Jason Gerra
jay.gerra@gmail.com
Time Series (Spring 2016)

A Time Series Model of US Electricity Prices

Introduction

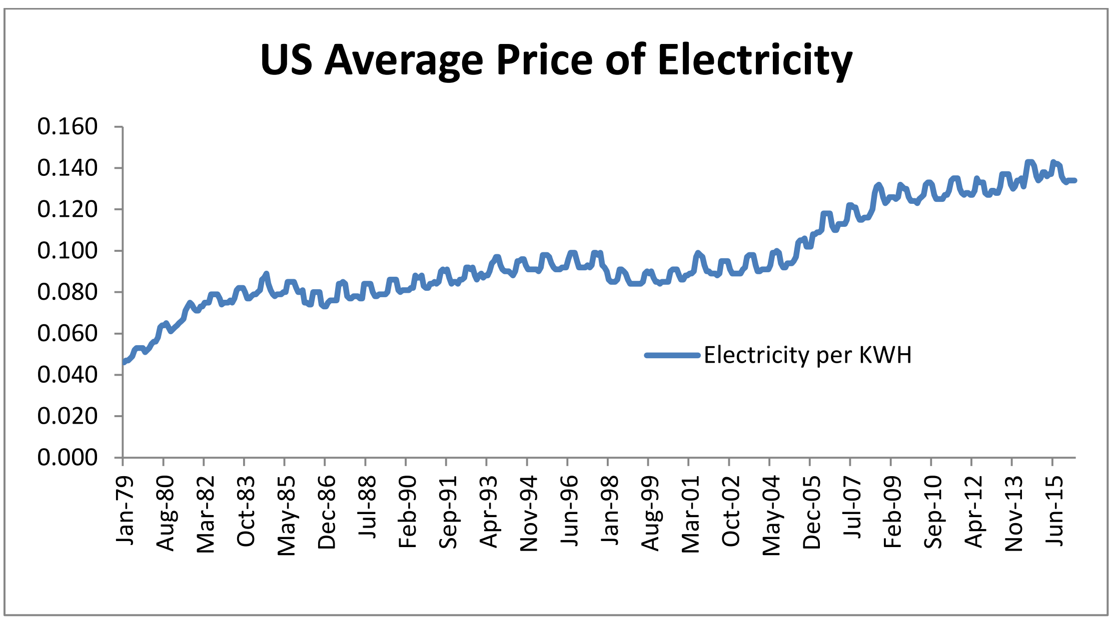
This project creates a time series model for the average price of electricity (per kilowatt Hour) in US cities based on data from January 1979 to April 2016.

Data

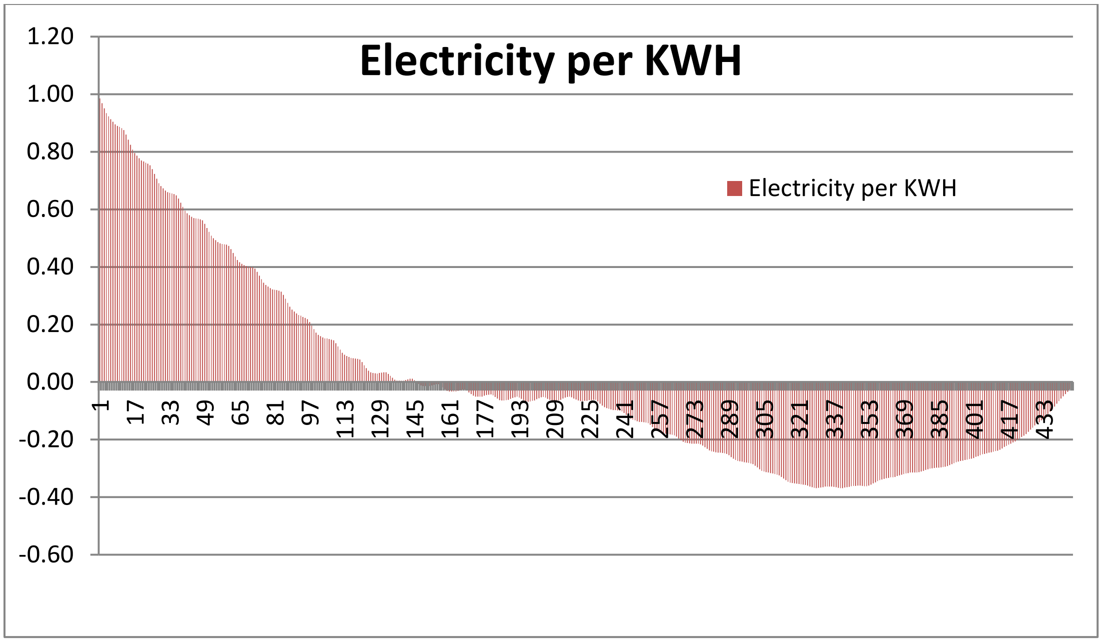
Historical data for the average price of electricity per kilowatt hour across US cities can be found here:

http://data.bls.gov/pdq/SurveyOutputServlet

Monthly data was collected dating back to Jan 1979. The plot of the price per kilowatt hour over time is below:

