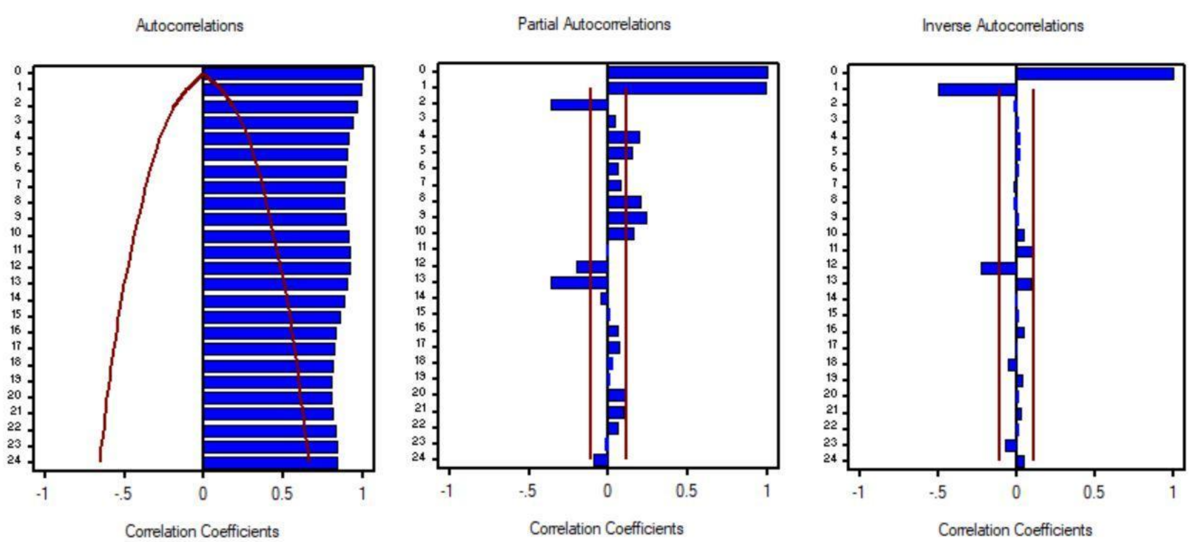

Figure 2.3

From above ACF plot (Figure 2.3), we can find out that it is not stationary.

There are total of 335 observations in our data, and we chose 36 observations as our hold-out samples in all models.

2.1 Deterministic Time Series Models (Seasonal Dummies & Trend) & Error model

2. 1.1 Linear Trend + Seasonal Dummy

By observing time series data, we assume that average retail price of electricity has a prominent seasonality. Thus, we add seasonal dummy based on linear model and get outputs below.

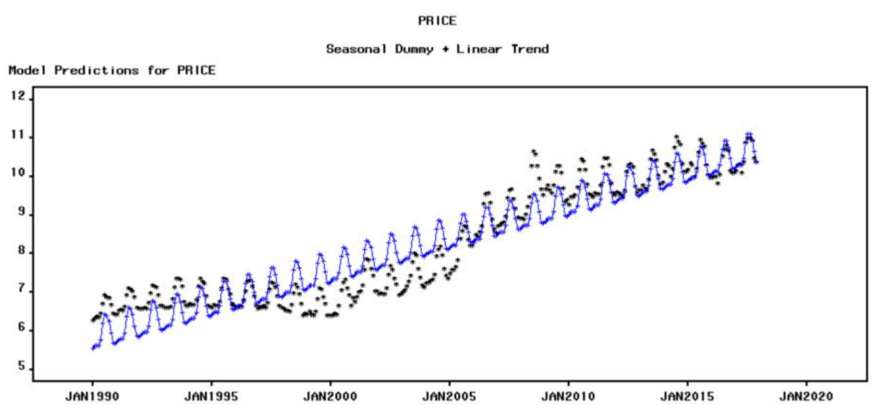

Figure 2.4

We could find that price data got fit better after linear trend added (Figure 2.4). The ACF plot shows that the residuals are not stationary (Figure 2.5) and the coefficients of seasonal dummy are not significantly different than 0 (Figure 2.6).

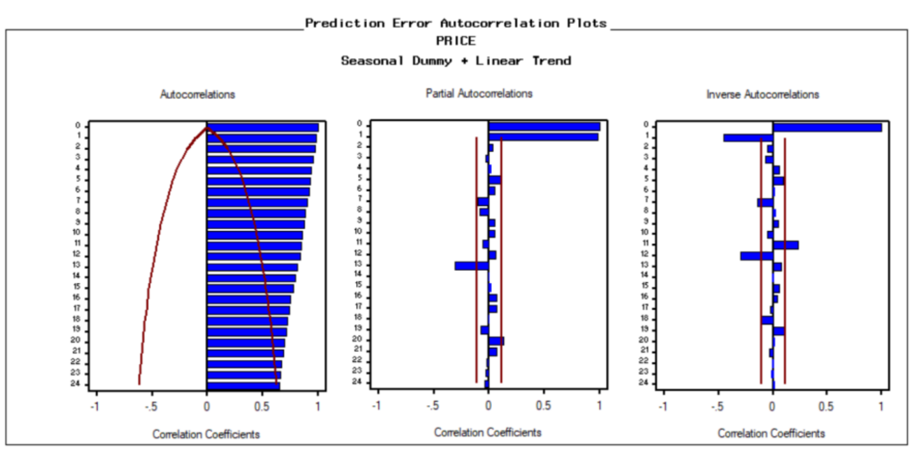

Figure 2.5

Parameter Estimates
PRICE: PRICE
Linear Trend + Seasonal Dummies

Model Parameter	Estimate	Std. Error	T	Prob>|T|
Intercept	5.49799	0.1251	43.9395	<.0001
Linear Trend	0.01449	0.000369	39.3079	<.0001
Seasonal Dummy 1	0.03679	0.1572	0.2341	0.8170
Seasonal Dummy 2	0.08070	0.1571	0.5135	0.6125
Seasonal Dummy 3	0.07661	0.1571	0.4875	0.6305
Seasonal Dummy 4	0.05932	0.1571	0.3775	0.7093
Seasonal Dummy 5	0.20882	0.1571	1.3289	0.1969
Seasonal Dummy 6	0.60913	0.1571	3.8763	0.0008
Seasonal Dummy 7	0.82224	0.1571	5.2325	<.0001
Seasonal Dummy 8	0.79015	0.1571	5.0282	<.0001
Seasonal Dummy 9	0.61246	0.1571	3.8974	0.0007
Seasonal Dummy 10	0.31197	0.1571	1.9852	0.0592
Seasonal Dummy 11	0.03628	0.1572	0.2308	0.8195
Model Variance (sigma squared)	0.30237	.	.	.

