DSA302 Financial Data Analytics

Prepared by: G2 Group 4

Student	Matriculation Number
Chia Dehan	01332535
Eugene Wong Wei Huan	01334933
Ng Kai Wen Jeremy	01336554
Lim Jing Zhe Austin	01335331
Tan You Joh	01327394

Prepared for:

Dr Tse Yiu Kuen
Professor, School of Economics
Singapore Management University

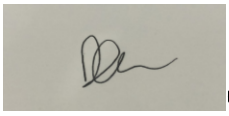
10 November 2020

Table of Contents

1. Declaration

- Each member of this group contributes honestly and fairly to the completion of this report.

Signed by:

Chia Dehan

Eugene Wong Wei Huan

Jeremy Ng Kai Wen

Lim Jing Zhe Austin

Tan You Joh

2. Part 1: Estimating the Yield Curve

In this section, we will attempt to estimate the yield curve of Zero-Coupon Bonds. Zero coupon bonds are priced according to the equation:

$$P(t) = 100e^{-r(t)t} \quad (1)$$

for a bond with a face value of 100 at time *t* where *r(t)* is the spot rate that we want to estimate.

Since the spot rate is related to prices through the forward rate, *f(t)*, in the equation:

$$r(t) = \frac{1}{t}\int_0^t f(s)ds \quad (2)$$

We must first estimate the forward rate for the data set that we are given.

2.1 Dataset:

The dataset we are given (ZCBP.txt) contains a set of zero-coupon bond prices (per 100 face value) with their corresponding maturities (in years) up to 30 years.

Exploratory Data Analysis:

1. We first checked if the data had any missing values

```
> table(is.na(ZCBP))
FALSE
 232
```

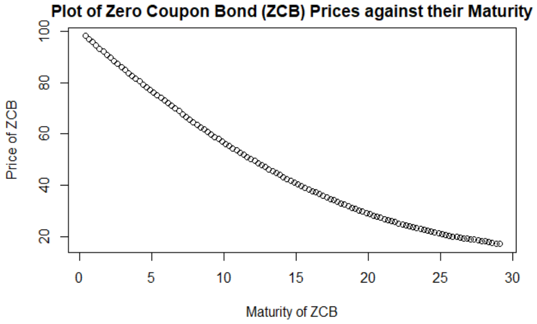
2. Next, we looked at the summary and first 5 rows of data and found that the data was not arranged by the time column.

```
> summary(ZCBP)
      time             price
 Min.   : 0.3699   Min.   :17.26
 1st Qu.: 7.5589   1st Qu.:25.55
 Median :14.7479   Median :41.50
 Mean   :14.7479   Mean   :47.16
 3rd Qu.:21.9370   3rd Qu.:65.98
 Max.   :29.1260   Max.   :98.16
```

```
> head(ZCBP)
  time    price
  <dbl>   <dbl>
1 0.6219  96.924
2 1.1260  94.511
3 1.6219  92.070
4 2.1260  89.644
5 2.6219  87.295
```

(1) Plot the bond prices versus their maturities.

We sorted the data by the time column first as this would help in the later parts before plotting the bond prices against the maturities to obtain the following scatterplot:



This plot shows that the zero-coupon bonds of longer maturities are priced lower than those with shorter maturities.

2.2 Forward Rates

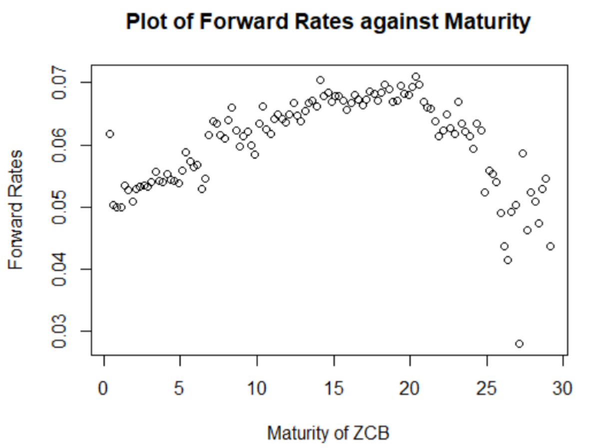
<u>(2) Plot the empirical forward rates as computed in equation (3) versus maturities.</u>