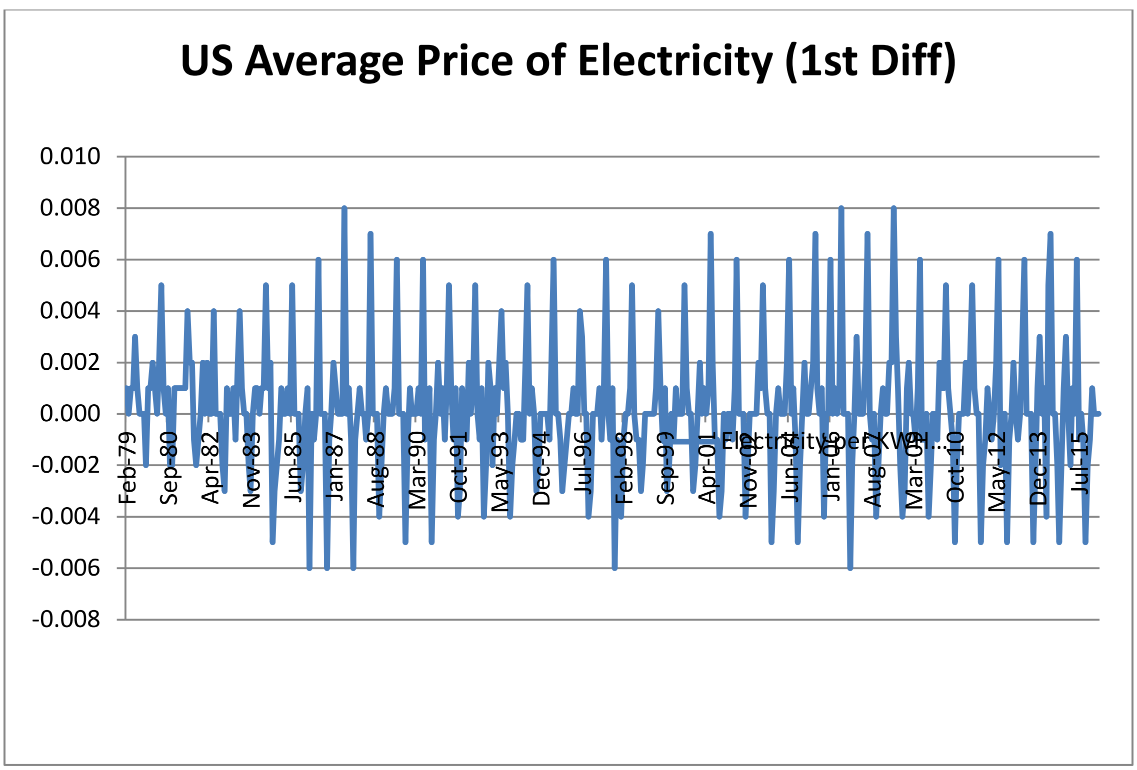


Given the increasing trend of the plot, this would seem to suggest that this time series is not stationary. We can look at the correlogram of this series to confirm.