Fit Range: JAN1990 to NOV2014

Figure 2.6

2.1.2 Linear Trend + Seasonal Dummy + Error Model

By observing the ACF and PACF of the linear trend and seasonal dummy model (Figure 2.5), we can see that the ACF is obviously not stationary and PACF chops off at lag 2. So, we can use an AR(1) model as the error model to further fit the series. The fitted series are shown below (Figure 2.7).

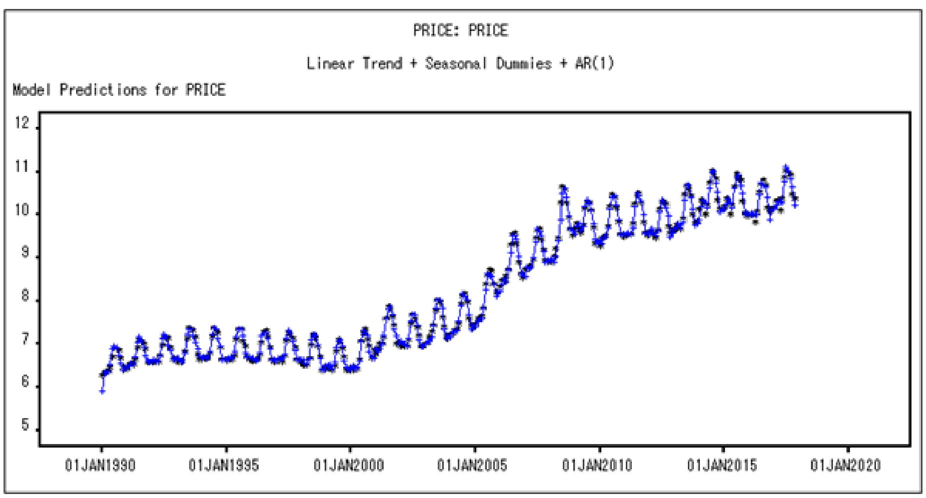

Figure 2.7

The model does not fit the series quite well (Figure 2.8). We see the ACF of this model does not act like a white noise. Because the seasonality still exists, we might need other models to better fitting the series.

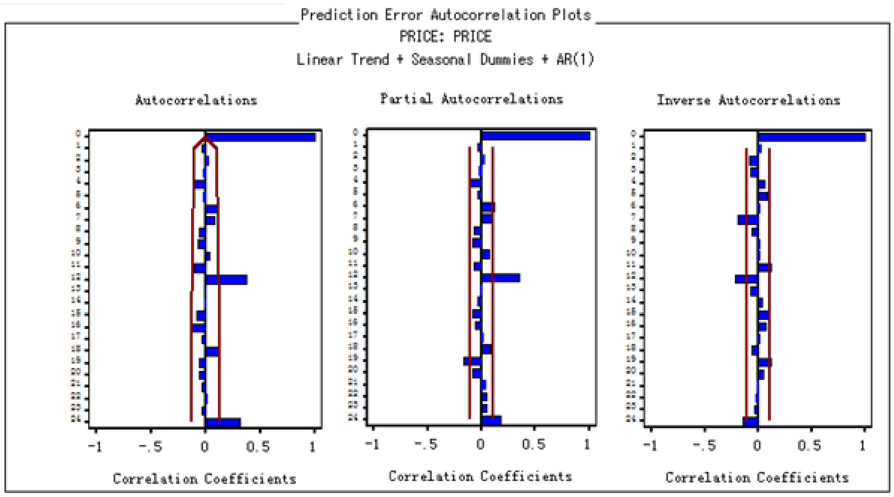

Figure 2.8

By the way, we also added the dummy variable, after of before 2000, but the result is not as good as the result of linear trend + seasonal dummy + error model.

2.2 Exponential Smoothing Models (Only The Relevant Ones)

2.2.1 Simple Exponential Smoothing

Given that the linear trend and seasonal dummy models are still not enough to fit our series, we need to estimate more models. Here is the results of a simple exponential smoothing model. From the fitted plot (Figure 2.9), we can see the series are well fitted.

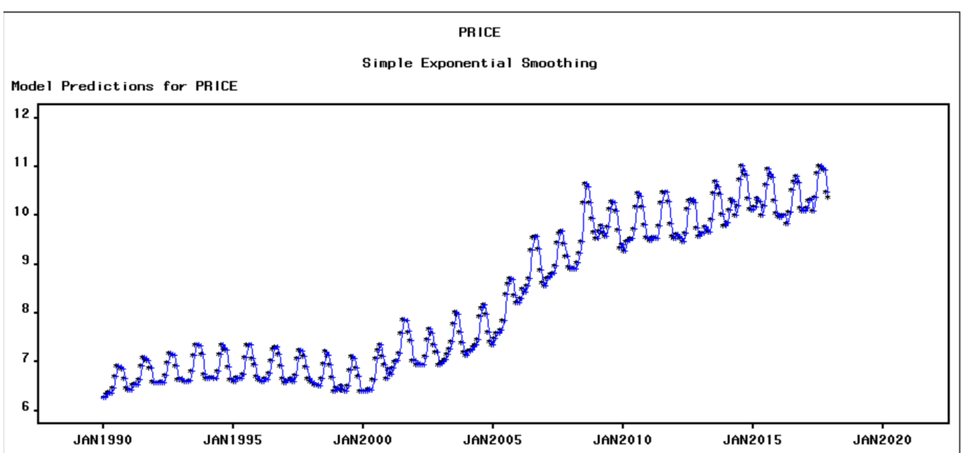

Figure 2.9

But, when we look at the residues and ACF (Figure 2.10), we can see that the seasonality exists in the residuals and we need to take it into consideration.