As mentioned earlier, we will first compute the empirical forward rates based on the dataset using the following formula:

$$f_j = -\frac{\log P_{j+1} - \log P_j}{t_{j+1} - t_j}, for\ j = 1, \dots, n-1 \quad (3)$$

With $f_0 = -\frac{\log 0.01 P_1}{t_1}$

Plotting the empirical forward rates against maturity, we get the following scatterplot:



<u>(3) Smooth the empirical forward rates using second order and third order polynomials.</u>

We then smooth the curve using either a quadratic or a cubic curve. To do so, we will first define a function to return the sum of squared errors (SSE) given the

```
sumSqMin1 <- function(par, data) {
  sum((frate-(par[1]+par[2]*data[,1]+par[3]*(data[,1])^2))^2)
}
```

empirical forward rate and a set of parameters that define the quadratic or cubic function. An example for the quadratic SSE is given:

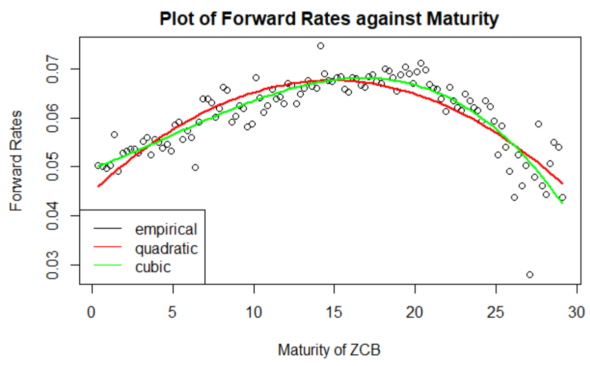
We then use the *optim()* function of R to find the parameters that will minimize the above function to obtain parameters of a quadratic or cubic function that minimizes the SSE. An example for the optimization for the quadratic is shown:

```
opt1 = optim(c(0,0,0),sumSqMin1,data=ZCBP,hessian=T,method="Nelder-Mead")
frquad = opt1$par[1]+opt1$par[2]*ZCBP$time+opt1$par[3]*(ZCBP$time)^2
lines(ZCBP$time,frquad, col="red",lwd=2)
```

Superimposing the smoothing functions on the same plot, we get the resulting plot:

2.3 Spot Rates

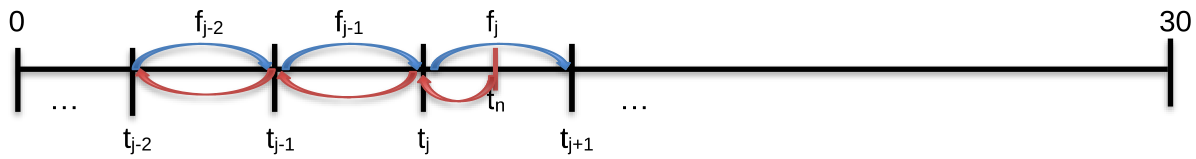
(4) Estimate the empirical spot rates for t in (t_1, t_n) using equation (4).



With the forward rate f(t), we can estimate the empirical spot rates using the equation:

$$r(t) = \frac{1}{t}\left[\sum_{i=1}^{j} f_{i-1}(t_i - t_{i-1}) + f_j(t - t_j)\right] \qquad (4)$$

Using a number line representation:

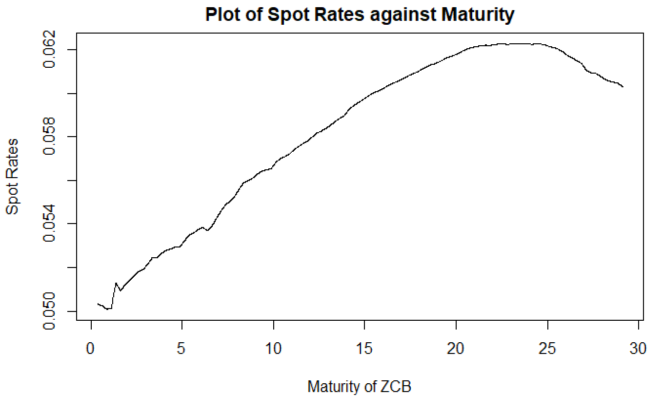


Spot rate at t_n will thus be 1 divided by the summation of the product of the blue arrows and red arrows up to t_j plus the product of the blue arrow and the difference between t_n and t_j

We will first create a grid of numbers using the function *seq()* within the minimum and maximum values of the time given in the data set as we will create the breaks (t_j) at the time in the data. The grid is set within the data as we only have the empirical forward rates for the respective t_j, we cannot extrapolate the values of f_{j-1} and f_{j+1} if the t_n lies outside the range of t_j.

To implement the formula in equation 4, we find the interval each number i, of the grid is in using *findInterval(grid[i],breaks)* where *breaks=c(0,ZCBP$time)*. After which, the formula can then be applied for each number in the grid as depicted by the figure above.

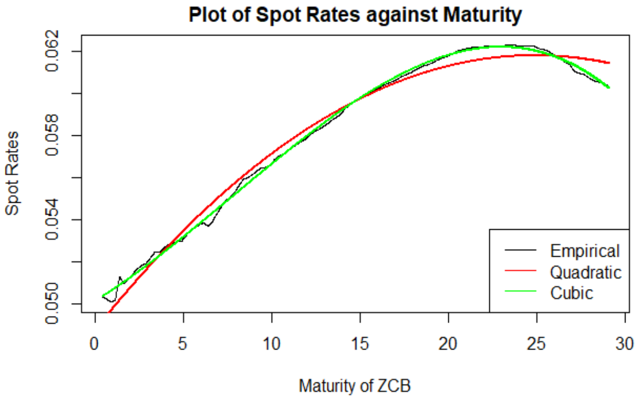
Plotting the empirical spot rates against maturity, we will finally obtain the yield curve as such:



(5) Smooth the empirical spot rates using second order and third order polynomials.

We follow a similar process as 2.2 (3) to smooth the curve using either a quadratic or a cubic function by first defining a function to return the SSE based on a set of parameters and empirical spot rates.

Superimposing the smoothing curves, the resultant plot is:



(6) Comment on your results.

We see that for both the empirical forward and empirical spot rate smoothing, the cubic curve fits the data better.

If we look at the value of the errors given by both optimizations:

```
> options(scipen = 999)
> opt3$value #quadratic
[1] 0.0006527684
> opt4$value #cubic
[1] 0.00007019978
```

The cubic curve gives a lower SSE. This is because as we increase the number of parameters to estimate, our estimated value will become more precise as each additional parameter can adjust the estimated value more, leading to lower bias.

However, it comes at a cost as the bias-variance trade-off suggests that having more parameters to estimate would cause greater variance. Hence, while increasing the number of parameters can lower the sum of squared error, it is important to understand this trade-off to select the optimal fitted model.

3. <u>Part 2: Fitting the Yield Curve</u>

In this section we will attempt to fit the yield curve into a model.

3.1 Dataset and Research Questions:

The data set we are given ("ZCBYF86.csv") contains the spot rate (in percent per annum) from 1986-02-01 to 2020-08-28 of US Treasury Zero Coupon bonds that have maturities of 1 to 30 years.