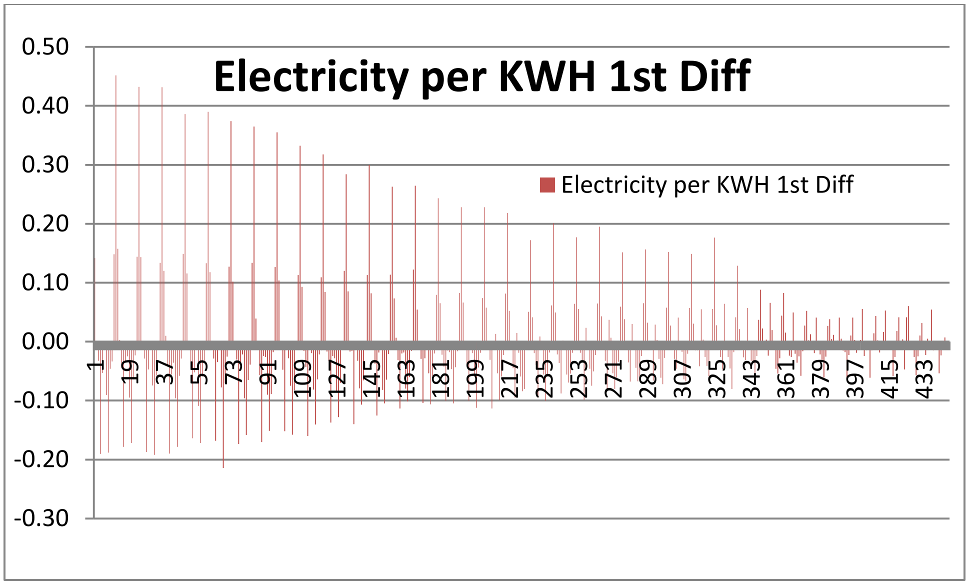


The correlations are zero around the 113th lag and then become negative. This confirms that this series is not stationary. We will take the 1st difference model to try to attain a stationary model.

Below is a plot of the first difference.

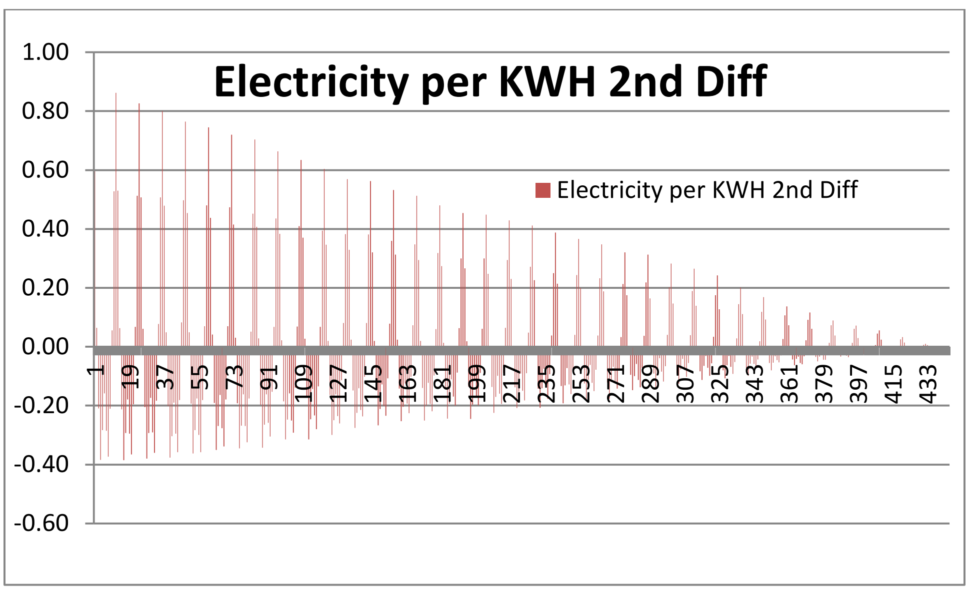


We can see that the plots seem to oscillate around 0 implying stationarity. We will review the correlogram of this series next.



The correlogram above oscillates around a mean and seems to decrease to zero as we move down the lags. Therefore this series is stationary and

Just out of curiosity, we can take a look at the 2nd difference correlogram to see if there is anything to gain from this.



The 2nd difference time series decays a bit quicker than the 1st difference but otherwise seems very similar to the 1st difference in that it oscillates around a mean value and then decays to 0. There doesn't seem to be any additional value taking the 2nd difference of this time series, however we will review 1st and 2nd differences to determine the best possible model.

Model Specification

We will review the following types of models:

ARI(1,1)
ARI(1,2)
ARI(1,3)
ARI(2,1)
ARI(2,2)
ARI(2,3)

The general equations for AR(1,i) and AR(2,i) models are:

$$AR(1):\ Y_t = \varphi Y_{t-1} + \varepsilon_t$$

$$AR(2):\ Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \varepsilon_t$$

We have calculated the coefficients using Excel's Regression tool. The results of the regression for all six models are below.