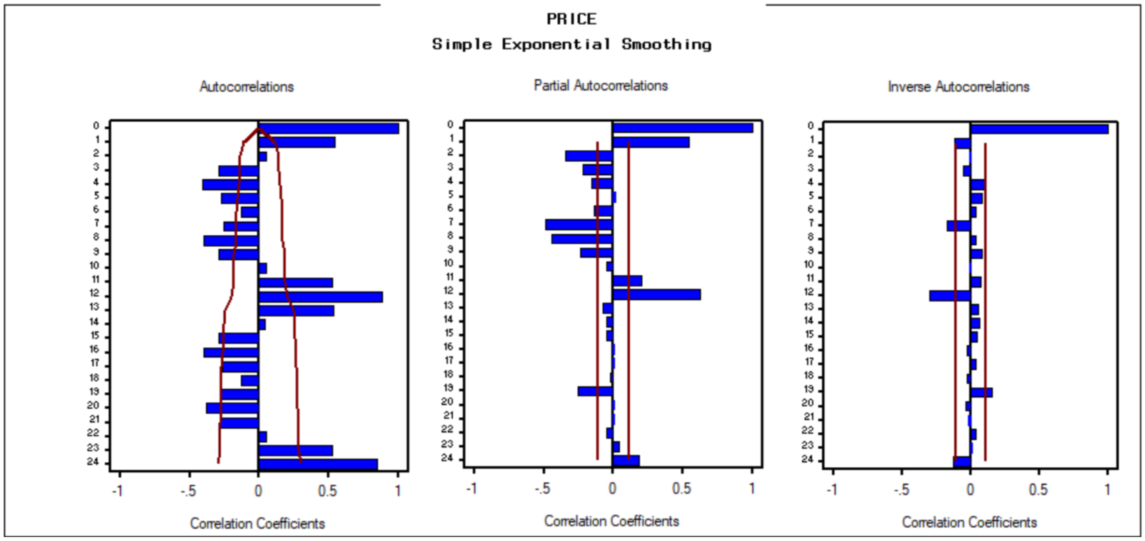

Figure 2.10

2.2.2 Seasonal Smoothing

To fit the series seasonality, we can use seasonal smoothing to model the series. The fitted series are shown below (Figure 2.11).

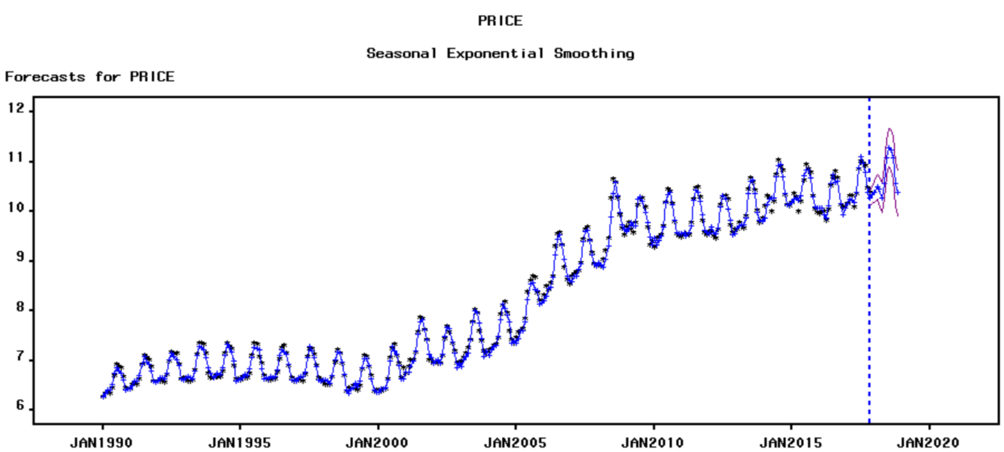

Figure 2.11

We can look at the residues and ACF. It seems seasonal smoothing model performs much better than simple smoothing.

PRICE
Seasonal Exponential Smoothing

Model Parameter	Estimate	Std. Error	T	Prob>|T|
LEVEL Smoothing Weight	0.79880	0.0316	25.3095	<.0001
SEASONAL Smoothing Weight	0.99900	0.1902	5.2525	<.0001
Residual Variance (sigma squared)	0.00682	.	.	.
Smoothed Level	10.39063	.	.	.
Smoothed Seasonal Factor 1	-0.18088	.	.	.
Smoothed Seasonal Factor 2	-0.11495	.	.	.
Smoothed Seasonal Factor 3	-0.18582	.	.	.
Smoothed Seasonal Factor 4	-0.24443	.	.	.
Smoothed Seasonal Factor 5	0.01608	.	.	.
Smoothed Seasonal Factor 6	0.50846	.	.	.
Smoothed Seasonal Factor 7	0.69763	.	.	.
Smoothed Seasonal Factor 8	0.60169	.	.	.
Smoothed Seasonal Factor 9	0.42381	.	.	.
Smoothed Seasonal Factor 10	-0.05263	.	.	.
Smoothed Seasonal Factor 11	-0.26063	.	.	.
Smoothed Seasonal Factor 12	-0.24185	.	.	.

Figure 2.12

The Seasonal Component is almost 1, which means there are almost no smoothing of seasonal factor. We can identify the seasonal factors in our model (Figure 2.12). We can conclude that the series is relatively higher through May to September.

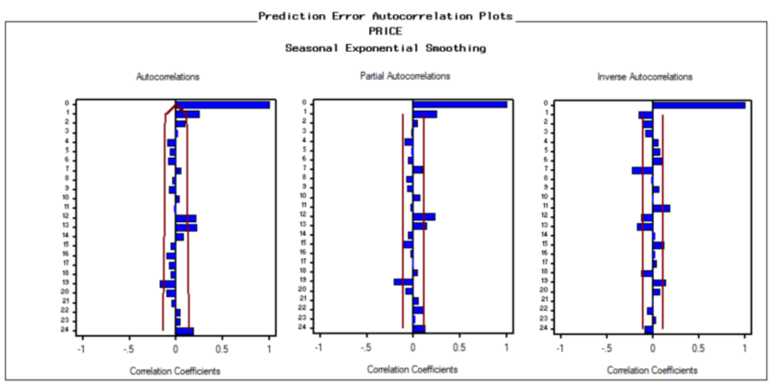

Figure 2.13

Seasonal smoothing model is a better model but it’s still not enough. The model didn’t capture all the seasonality behaviors in series. And, the series is not transformed to a white noise through the model (Figure 2.13).

2.3 ARIMA Models (With Seasonal ARIMA Components If Relevant)

2.3.1 Series Differencing

2.3.1.1 Original Series

As we discussed above, the ACF and PACF of original series are not stationary and have seasonal pattern (Figure 2.3). We will try to do differencing and seasonal differencing to the series.

2.3.1.2 First Difference

First, we do first difference to the series. We can tell an obvious seasonal trend for every 12 month from the plots below (Figure 2.14).

