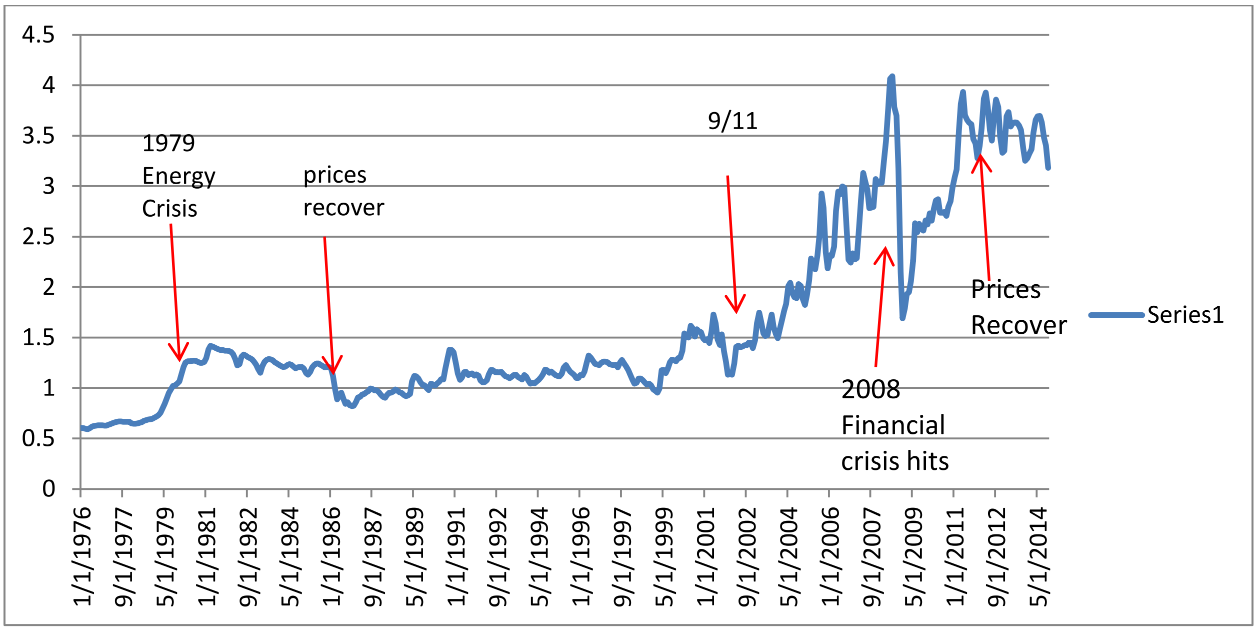
Introduction

The purpose of this project is to come up with a time series model that explains and predicts fluctuations in gasoline prices over time. Gasoline price data was obtained from the US Energy Information Administration website @ http://www.eia.gov/. Inflation data used in this project can be found @ http://inflationdata.com/. All of the calculations and data for this project are in the accompanying excel file.

Data

My first step was to chart and analyze data for gasoline prices. I looked at the US city Average retail price for regular unleaded gasoline. I chose regular unleaded because it was sold continuously since 1976. (Prior to 1976, gasoline on the market was leaded) The data can be seen from the chart below.

