Janani Sundaresan

COMP 6212
MSc Data Science
Assignment 2 – Computational Finance
University of Southampton
js10n18@soton.ac.uk

ABSTRACT

Derivatives Pricing and Time Series Analysis using Kalman Filter is performed here using the S & P Index 100 data.

1.a DERIVATIVES PRICING

1.1 Derivative of Normal Distribution

In a Normal Distribution $\mathcal{N}(x)$, the data points cluster around the centre(mean). $\mathcal{N}(x)$ is the cumulative distribution function for a normal distribution.

The derivative of $\mathcal{N}(x)$ is denoted by $\mathcal{N}'(x)$ .As standard deviation goes up, the term 'd1' goes up and the term d2 goes down.

$$\mathcal{N}(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Differentiation will cancel the integration and hence,

$$\boxed{\boldsymbol{\mathcal{N}'(x) = \frac{\partial \mathrm{d1}}{\partial \mathrm{x}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}}}$$

1.2 Proof -> $SN'(d1) = Ke^{-r(T-t)}N'(d2)$

We know $C = SN(d1) - Ke^{-r\tau}N(d2)$ and from the above section we know that,

$$\mathcal{N}'(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

$$\mathcal{N}'(\mathrm{d1}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{d1^2}{2}}$$

$$\mathcal{N}'(\mathrm{d2}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{d2^2}{2}}$$

Differentiating the LHS of the call option equation w.r.t time

$$\frac{\partial[SN(d_1)]}{\partial S} = \frac{\partial S}{\partial S} N(d_1) + S\frac{\partial N(d_1)}{\partial S}$$

$$= N(d_1) + \boldsymbol{S}\frac{\partial N(d_1)}{\partial S}$$

$$\frac{\partial N(\mathrm{d1})}{\partial S} = \frac{\partial N(d_1)}{\partial d1} \cdot \frac{\partial \mathrm{d1}}{\partial S}$$

$$\frac{\partial N(\mathrm{d2})}{\partial S} = \frac{\partial N(d2)}{\partial d2} \cdot \frac{\partial \mathrm{d2}}{\partial S}$$

Substituting the values for $\frac{\partial N(d_1)}{\partial d1}$ and $\frac{\partial N(d2)}{\partial S}$, we get

$$\boxed{\boldsymbol{SN'(d1) = Ke^{-r\tau}N'(d2)}}$$

1.3 Derivative of d1 and d2 w.r.t Asset S

$$\boldsymbol{d1 = \frac{log\left(\frac{S}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{(T-t)}}}$$

$$\boldsymbol{d2 = d1 - \sigma\sqrt{(T-t)}}$$

Differentiating d1 w.r.t S,

$$\boldsymbol{\frac{\partial d1}{\partial S} = \frac{\partial}{\partial S}\left(\frac{\log\left(\frac{S}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{(T-t)}}\right)}$$

Taking the constant in the denominator outside the differentiation, we get the following equation.

$$\boldsymbol{\frac{\partial d1}{\partial S} = \frac{1}{\sigma\sqrt{(T-t)}} \frac{\partial}{\partial S}\left(log\left(\frac{S}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)(T-t)\right)}$$

$$\boldsymbol{= \frac{1}{\sigma\sqrt{(T-t)}}\left|\frac{\partial}{\partial S}\left(log\left(\frac{S}{K}\right)\right) + \frac{\partial}{\partial S}\left(r + \frac{\sigma^2}{2}\right)(T-t)\right|}$$

The second term in the equation becomes zero since it is a constant.

We know that, $\log\frac{S}{K} = \log S + \log\frac{1}{K}$, Log S = $\frac{1}{S}$ and

Log $\frac{1}{K} = K$

Hence substituting the above values, we get the below equation.

$$\frac{\partial d1}{\partial S} = \frac{1}{S\sigma\sqrt{(T-t)}}$$

Differentiating $d2$ w.r.t S by substituting the value of $d1$,

$$\frac{\partial d2}{\partial S} = \frac{1}{S\sigma\sqrt{(T-t)}} - \sigma\sqrt{(T-t)}$$

The second term of above equation is zero since it has no 'S' term.

$$\frac{\partial d2}{\partial S} = \frac{1}{S\sigma\sqrt{(T-t)}}$$

Therefore

$$\boxed{\begin{aligned}\frac{\partial d1}{\partial S} &= \frac{\partial d2}{\partial S} \\ &= \frac{1}{S\sigma\sqrt{(T-t)}}\end{aligned}}$$