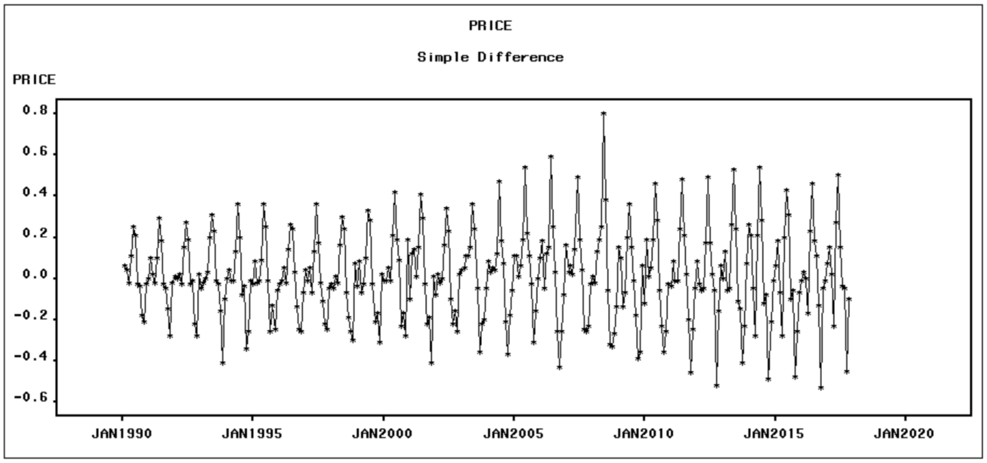

Figure 2.14

The obvious seasonal behavior can be seen in ACF plot (Figure 2.15). The seasonal difference is still needed.

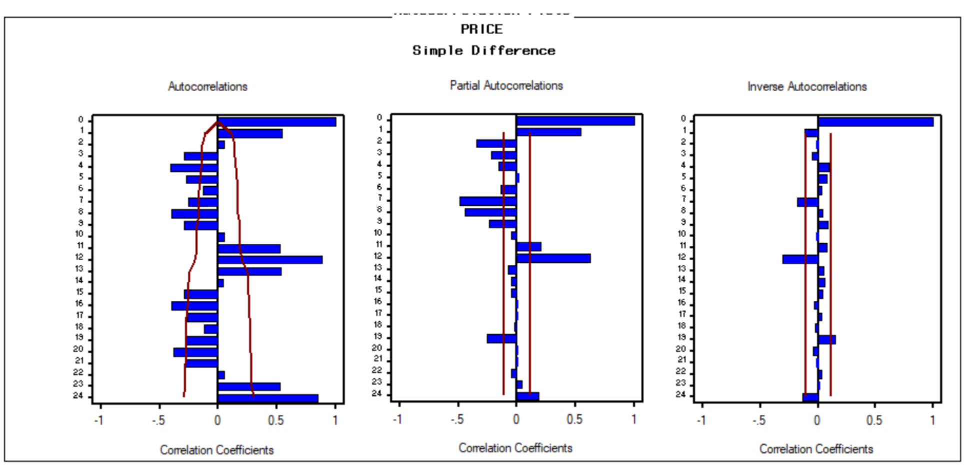

Figure 2.15

2.3.1.3 First Simple Difference And Seasonal Difference

After first and seasonal difference, we can find that the series became stationary (Figure 2.16). The ACF and PACF is chopped to 0 after lag 0. For the seasonal factor, ACF cut off to 0 after first 12 lags. It suggests a pattern of ARIMA(0,1,0)(0,1,1)$_{12}$ model (Figure 2.17).

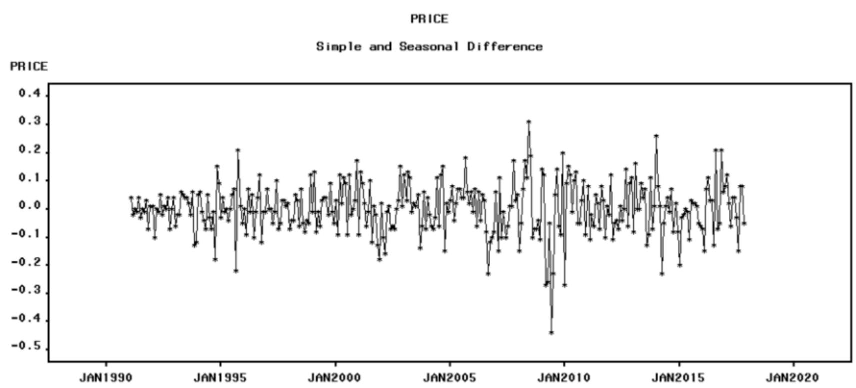

Figure 2.16

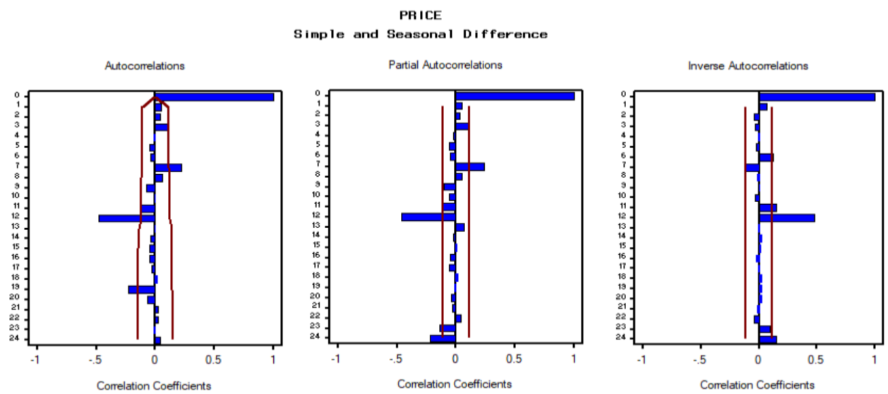

Figure 2.17

2.3.1.4 ARIMA Model

We fitted the series with ARIMA(0,1,0)(0,1,1)$_{12}$ model. The model performs well in fitting the series (Figure 2.18). And the residue series is almost a white noise (Figure 2.19). The fitted series and ACF plots are shown below.

