MATH 642 - Data Project - yl714 – 15%

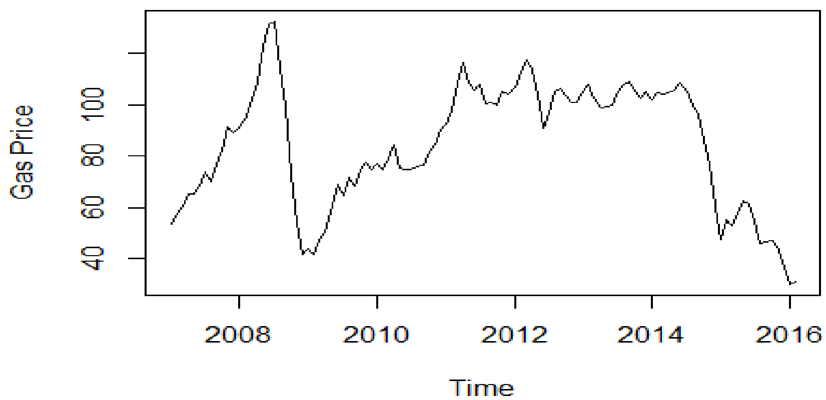
Neal (Yuchen) Li

April 13, 2016

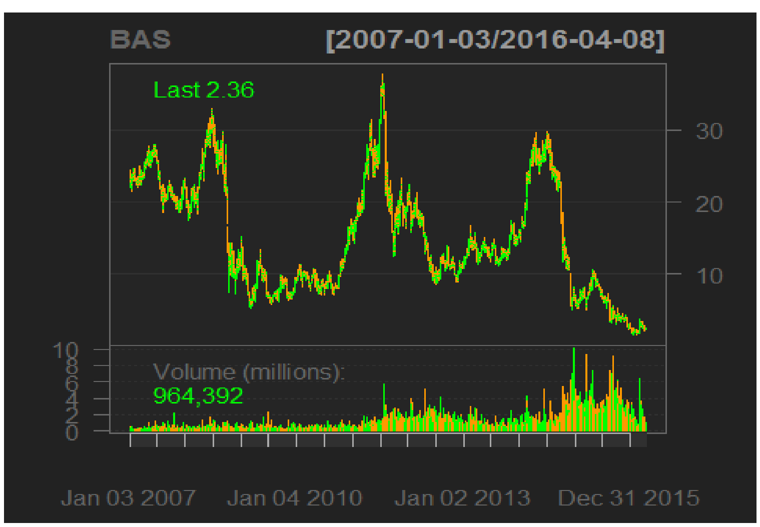
Is Now The Right Time To Buy Oil Stocks? Insight from Historical Data

Project Background:

It is universally agreed that crude oil is the most important commodities, hence its price movement becomes a crucial financial determinant. The following graph shows the oil price from January 2007 till Present (US Dollar/Barrel).



BAS (Basic Energy) is one of the oil stocks I am currently investing in, and the following graph is the price of BAS from January 2007 till Present, which bears great resemblance to oil price.



Project Goals and Customers:

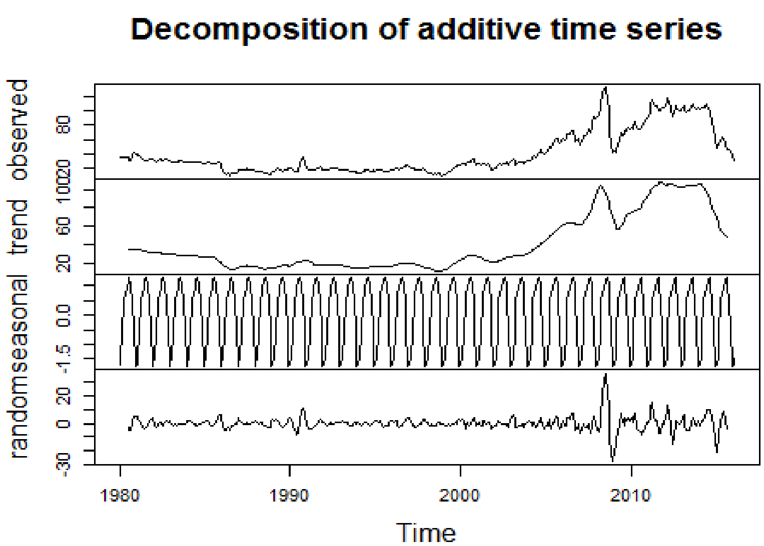
To provide short-term forecasts of oil prices based on historical prices from January 1980 to February 2016, and, to give recommendations (i.e. buy, hold or sell) for oil stock investors.