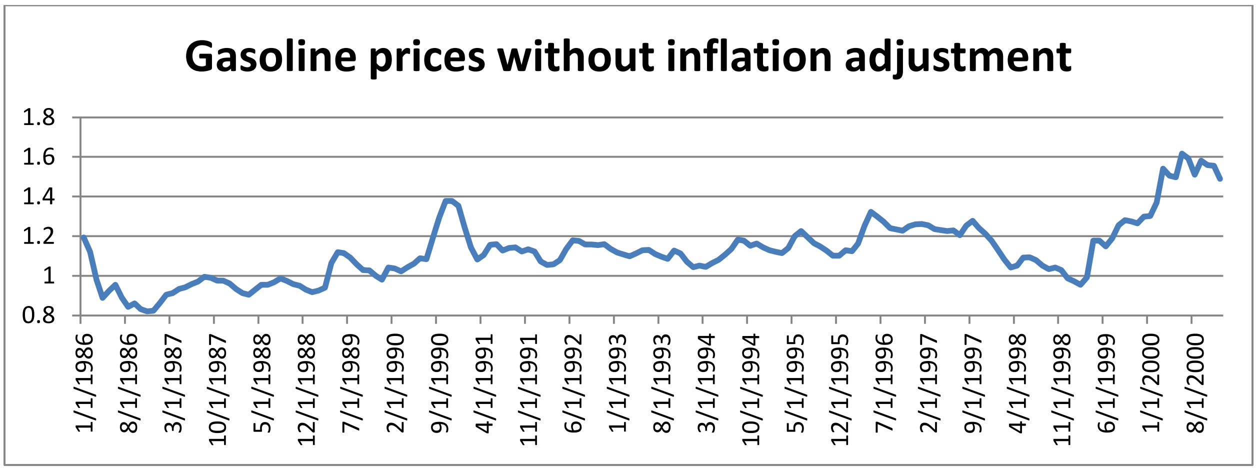


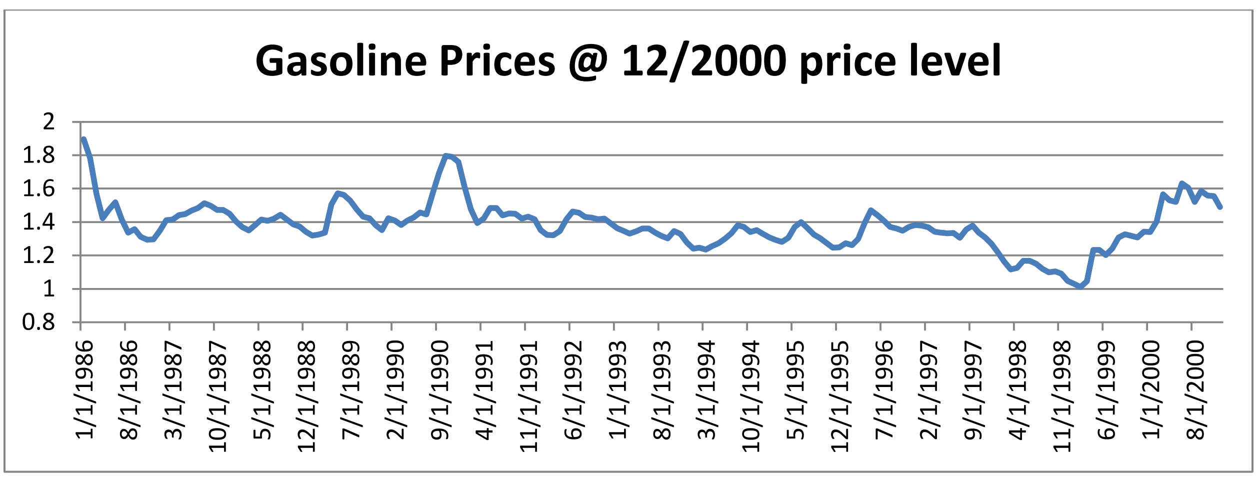
As can be seen from the graph, gasoline prices have gone through various phases over the past few decades. The first important event on the graph is the jump in prices during the 1979 energy crisis, which was largely connected to political unrest in Iran. A recession followed the energy crisis, and prices got back to roughly their pre-1979 prices after the recession ended in the mid 80s. This was followed by a period of stability that ended on Sept. 11th 2001. Following 9/11, political instability, and the declining value of the dollar lead to a huge spike in oil prices. The financial crisis in 2008 caused a huge drop in gasoline prices, due to less demand for gasoline for a nation entering a deep recession (investor speculation of a drop in demand probably also accelerated the price drop) Oil prices recovered after the worst years of the financial crisis past, and have been relatively stable since.

I found that because gasoline prices have gone through many phases, it did not make sense to use a single model for 1976 and on. I decided to focus on the period of 1986 – 2000, because this data was the most stable, and probably best explained by a non-complex time series model.