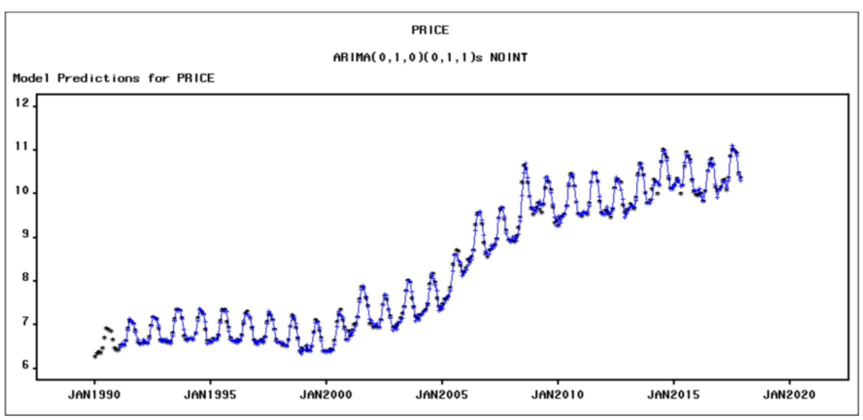

Figure 2.18

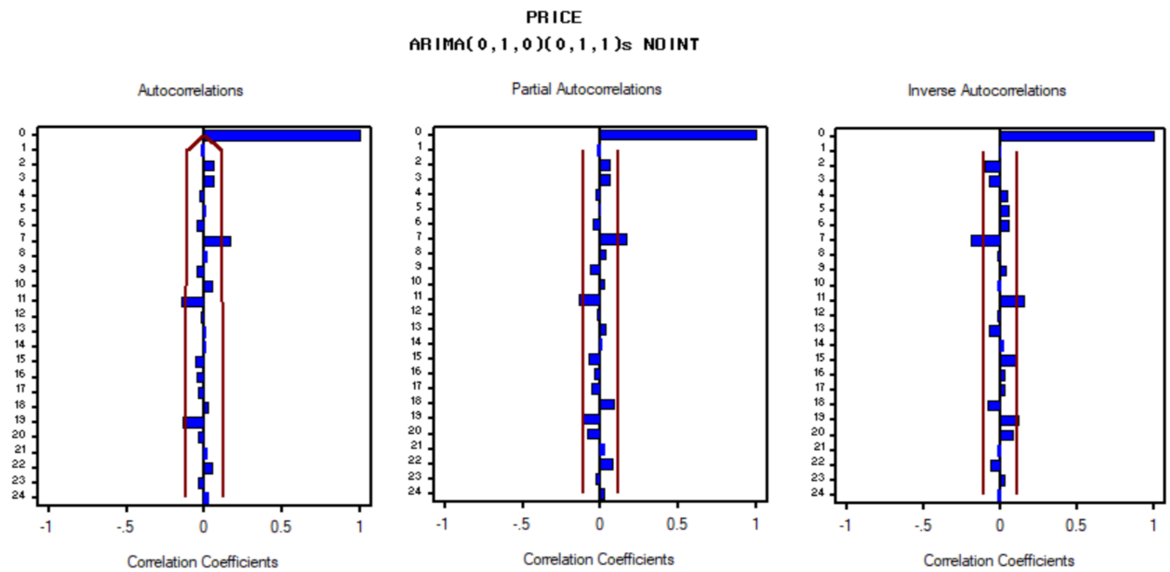

Figure 2.19

We can ignore the ACFs that are slightly higher than the bar and conclude that the series is white noise.

The coefficient of MA parameter is significantly different than 0 (Figure 2.20). The model has an MAPE of only 0.58. The model performs really well and achieved low values in both MAPE and RMSE (Figure 2.21).

PRICE
ARIMA(0,1,0)(0,1,1)s NOINT

Model Parameter	Estimate	Std. Error	T	Prob>|T|
Seasonal Moving Average, Lag 12	0.65594	0.0513	12.7894	<.0001
Model Variance (sigma squared)	0.00583	.	.	.

Figure 2.20

Statistic of Fit	Value
Mean Square Error	0.0056614
Root Mean Square Error	0.07524
Mean Absolute Percent Error	0.58005
Mean Absolute Error	0.06015
R-Square	0.952

Figure 2.21

Here is the prediction made by ARIMA model (Figure 2.22).

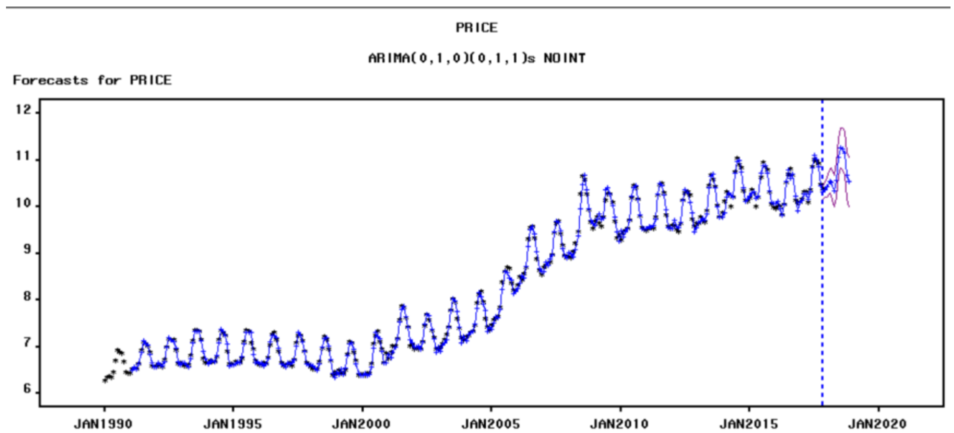

Figure 2.22

2.4 Comparison of Models (In Terms of Fit And Validation)

After analyzing the smoothing models and ARIMA models, we can find that the Seasonal Exponential Smoothing model and ARIMA(0,1,0)(0,1,1)$_{12}$ are the potential best model to fit and forecast the series.

From the forecast graph of the both models (Figure 2.11 and 2.22), we can conclude that they did a good job in forecasting for the hold-out samples.

Forecast Model	Model Title	Root Mean Square Error
☐	Seasonal Exponential Smoothing	0.08232
☑	ARIMA(0,1,0)(0,1,1)s NOINT	0.07641

Figure 2.23

Forecast Model	Model Title	Mean Absolute Percent Error
☐	Seasonal Exponential Smoothing	0.77970
☑	ARIMA(0,1,0)(0,1,1)s NOINT	0.70110

Figure 2.24

(Fit-model-based variance)

Forecast Model	Model Title	Root Mean Square Error
☐	Seasonal Exponential Smoothing	0.08965
☑	ARIMA(0,1,0)(0,1,1)s NOINT	0.07524

Figure 2.25

Forecast Model	Model Title	Mean Absolute Percent Error
☐	Seasonal Exponential Smoothing	0.69940
☑	ARIMA(0,1,0)(0,1,1)s NOINT	0.58005

Figure 2.26

(Evaluation-based variance)

When Root Mean Square Error and Mean Absolute Percent Error are used to evaluate the two models, the ARIMA(0,1,0)(0,1,1)$_{12}$ model is the better model for it has lower errors (Figure 2.23 - Figure 2.26).