1.4 First Order Derivative of Call Price w.r.t Time

We know $C = SN(d1) - Ke^{-r\tau}N(d2)$

Substituting $\tau = T - t$ and differentiating it w.r.t to Asset,

$$\begin{aligned}\frac{\partial C}{\partial S} &= N(d1) + S\frac{\partial N(d_1)}{\partial S} - Ke^{-r\tau}\frac{\partial N(d2)}{\partial S} \\ &= N(d1) + S\frac{\partial N(d1)}{\partial d1}\frac{\partial d1}{\partial S} - Ke^{-r\tau}\frac{\partial N(d2)}{\partial d2}\frac{\partial d2}{\partial S} \\ &= N(d1) + S\frac{1}{\sqrt{2\pi}}e^{-\frac{d1^2}{2}}\cdot\frac{1}{S\sigma\sqrt{\tau}} - Ke^{-r\tau}\frac{1}{\sqrt{2\pi}}e^{-\frac{d_1^2}{2}}\cdot\frac{S}{K} \\ &\quad \cdot e^{r\tau}\cdot\frac{1}{S\sigma\sqrt{\tau}} \\ &= N(d1) + S\frac{1}{S\sigma\sqrt{2\pi\tau}}e^{-\frac{d1^2}{2}} - S\frac{1}{S\sigma\sqrt{2\pi\tau}}e^{-\frac{d1^2}{2}} \\ &= N(d1)\ [4]\end{aligned}$$

Hence proved,

$$\boxed{\frac{\partial C}{\partial S} = N(d1)}$$

$$\frac{\partial C}{\partial \tau} = SN'(d1)\frac{\partial d1}{\partial \tau} + rKe^{-r\tau}N(d2) - Ke^{-r\tau}N'(d2)\frac{\partial d2}{\partial \tau}$$

Substituting $\frac{\partial d2}{\partial \tau} = \frac{\partial d1}{\partial \tau} - \frac{\sigma}{2\sqrt{\tau}}$

$$\frac{\partial C}{\partial \tau} = SN'(d1)\frac{\partial d1}{\partial \tau} + rKe^{-r\tau}N(d2) - Ke^{-r\tau}N'(d2)\frac{\partial d1}{\partial \tau} + Ke^{-r\tau}N'(d2)\frac{\sigma}{2\sqrt{\tau}}$$

Substituting the value for $SN'^{(d1)}$, we get

$$\boxed{\frac{\partial C}{\partial \tau} = -rKe^{-r\tau}N(d2) - SN'(d1)\frac{\sigma}{2\sqrt{\tau}}}$$

1.5 Second Order Derivative of Call Price w.r.t Time

From the previous section, we know

$$\frac{\partial C}{\partial S} = N(d1)$$

Again differentiating it w.r.t to Asset

$$\frac{\partial^2 C}{\partial S^2} = \frac{\partial}{\partial S}\left(\frac{\partial C}{\partial S}\right)$$

$$\frac{\partial^2 C}{\partial S^2} = \frac{\partial N(d1)}{\partial S}$$

Using the chain rule,

$$\frac{\partial N(d1)}{\partial S} = \frac{\partial N(d1)}{\partial d1}.\frac{\partial d1}{\partial S}$$

From the section 1.1, we know,

$$\frac{\partial N(d1)}{\partial d1} = \frac{1}{\sqrt{2\pi}}.e^{-\frac{-d1^2}{2}} = N'(d1)$$

$$\frac{\partial d1}{\partial S} = \frac{1}{S\sigma\sqrt{(T-t)}}$$

Therefore,

$$\boxed{\frac{\partial^2 C}{\partial S^2} = N'(d1).\frac{1}{S\sigma\sqrt{(T-t)}}}$$

Black Scholes Differential equation is given as,

$$\boxed{\frac{\partial C}{\partial \tau} + \frac{1}{2}\sigma^2 S^2\frac{\partial^2 C}{\partial S^2} + rS\frac{\partial C}{\partial S} - rC = 0}$$

Substituting the values for all the terms based on the previously defined equations, we get zero thereby proving Black Scholes differential equation is same as call option price equation.

1.b Black Scholes Option Pricing

European Call option: The individual possesses the right but there is no obligation to buy the stock for a price P at a time T.