Holt-Winters Exponential Smoothing

If you have a time series that can be described using an additive model with increasing or decreasing trend and seasonality, you can use Holt-Winters exponential smoothing to make short-term forecasts.

Holt-Winters exponential smoothing estimates the level, slope and seasonal component at the current time point. Smoothing is controlled by three parameters: alpha, beta, and gamma, for the estimates of the level, slope b of the trend component, and the seasonal component, respectively, at the current time point. The parameters alpha, beta and gamma all have values between 0 and 1, and values that are close to 0 mean that relatively little weight is placed on the most recent observations when making forecasts of future values.

First, let's check if oil price can be described using an additive model with trend and seasonality.



The plot shows that historical oil price can be described as an additive model with increasing trend, seasonality and white noise.

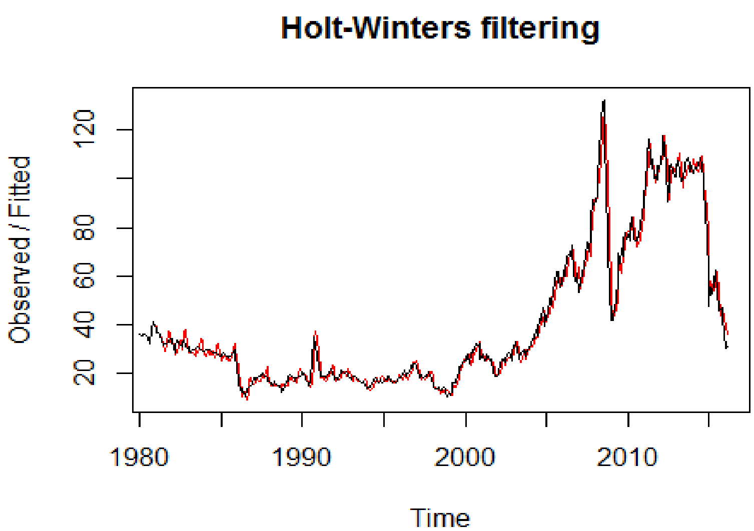
To make forecasts, we can fit a predictive model using the HoltWinters() functions.

```
## Holt-Winters exponential smoothing with trend and additive seasonal component.
##
## Call:
## HoltWinters(x = oil.price.ts)
##
## Smoothing parameters:
##  alpha: 0.8857118
```

```
##  beta : 0
##  gamma: 1
```

The estimated value of alpha, beta and gamma are 0.885, 0, and 1, respectively. The value of alpha (0.885) is relatively high, indicating that the estimate of the level at the current time point is based more upon recent observations than observations in the more distant past. The value of beta is zero, indicating that the estimate of the slope beta of the trend component is not updated over the time series, and instead is set equal to its initial value. This makes good intuitive sense, as the level changes quite a bit over the time series, but the slope beta of the trend component remains roughly the same. In contrast, the value of gamma (1) is high, indicating that the estimate of the seasonal component at the current time point is just based upon very recent observations.

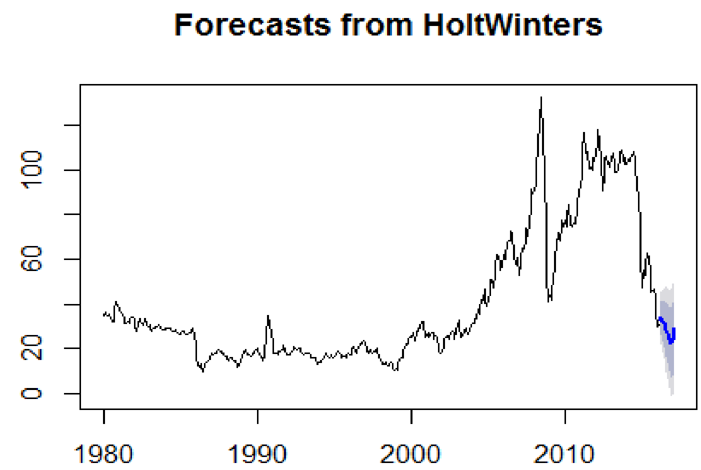
As for Holt Winter's exponential smoothing, we can plot the original time series as a black line, with the forecasted values as a red line on top of that:



(a) Forecasting using Holt-winters exponential smoothing

To make forecasts for future times not included in the original time series, we use the "forecast.HoltWinters()" function in the "forecast" package. For example, the original data for the oil prices is from January 1980 to February 2016. If we wanted to make forecasts for March 2016 and the following three month, we would have:

```
##          Point Forecast     Lo 80    Hi 80      Lo 95    Hi 95
## Mar 2016       32.60371 27.098737 38.10868 24.1845827 41.02283
## Apr 2016       33.91432 26.560523 41.26811 22.6676603 45.16097
## May 2016       32.41154 23.588172 41.23491 18.9173622 45.90572
## Jun 2016       31.28251 21.201563 41.36345 15.8650337 46.69998
```



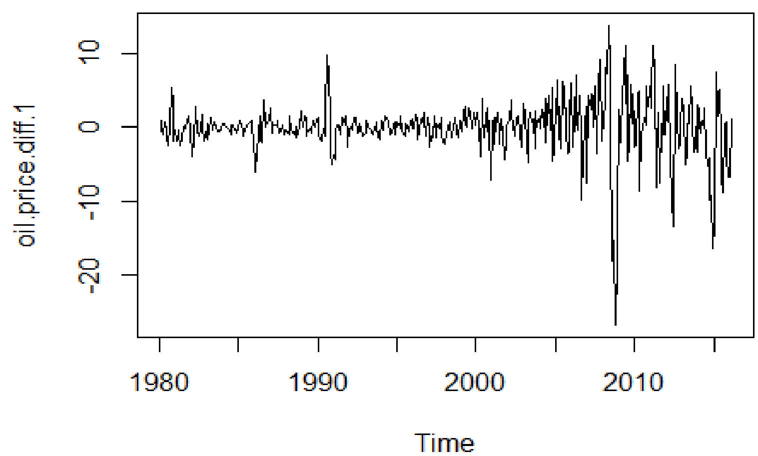
The forecasts are shown as a blue line, and the blue and grey shaded areas show 80% and 95% prediction intervals, respectively.

ARIMA Models

Since the forecast errors show more drift from zero than before, let's consider Autoregressive Integrated Moving Average(ARIMA) models which allow correlated error terms.

Because ARIMA models are defined for stationary time series. Therefore, if oil price is non-stationary, you will first need to 'difference' the time series until you obtain a stationary time series. If you have to difference d times to obtain a stationary time series, then you have an ARIMA(p,d,q) mode, where d is the order of differencing used.

Let's first difference oil price once, and plot the differenced series:



It seems than the first difference of oil price is stationary in mean, which centered around 0.

(a) Selecting a suitable ARIMA model

Since we have loosely obtained stationarity by observing the differencing plot, it's time to figure out an appropriate ARIMA model, ARIMA(p,d,q), where q and q are undetermined. Luckily, auto.arima() function can be used to find the appropriate ARIMA model and its output will suggest the value of p,d and q.

```
## Series: oil.price.ts
## ARIMA(2,1,2)(0,0,2)[12]
##
## Coefficients:
##          ar1      ar2      ma1     ma2    sma1     sma2
##       1.4723  -0.5867  -1.0334  0.1538  0.0914  -0.1301
## s.e.  0.0870   0.0785   0.1040  0.0959  0.0535   0.0515
##
## sigma^2 estimated as 11.56:  log likelihood=-1141.68
## AIC=2297.37   AICc=2297.63   BIC=2325.86
```

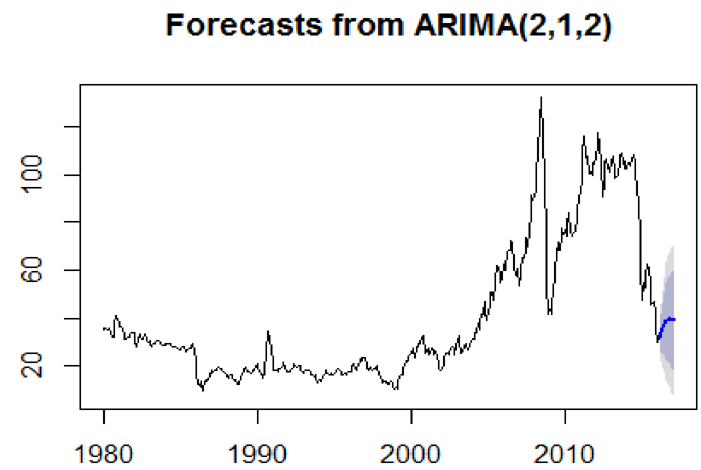
Since p,d,q are 2,1,2 respectively, we can conclude that the first difference of oil price is indeed a stationary time series.

(b) Forecasting using ARIMA(2,1,2) model

First, we fit an ARIMA(2,1,2) model to oil price

Now let's forecast the price of oil using the model we just obtained.