As the results showed above, ARIMA(0,1,0)(0,1,1)$_{12}$ is the best model for univariate time series forecasting for US electricity price.

3. Multivariate Time-series Models

Based on the analysis above, some univariate models were good in fitting Price series but those models didn't fully capture the behavior of the series. As a result, we may need a better model with more regressors added to fit our series. We chose using CPI and Natural Gas Import as regressors to further conduct our analysis.

In this part, we exclude the 36 hold-out sample to fit the model and hold-out sample are used for assessing the forecast.

3.1 Price VS. CPI

As shown in the ARIMA part, we can do a simple differencing and a seasonal differencing to make the Price series stationary. So, we can use differenced price series as our response series in TF model.

3.1.1 Check for Stationarity

In a TF model, we need to prewhiten our input series. So, we start with stationary check for CPI series. We can see from the plots below that the CPI is obviously not stationary and has a clear up going trend (Figure 2.27). So, we need to do differencing to make the series stationary.

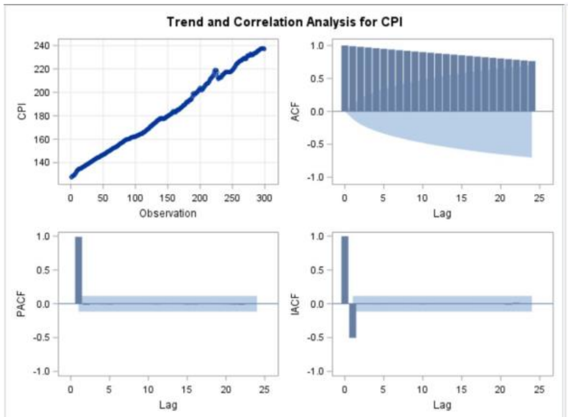

Figure 2.27

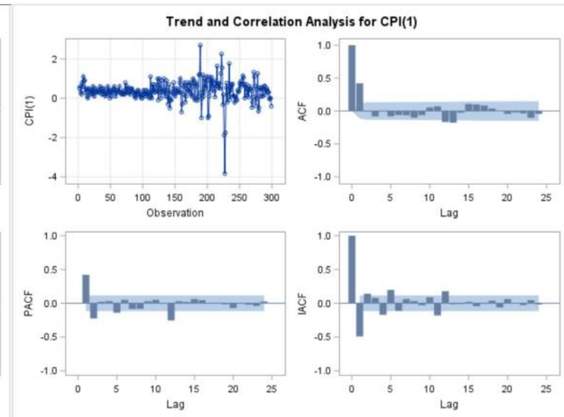

Figure 2.28

After first differencing (Figure 2.28), we can get a series shown below. It has some extreme values which our method might not be able to capture. But the series looks stationary now.

3.1.2 Pre-Whitening Of CPI

By observing the decreasing and chop-off pattern of ACF and PACF, we can identify a moving average pattern in the differencing series. It is further confirmed with IACF. And, we can not ignore the seasonal patterns in the ACF at lag of 12 which is a behavior of seasonal moving average. So, we decide to pre-whiten the CPI series with the model: ARIMA(0, 1, 1)(0, 0, 1)$_{12}$. The results is shown below (Figure 2.29). We can say that the input CPI series is white noise after pre-whitening process.

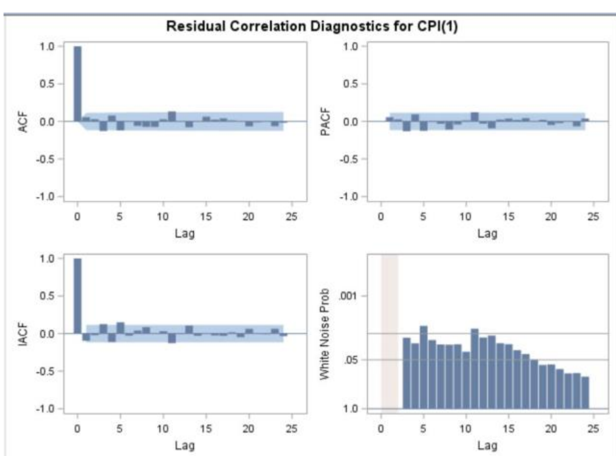

Figure 2.29

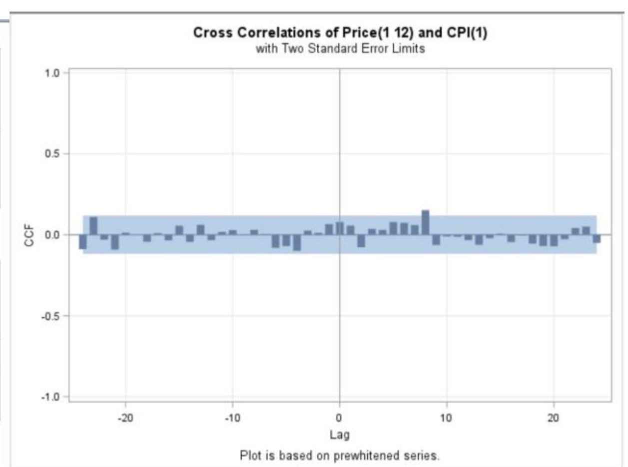

Figure 2.30

3.1.3 Identify TF By Sample CCF

With the prepared series, we can use CCF to identify TF models. From plot above (Figure 2.30), we can ignore the ACF at negative lags, and conclude that the first significant response in CCF is at lag 8, which means b = 8.

Given that the CCF cuts off at lag 9, we can conclude that r = 0. And, we can see the impulse response is significant only at lag 8 without other terms. It is a behavior of s = 0.

3.1.4 Estimate the TF model

With the above information, we can fit a TF model to our series with b=8 and r=s=0. And based the ACF of the error, a seasonal AR(1) should be applied in the error model (Figure 2.31).