Trend

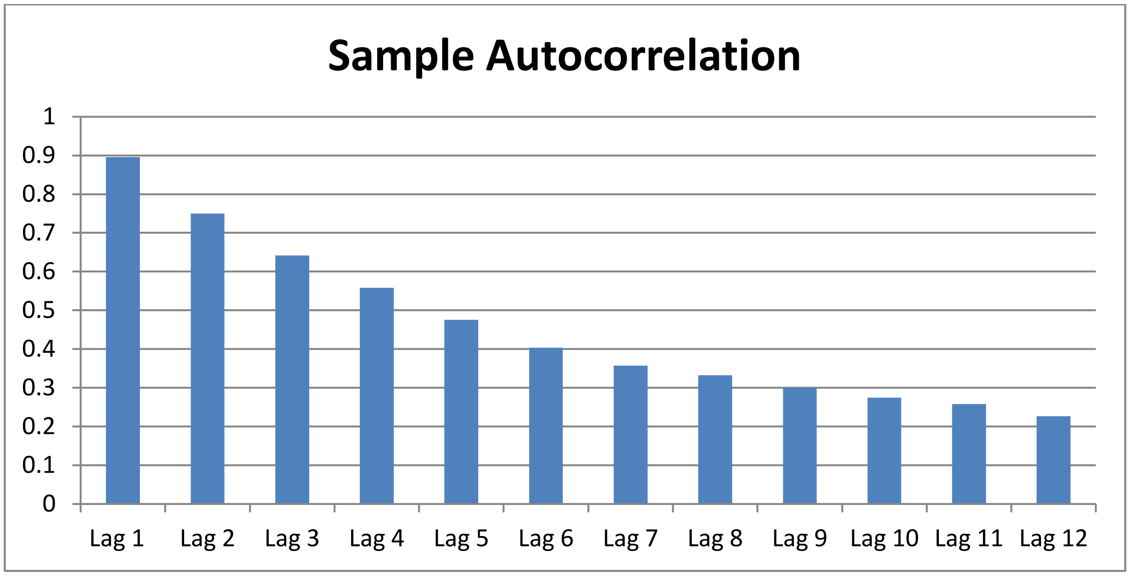
As can be seen from the above graph, gasoline prices have a positive trend. This trend can be seen even during the relatively stable years from 1986-2000. To make the series stationary, I remove the trend by applying a factor from an inflation index. As can be seen from the tables below, the inflation adjustment removed the positive trend from the data.





Seasonality

When beginning the project, I assumed that there would be some seasonality in gasoline prices. I believed that increased travel during the summer months would lead to greater demand, causing higher gasoline prices. To check for evidence of seasonality, I looked at the sample autocorrelations from the data at various lags. The results are displayed below.



This table does not show evidence of seasonality. If there was a strong seasonality effect, then there would have been an increase in the sample autocorrelations as the lag approached 12.

To further test for seasonality, and to look for the best model, I tried various AR models using excel's linear regression tool.

Models

My first model was an AR(1) model. To come up for parameters for this model, I regressed gasoline prices on the gasoline prices from a month prior. This resulted in the following model.

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.929815							
R Square	0.864556							
Adjusted F	0.863791							
Standard E	0.048897							
Observatio	179							
ANOVA								
	df	*SS*	*MS*	*F*	*ignificance F*			
Regressior	1	2.701242	2.701242	1129.811	9.32E-79			
Residual	177	0.423186	0.002391					
Total	178	3.124428						
	Coefficients	*tandard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*Ipper 95.0%*
Intercept	0.144011	0.036822	3.911038	0.000131	0.071345	0.216677	0.071345	0.216677
X Variable	0.893833	0.026592	33.61266	9.32E-79	0.841354	0.946311	0.841354	0.946311

The Φ value was for this model is .89

I tried an AR(2) model using the following regression

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.941505							
R Square	0.886432							
Adjusted F	0.885134							
Standard E	0.04381							
Observatio	178							
ANOVA								
	df	*SS*	*MS*	*F*	*ignificance F*			
Regressior	2	2.621637	1.310818	682.9637	2.16E-83			
Residual	175	0.335879	0.001919					
Total	177	2.957516						
	Coefficients	*tandard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*Ipper 95.0%*
Intercept	0.153232	0.034462	4.446417	1.55E-05	0.085217	0.221246	0.085217	0.221246
X Variable	-0.43028	0.065058	-6.61389	4.38E-10	-0.55868	-0.30188	-0.55868	-0.30188
X Variable	1.318236	0.067503	19.52855	8.08E-46	1.185012	1.451461	1.185012	1.451461

For this AR(2) model, the Φ1 value is 1.31, and the Φ2 value is -.43028. This model is stationary, because

-.43028 + 1.31 < 1 and

-.43028 - 1.31 < 1 and

.43028 < 1.