European Put option: The individual possesses the right but there is no obligation to sell the stock for a price P at a time T.

The value of the stock at a given time T depends on many factors, predominantly on Strike Price, how much time is left for the stock to mature, rate of interest, dividends paid and the annual volatility. Usually price increases with volatility

Let us consider a European Call and Put option for Strike Prices 6800 and 7600 respectively. Stock options are very valuable when you are dealing with a stock that has high standard deviation.

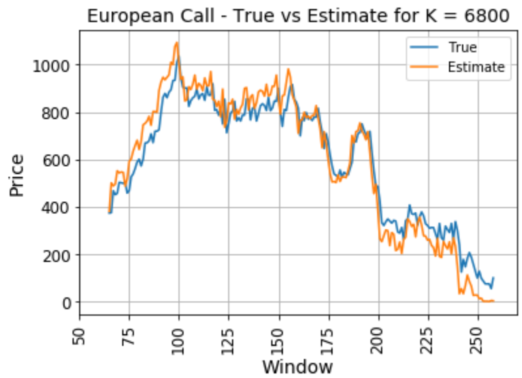

Fig 1.b – 1 – Call Option K = 6800

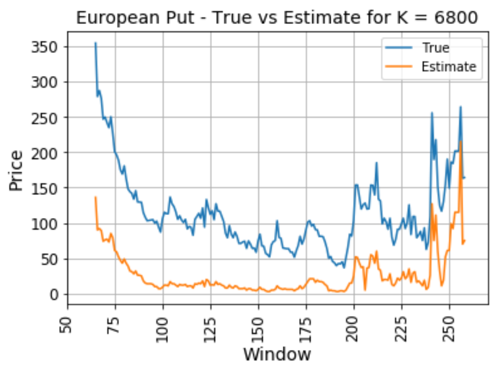

Fig 1.b – 2 – Call Option K = 6800

The Black Scholes model is used to calculate the Call and Put Option Price for Strike rate 6800 for the period $\frac{T}{4} + 1$ to T. The comparison between the true and the estimated prices are shown in the figures Fig 1.b – 1 and In Fig 1.b – 2. We can see the estimate price is almost same as true price for call option and it is almost same when the days to expire is between 160th and 200th days but in case of put price there is a large deviation between true and estimated price. But the deviation decreases as the days to expire decreases.

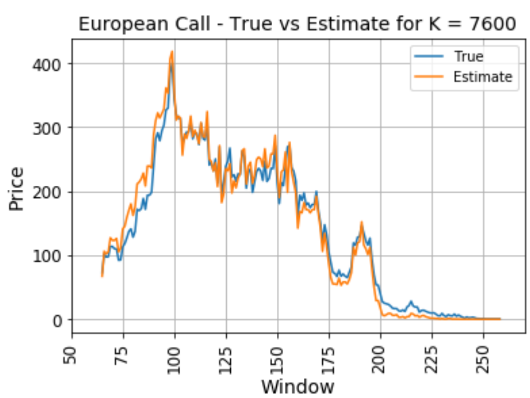

Fig 1.b – 3 – Call Option K = 7600

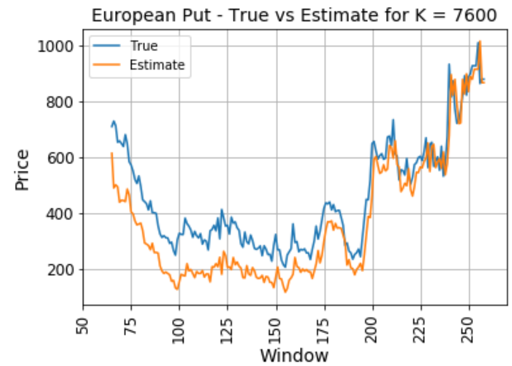

Fig 1.b – 4 – Call Option K = 7600

The comparison between the true and the estimated prices are shown in the figures Fig 1.b – 3 and In Fig 1.b – 4. We can see the estimate price is almost same as true price for call option throughout the period under consideration but in case of put price there is a large deviation between true and estimated price. But the deviation decreases as the days to expire decreases and true and estimate price are almost same.

1.c Volatility Smile

Implied Volatility is defined as the standard deviation of the asset price that makes the strike price of the call/put option same as the price computed using the Black Scholes. Here, implied volatility is calculated for a 30-day period starting from the 231st day to 260th day in the period $\frac{T}{4} + 1$ to T under consideration for analysis.