Model	Adjusted R^2	Intercept	ϕ_1	ϕ_2	ϕ_3	$\phi_2 + \phi_1 < 1$	$\phi_2 - \phi_1 < 1$	$\phi_3 + \phi_2 + \phi_1 < 1$
AR(1,1)	0.05893	0.00015	0.24705					
AR(1,2)	0.05994	0.00016	0.25802	(0.04192)		0.21610	(0.29995)	
AR(1,3)	0.06037	0.00016	0.26315	(0.06551)	(0.02149)			0.17615
AR(2,1)	0.35440	0.00016	0.59653					
AR(2,2)	0.48556	0.00023	0.86659	(0.45272)		0.41387	(1.31931)	
AR(2,3)	0.48451	0.00023	0.87135	(0.46175)	0.01045			0.42006

The adjusted R^2 much higher for the 2nd difference models, however this doesn't necessarily imply that these models are the best ones to use. We will use the Durbin Watson test for residual correlation to better analyze the models and determine the most appropriate model to use.

Model Diagnostics

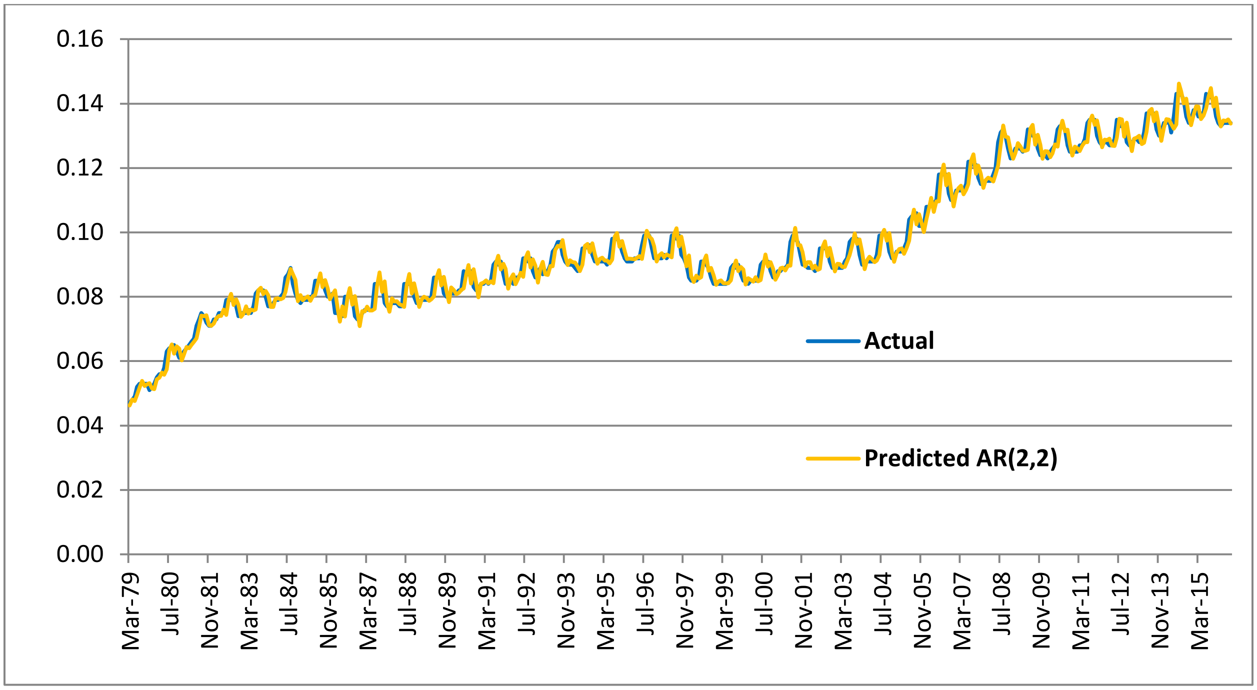
Using the residual output from the Excel Regression too, we can easily calculate the Durbin Watson statistic for each model. We take the ratio of the sum of the squared difference between the t^{th} residual and the *t-1* residual and sum of the residuals squared. A ratio that is close to 2 implies there is no material correlation between residuals. Below are the results.

Model	Durbin Watson
AR(1,1)	1.9643
AR(1,2)	1.9887
AR(1,3)	2.0194
AR(2,1)	1.4596
AR(2,2)	1.9903
AR(2,3)	1.9926

With the exception of the ARI(2,1) model, all of the models are very close to 2. The model closest to 2 is the ARI(2,3) model. However, using the principle of parsimony we will use the ARI(2,2) model as our fitted model.

Final Model and Fit

The DW test results have led us to the ARI(2,2) model (2nd lag and 2nd difference model). Below is a comparison between the actual vs. predicted. The results indicate that this model fits very well.



Forecasting

Using the AR(2,2) model we can forecast the next 3 months. Below are the results. The results show moderate increases in values as opposed to the up and down jumps seen in prior results. This is probably due to the fact that model is based on the prior 2 actual values. In this case they were fairly level so the forecasting is as such too.