I prefer this model because the adjusted R Square value is higher than the first model.

Seasonal Model

I tried some seasonal models to see if a model that incorporated seasonality, would be a better predictor of future prices. My first model was an AR model, with a Φ1 variable, along with a Φ12 variable for the seasonal effect. I used excels regression tool and regressed gas prices on prior prices with lags of 1 month and 12 months. It resulted in the following model.

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.935811							
R Square	0.875742							
Adjusted F	0.874236							
Standard E	0.045605							
Observatio	168							
ANOVA								
	df	*SS*	*MS*	*F*	*ignificance F*			
Regressior	2	2.418563	1.209281	581.4407	1.91E-75			
Residual	165	0.343167	0.00208					
Total	167	2.76173						
	Coefficients	*tandard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	0.063166	0.046339	1.363146	0.174694	-0.02833	0.154659	-0.02833	0.154659
X Variable	0.869458	0.026933	32.28172	3.33E-73	0.816279	0.922637	0.816279	0.922637
X Variable	0.08173	0.027745	2.94577	0.003688	0.026949	0.136511	0.026949	0.136511

The $\Phi 1$ value is .869, and the $\Phi 2$ value is -.082. Although the adjusted R squared was higher than the AR(1) model, it was not higher than the AR(2) model.

I also tried two additional linear regression models (These are not time series models that were discussed in the course. However I wanted to see if other models that incorporate seasonality can be a better predictor of gas prices.) These linear regression models use three dummy variables to represent either winter, spring, summer, or fall. The first model regressed on the prior month, while the second regressed on the prior two months. Neither model produced better results than the AR(2) model shown earlier. The regression models are below.

Linear Regression using prior month variable, and 3 seasonal dummy variables:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.939146
R Square	0.881995
Adjusted F	0.879283
Standard E	0.046032
Observatio	179

ANOVA

	df	*SS*	*MS*	*F*	*ignificance F*
Regressior	4	2.755731	0.688933	325.1295	1.4E-79
Residual	174	0.368697	0.002119		
Total	178	3.124428			

	Coefficients	*tandard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*Ipper 95.0%*
Intercept	0.124234	0.034929	3.556728	0.000484	0.055294	0.193174	0.055294	0.193174
X Variable	0.894644	0.025273	35.39872	2E-81	0.844763	0.944526	0.844763	0.944526
X Variable	0.046468	0.009763	4.759732	4.06E-06	0.027199	0.065737	0.027199	0.065737
X Variable	0.018944	0.00983	1.927083	0.055598	-0.00046	0.038347	-0.00046	0.038347
X Variable	0.008807	0.009802	0.898477	0.370173	-0.01054	0.028152	-0.01054	0.028152

Linear regression using 2 variables for prior months, and seasonal dummy variables:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.979638
R Square	0.959691
Adjusted R Square	0.958519
Standard Error	0.027004
Observations	178

ANOVA

	df	*SS*	*MS*	*F*	*ignificance F*
Regression	5	2.986068	0.597214	818.9975	6.7E-118
Residual	172	0.125423	0.000729		
Total	177	3.111491			

	Coefficients	*tandard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*Ipper 95.0%*
Intercept	-0.03692	0.022342	-1.65246	0.100266	-0.08102	0.007181	-0.08102	0.007181
X Variable 1	0.518253	0.025542	20.29	1.67E-47	0.467836	0.56867	0.467836	0.56867
X Variable 2	0.501706	0.027537	18.21931	5.29E-42	0.447352	0.55606	0.447352	0.55606
X Variable 3	0.01776	0.00594	2.990083	0.003198	0.006036	0.029484	0.006036	0.029484
X Variable 4	0.007574	0.0058	1.30574	0.193385	-0.00388	0.019023	-0.00388	0.019023
X Variable 5	0.011758	0.005779	2.034797	0.043407	0.000352	0.023164	0.000352	0.023164

Conclusion

From the above models, I prefer the AR(2) model, because the regression used to create the model has the highest R Squared from all of the choices. However, this model does not completely explain the data. I created a quantile plot to test if the residuals are normal. Although the plot seems close to normal, there is some evidence of a thick tail.