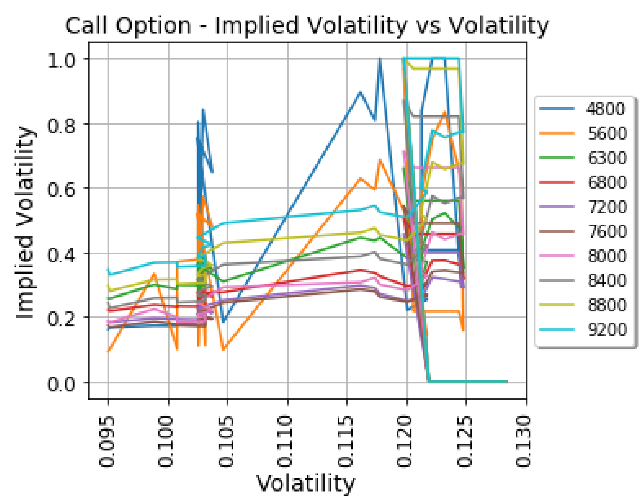

Fig 1.c – 1 – Call Option Implied Volatility vs Volatility

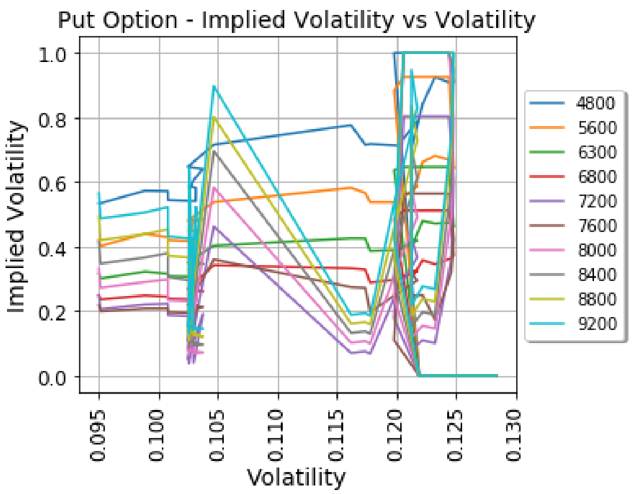

Fig 1.c – 2 – Call Option Implied Volatility vs Volatility

In Fig 1.c – 1 and In Fig 1.c – 2, we can see the volatility plotted again the implied volatility for each of the Strike Prices for Call and Put options.

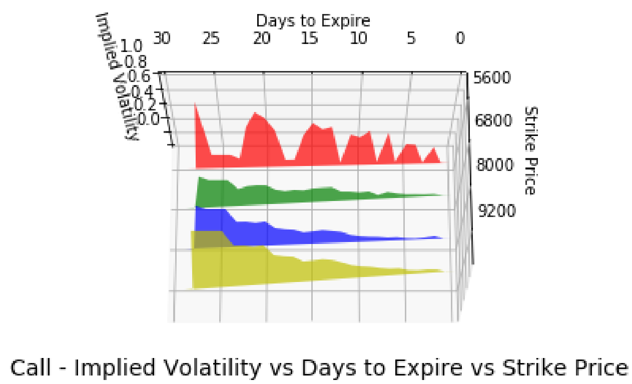

Fig 1.c – 3 – Call Option Volatility Curve

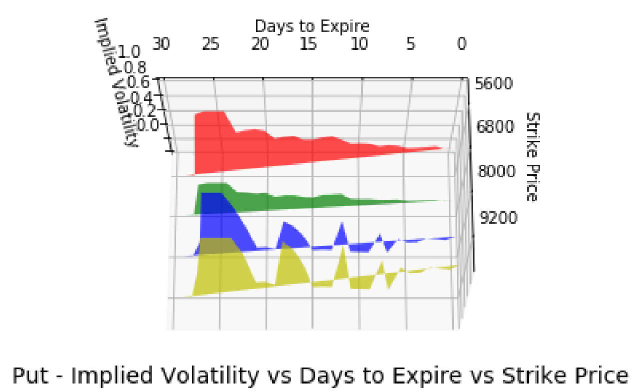

Fig 1.c – 4 – Put Option Volatility Curve

Implied Volatility has a linear relationship with maturity. As the number of days to mature increases, the implied volatility also increases. We can see that the implied volatility is reducing as the number of days to expire reduces. In Fig 1.c – 3 and In Fig 1.c – 4, the 3D plot for implied volatility, strike price and days to expire are plotted for put and call options. In case of Call options, we can see the implied volatility curve becomes smooth as the Strike price increases and the implied volatility reduces as the time to expire reduced for each of the Strike Price. In case of Put options, we can see the implied volatility curve becomes less smooth as the Strike price increases and the implied volatility reduces as the time to expire reduces for each of the Strike Price.