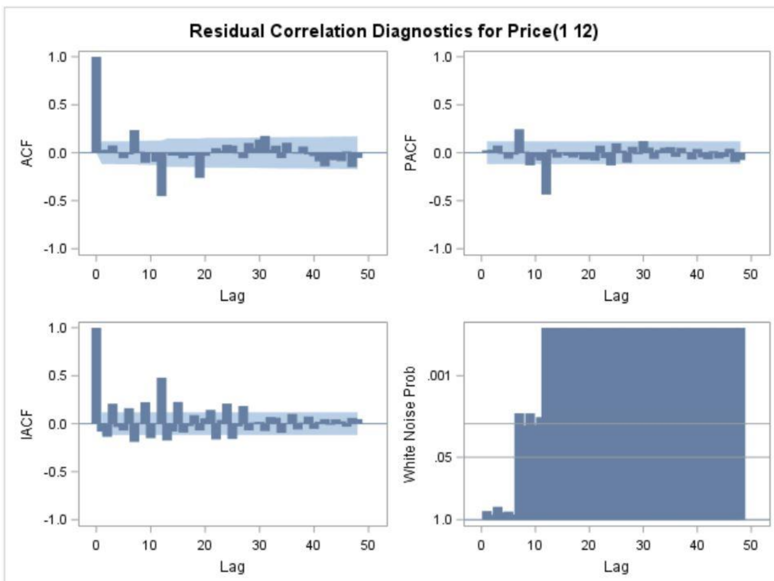

Figure 2.31

After adding the error model, the result is shown below (Figure 2.32). The coefficients are significant.

Autoregressive Factors	
Factor 1:	1 + 0.48252 B**(12)

Unconditional Least Squares Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag	Variable	Shift
AR1,1	-0.48252	0	-Infty	<.0001	12	Price	0
NUM1	0.03114	0	Infty	<.0001	0	CPI	8

Input Number 1	
Input Variable	CPI
Shift	8
Period(s) of Differencing	1
Overall Regression Factor	0.031141

Figure 2.32

Given the parameters above, we can write the formulas for our model:

$$(1-B)(1-B^{12})PRICE_t = 0.03114(1-B)CPI_{t-8} + \varepsilon_t$$

$$(1+0.48252B^{12})\varepsilon_t = a_t$$

(a_t is white noise)

3.1.5 Check of Model Adequacy

The ACF for model residues are shown below. ACF are slightly significant at only few lags. We can ignore those values and conclude that the residue is almost white noise.

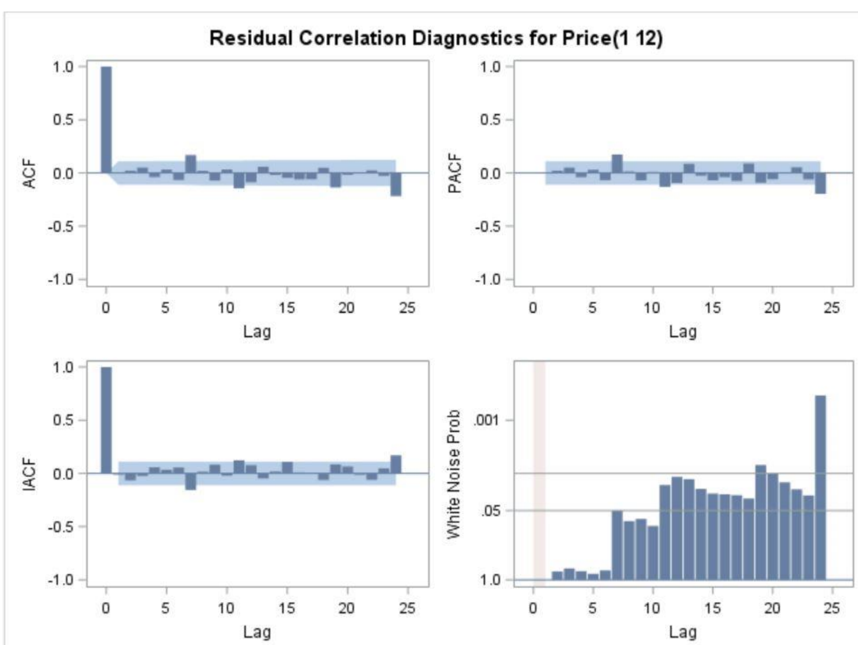

Figure 2.33

We can also see the model results directly from the table below (Figure 2.34).

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	5.64	5	0.3434	0.023	0.068	0.086	0.003	0.079	-0.023
12	24.00	11	0.0127	0.205	0.029	-0.055	0.085	-0.085	-0.040
18	28.21	17	0.0425	0.095	0.012	0.019	0.016	-0.002	0.064
24	40.60	23	0.0131	-0.121	0.007	0.023	0.082	0.074	-0.111
30	62.56	29	0.0003	0.143	0.046	-0.067	0.135	0.046	0.146
36	75.82	35	<.0001	0.110	0.036	0.096	-0.006	0.130	0.032
42	81.59	41	0.0002	0.019	0.094	0.033	0.005	0.014	-0.082
48	85.60	47	0.0005	-0.001	-0.061	-0.064	0.038	-0.047	0.010

Crosscorrelation Check of Residuals with Input CPI									
To Lag	Chi-Square	DF	Pr > ChiSq	Crosscorrelations					
5	4.28	5	0.5092	-0.009	-0.103	-0.048	0.022	-0.009	-0.043
11	8.88	11	0.6328	-0.070	0.014	-0.018	0.023	-0.017	-0.102
17	17.51	17	0.4203	-0.106	-0.071	0.033	0.054	-0.073	-0.073
23	21.77	23	0.5344	0.019	0.044	-0.110	-0.007	-0.027	-0.000
29	27.89	29	0.5237	-0.093	0.042	0.063	-0.065	0.056	-0.017
35	34.23	35	0.5053	-0.042	-0.063	0.104	0.077	0.011	0.014
41	47.57	41	0.2227	0.046	-0.094	-0.142	-0.029	0.049	-0.117
47	50.86	47	0.3243	-0.061	0.011	-0.040	0.080	-0.001	0.004

Figure 2.34

3.1.6 Forecast

The forecast from our model is shown below. It displays an upward trending for price (Figure 2.35). However, for the hold-out sample, the actual values are almost in around the lower bound of predictions.

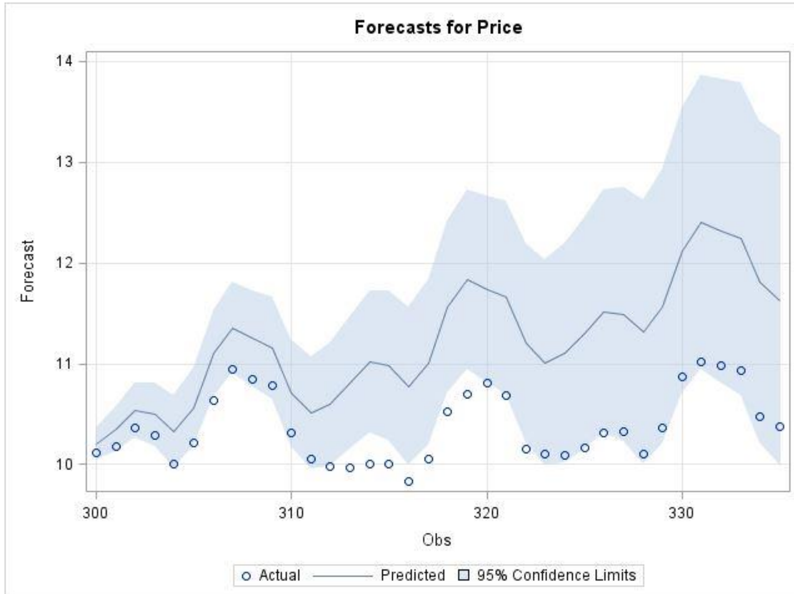

Figure 2.35

3.2 Price VS. GASIMPORTS

3.2.1 Check For Stationarity

Similarly, we get started with checking stationary of GASIMPORTS. We can see from the plots below (Figure 2.36), that the GASIMPORTS is obviously not stationary and has a clear up going trend at the beginning and has a decreasing trend recently. So, we need to do differencing to make the series stationary.