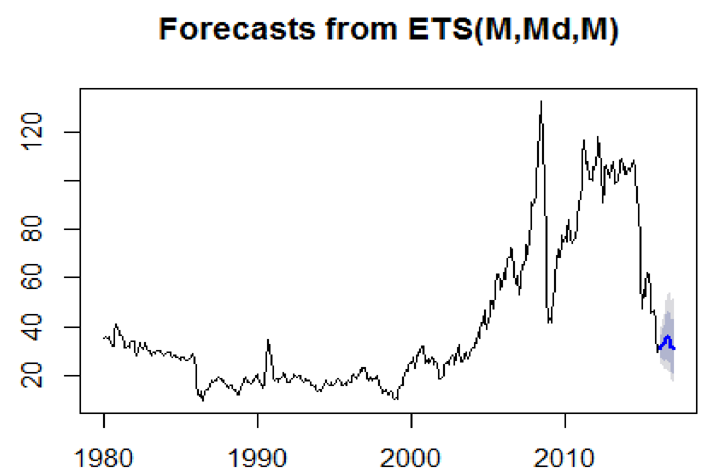
```
## Point          Forecast    Lo 80    Hi 80     Lo 95    Hi 95
## Mar 2016       32.23757 27.85420 36.62094 25.533785 38.94135
## Apr 2016       33.95556 26.35205 41.55907 22.326999 45.58413
## May 2016       35.71643 25.30002 46.13285 19.785908 51.64696
## Jun 2016       37.22481 24.48382 49.96580 17.739142 56.71048
```



ETS

ETS is a more recent R package developed by Dr. Rob Hyndman, ETS (Exponential Smoothing State Space Model) gained its popularity over the year through his famous paper: Automatic Time Series Forecasting: The Forecast Package for R.

(a) Forcasting using ETS()



Model Selection

Time series cross-validation answers two important questions:

1. We have used Holtwinters, ARIMA and ETS model, which one is the best?

2. Some of models may require tuning during model training, which tuning parameter values should we choose?

Here I compare the Mean Absolute Error(MAE) of each model on different horizons.

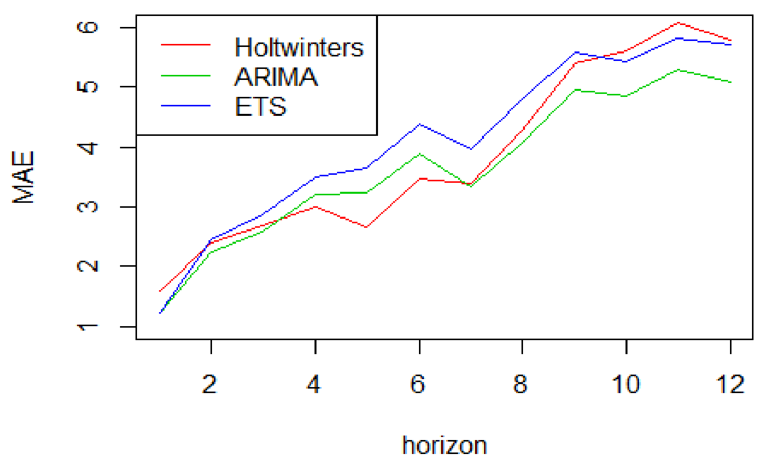
According to Dr. Hyndman, time-series cross-validation follows the following steps:

Assume k is the minimum number of observations for a training set.

(1) Select observation k+i for test set, and use observations at times 1,2,..., k+i-1 to estimate model. Compute error on forecast for time k+i.

(2) Repeat for i = 0,1,..., T-k where T is total number of observations.

(3) Compute accuracy measure (MAE) over all errors. 3.12



The MAE plot shows that all three models (Holtwinters, ARIMA, ETS) has increasing MAE as horizon goes up, which makes perfect sense because the further into the future, the less forecasting power a model has. In terms of selecting model, ETS has the smallest MAE before 4 horizons, while holtwinters has between 4 and 7, ETS between 8 and 9, and ARIMA after 9. As a result, in terms of short term(less than 4 horizons (month)) forecasting, ETS beats all other model regarding MAE obtained by cross validation.

```
##           Jan      Feb      Mar      Apr      May      Jun      Jul
## 2016                   31.01902 31.98063 32.86507 33.10071 33.84051
## 2017 31.55979 30.91661
##           Aug      Sep      Oct      Nov      Dec
## 2016 35.32140 35.71304 35.37855 34.07092 31.73898
## 2017
```

ETS model predicts that oil price will level off in March, April, May and slightly decrease in June, in conclusion, oil price will be facing continuous downward pressure, hence investor should carefully consider investment decision and lean towards selling.