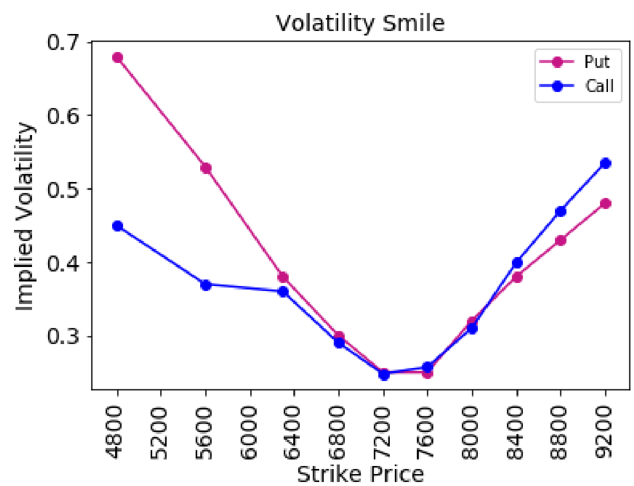

Fig 1.c – 5 –Implied Volatility vs Strike Price

The average implied volatility for each strike price was calculated and plotted. From the above Fig 1.c – 5, we can observe a nice volatility smile for Call and Put Options for various Strike Prices indicating the Call options are traded for higher prices. Also, we can see that the implied volatility for the Put option was higher than the Call option up to a Strike Price of 7200 after which the implied volatility for Call option is higher than the implied volatility for Put option. The volatility smile does not have any skewness and the smile is symmetry.

1.d Black Scholes vs Binomial Lattice

In Binomial Lattice, “we model the price of a stock in discrete time by a Markov chain of the recursive form Sn+1 = SnYn+1, n ≥ 0, where the {Yi} are iid with distribution P(Y = u) = p, P(Y = d) = 1−p.” [2]

In Black Scholes model, “the option’s value is a function of five variables, there are current security price (S), exercise price (X), time to expiration (T), risk-free rate (RFR), and security price volatility (σ). Hence, the Black-Scholes model holds that $c = f(S, X, T, RFR, \sigma)$. *S* and *RFR* are observable market prices, and *X* and *T* are defined by the contract itself.” [3]

In Fig 1.d -1 and Fig.1.d – 2, we can see the plot for the Call and Put option prices calculated using Black Scholes and Binomial models and plotted against the true prices. In both plots, we can see Black Scholes model predicted the dip and the highs before the binomial model. We can observe in the plots that the call price reduces as the number of days to expire reduces and the put price increases as the number of days to expire decreases. [5]