After simple differencing (Figure 2.37), the GASIMPORTS series became stationary.

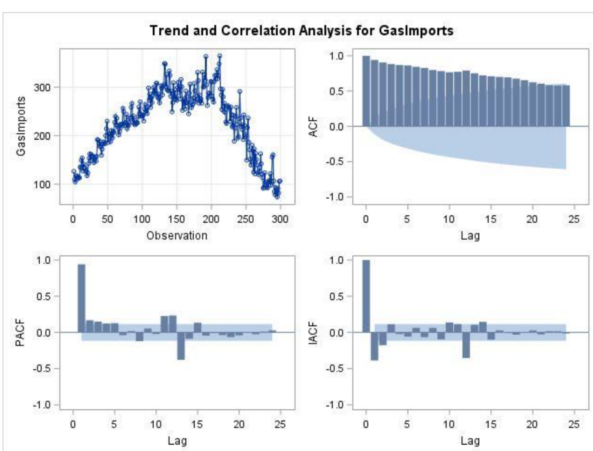

Figure 2.36

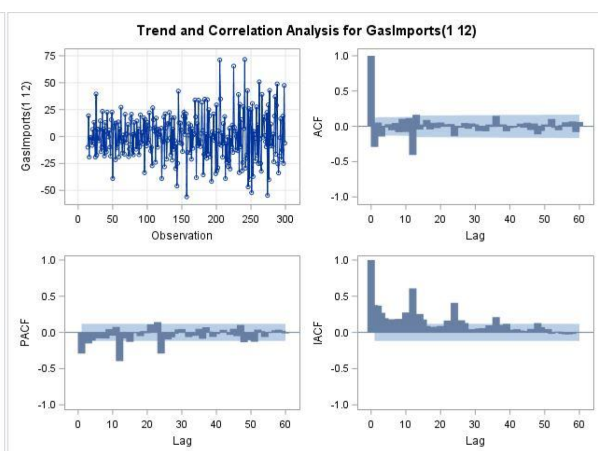

Figure 2.37

3.2.2 Pre-Whitening Of The GASIMPORTS

By observing the decreasing and chop-off pattern of ACF and PACF, we can identify a moving average pattern. And, we can also identify the seasonal patterns in the ACF at lag of 12 which is a behavior of seasonal moving average. So, we decide to pre-whiten the GASIMPORTS series with the model: ARIMA(0, 1, 1)(0, 1, 1)$_{12}$. The results is shown below (Figure 2.38). We can say that the input GASIMPORTS series is white noise after pre-whiten process.

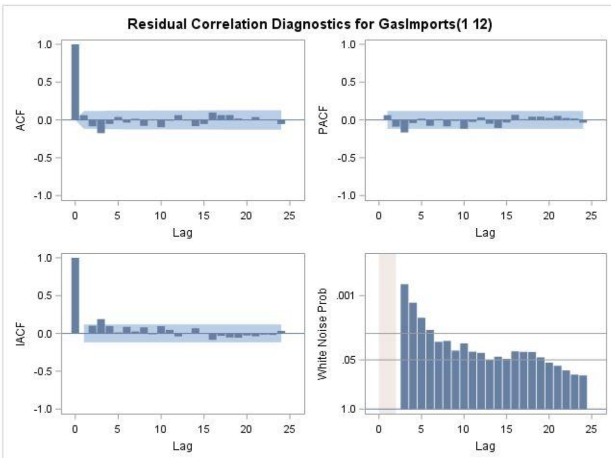

Figure 2.38

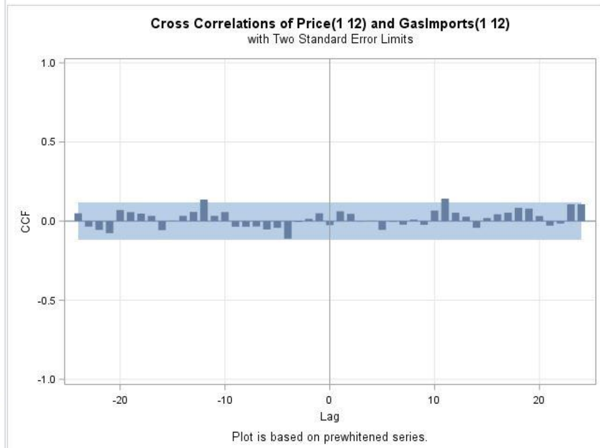

Figure 2.39

3.2.3 Identify TF By Sample CCF

By observing the CCF (Figure 2.39), we can say that there is almost no significant response in CCF. There is no significant response within lag of -20 to lag of 20. So, it's not appropriate to use TF model to estimate a model between Price and GasImports.

4. Conclusions

According to above univariate analysis, we can detect that our target variable monthly average retail electricity price has a upward trend and obvious seasonality by fitting deterministic model and exponential smoothing model. Especially in exponential smoothing model, the seasonal smoothing components are almost equal to 1 which indicate the strong seasonal fluctuation. In this way, we decided to make difference and seasonal difference before fitting ARIMA model and found ARIMA(0,1,0)x(0,1,1)$_{12}$ is the best fitted model which has the lowest values of RMSE and mean absolute percent error.

When we look at multivariate analysis, our target is to investigate binary correlation between the predictor and electricity price. We select CPI and GASIMPORTS as predictors and research if each of them has correlation with price respectively. As a result, after pre-whitening of predictors and identified transfer functions by sample cross-correlation function, we can not conclude that transfer function is appropriate to estimate a model to depict correlation for predictors and electricity price.

Among all the models above, we believe the Seasonal ARIMA model is the best to describe and predict the electricity price in the USA.