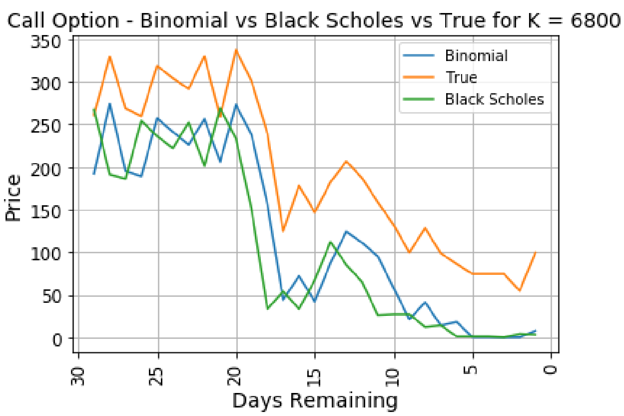

Fig 1.d – 1 – Binomial vs Black Scholes for Call Option

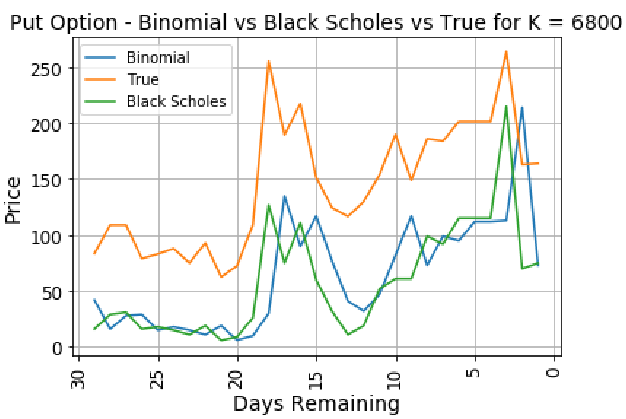

Fig 1.d – 2 – Binomial vs Black Scholes for Put Option

3 TIME SERIES ANALYSIS

3.1 Auto Regressive Model

Here Time Series Analysis is performed on the S&P Index 500 data from 1995 to end of 2015 at monthly interval. Once the data is read, it was converted into Time Series by mentioning the frequency of the data as 12, the start year as 1995. Usually to perform time series analysis that data should be stationary. In the Figures 3.1 - a and 3.1 - b, we can see the S&P Index 500 data for the time period 1995 to 2015. The blue points represent the point of time in which the index dipped and the red points represent the point of time in which the index price increased. We can infer the stock price dipped and increased at irregular intervals indicating high volatility and non-stationary behavior.

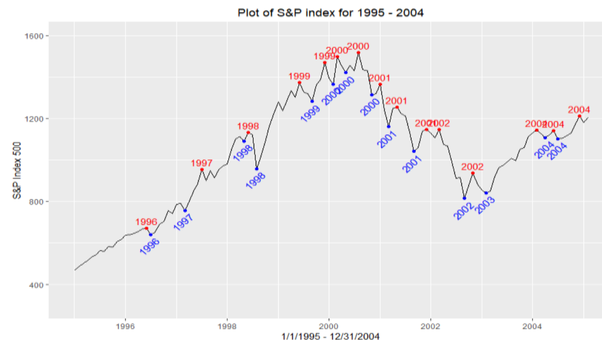

Fig 3.1 – a – Plot of S&P index from 1995 to 2004

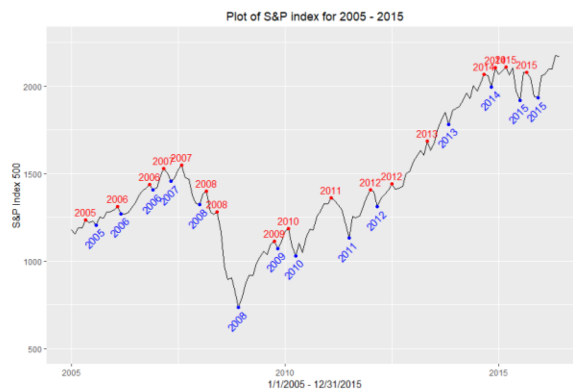

Fig 3.1 – b – Plot of S&P index from 2005 to 2015

ADF were performed to identify if the time series is non-stationary. The p-value obtained was than 0.05 indicating the data is not stationary. Hence the data has to be differenced till the p-value falls below 0.05. The ADF test was again performed on the first order differenced data and the value of p obtained was less than 0.05. The first order differenced data is shown in Fig 3.1- c.

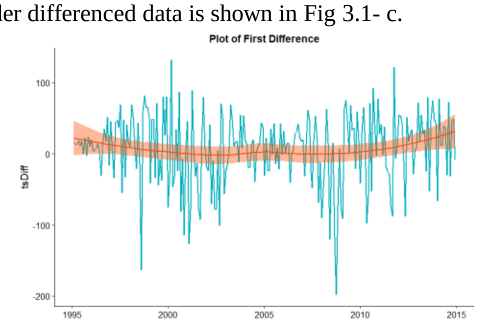

Fig 3.1 – c – First order Difference plot of S&P index

The ACF and decomposition plots were plotted to check the stationarity of the data. In Fig 3.1-d, we can see ACF plot before differencing has positive values for higher number of lags indicating a presence of a trend and highly non-stationary behavior since it does not drop off to zero quickly after few lags and therefore requires higher order of differencing to make it stationary and ACF plot after differencing displays stationarity. Once the data is stationary, it was fitted using second and third order auto regressive model.

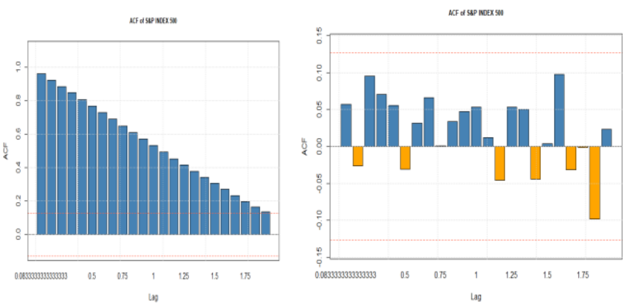

Fig 3.1 – d – ACF Plot before and after differencing

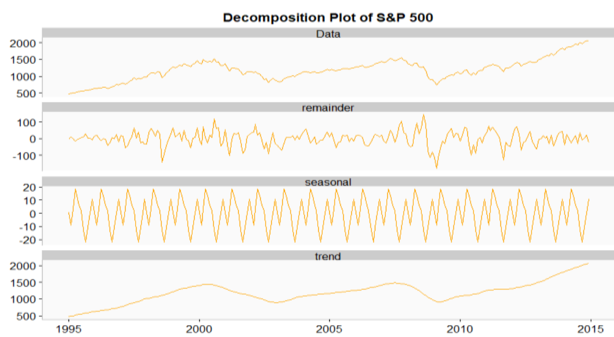

Fig 3.1 – e – Decomposition plot before differencing

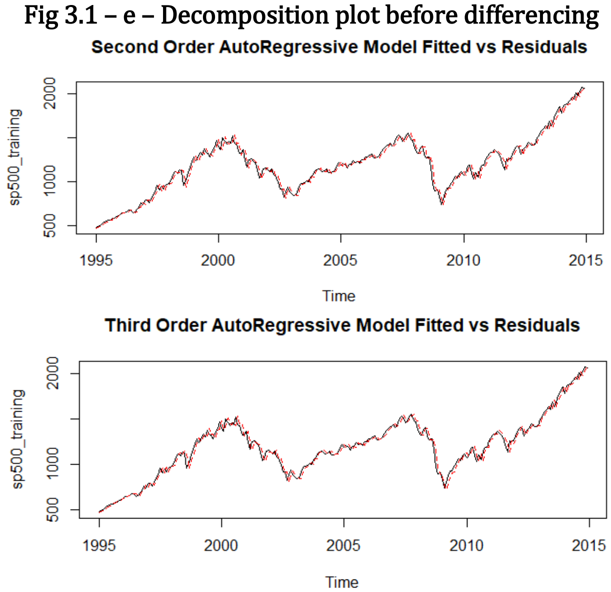

Fig 3.1 – f – Fitted vs Residual Plot

3.2 Estimation of AR using Kalman Filter

First modelled the residuals of an AR model. Used “tseries” library in R to analyze the residuals by the ARMA functions. The model fitted with residuals. Kalman Filter was used to estimate the unobserved component in the model. Used the “KalmanSmooth function” from the stats package in R by giving SS model as a parameter and the function returned the estimated AR model parameters recursively.

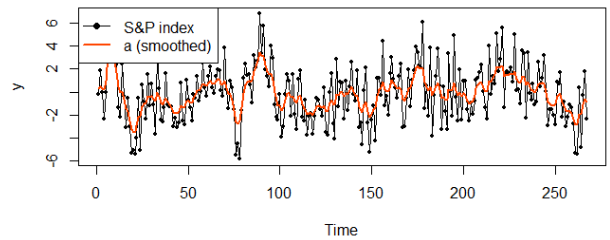

Fig 3.2 – a – Kalman Filter smoothed Plot of S&P index

3.3 Estimation of AR using Kalman Filter

In [3], the author tries to explain about usefulness of Kalman filter in giving the residual information. The data used in the paper is used here for analysis as well and the missing data rows where imputed with mean values. The S&P index 500 data is fitted using ARMA and the residuals obtained are plotted.

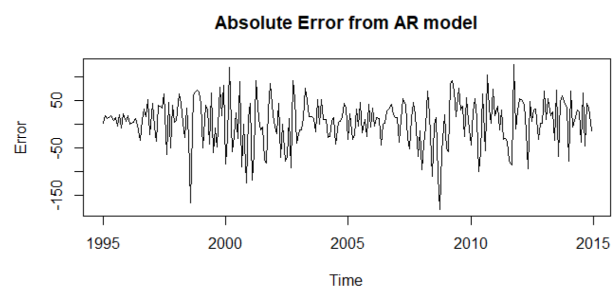

Fig 3.3 – a –Residual Plot

Python’s “LassoLarsCV” package was used to select features. From the below pictures we can see that PMI and Unemployment were removed at the end.

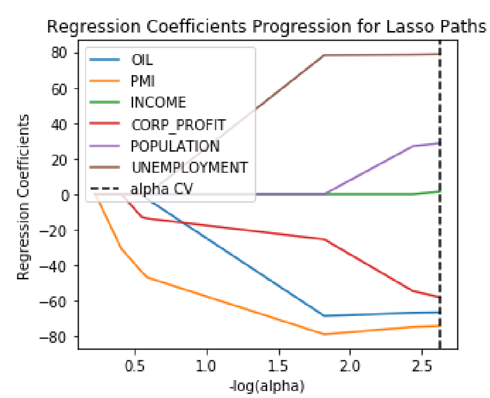

Fig 3.3 – b – Feature Selection using Lasso

The mean squared error is plotted for each of the ten-fold cross validation is shown in Fig 3.3 – b.

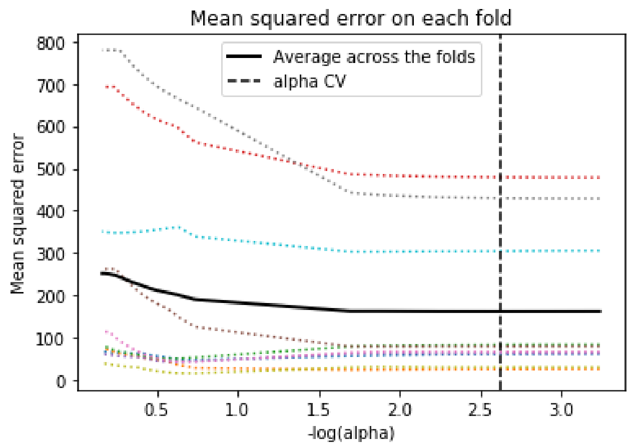

Fig 3.3 – c – MSE for 10—fold CV

The actual index values, the prediction from AR model and Kalman filter were plotted together by taking log transformation as shown in Fig 3.3 – c. We can see that both models predict close to the true value for the time period considered. [6]

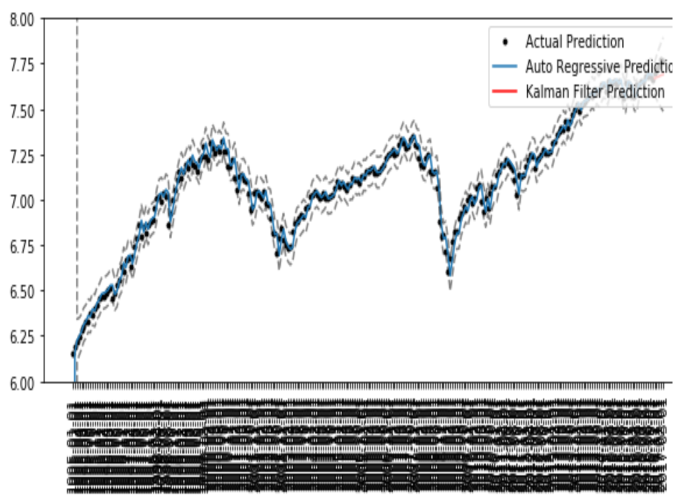

Fig 3.3 – d – S&P index vs Kalman Filter vs AR model

REFERENCES

[1] Sparse Index Tracking with R, Prof. Daniel P. Palomar

[2] Binomial lattice model for stock prices, by Karl Sigman, Columbia - http://www.columbia.edu/~ks20/FE-Notes/4700-07-Notes-BLM.pdf

[3] N. Mahler, \Modeling the S & P 500 index using the Kalman filter and the LagLasso," in Machine Learning for Signal Processing, 2009. MLSP 2009. IEEE International Workshop on, Sept 2009, pp.1{6.

[4] http://www.mkaranasos.com/FEGreeks.pdf

[5] “Review of Applying European Option Pricing Models”, Haochen Guo

[6] http://www.chadfulton.com/topics/state_space_python.html