We will attempt to answer the following research questions:

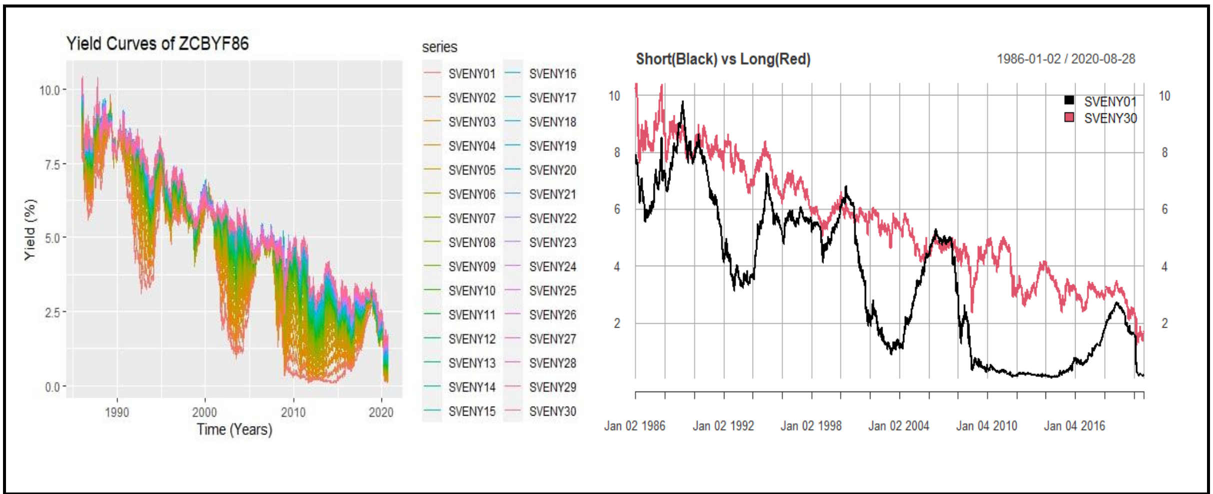
1. *Present the data in a suitable way.*
2. *Comparing the Nelson-Siegel and Nelson-Siegel-Svensson Models*
3. *Impact of Parameters of the Model*

3.2 Present the data in a suitable way.

The dataset has 3 dimensions available for plotting:

1. Spot Rates
2. Time
3. Maturity

Plotting Spot Rate against Time reveals the long-term trend of how daily spot rates change over time from 1986-02-01 to 2020-08-28 as shown below (left). However, as there are 30 different bonds of different maturities, it would be more intuitive to make a comparison between short term (1-year TTM) and long term (30-year TTM) maturity bonds, with the plot shown below (right).



We observe the following trends:

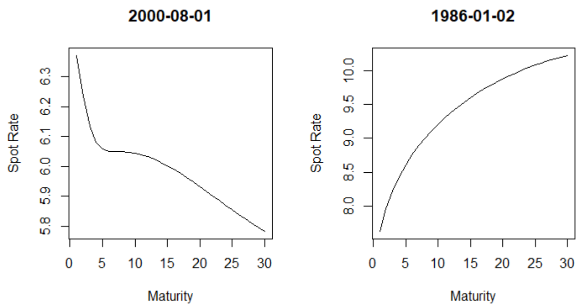
1. Shorter term bonds have generally lower spot rates than longer term bonds
2. Shorter term maturity bonds display more volatility in spot rates than longer term bonds
3. An overall downwards trend of spot rates over time, for both short- and long-term bonds.

Additionally, we observe how short- and long-term bond spot rates may act as a predictor of recessions. A normal yield curve is shown when short term debt instruments have a lower yield than longer term debt. Hence, its shape will be upwards sloping. This shape is typically associated with long-term growth. A yield curve that does not follow this is inverted.

An inverted Treasury yield curve is one of the most reliable leading indicators of an impending recession (McWhinney, 2020). Investors instead turn their money towards long term treasuries as a haven from falling equities. As a result, the increased demand for long term bonds causes their returns to fall below short-term rates. This causes the yield curve to invert (McWhinney, 2020). Inverted yield curves have preceded every recession since 1956. Looking at the provided data, we can see that this was true for the recessions caused by the bursting of the dotcom and subprime mortgage bubbles.

There are 3 prolonged periods where the spot rate of SVEN01 was greater than that of SVEN30. The 3 most recent periods of recession that coincide were from the savings & loans crisis (1990-1991), dot-com bubble collapse (2001-2001), & the subprime housing bubble (2007-2009). In these events, we can observe that the spot rate of SVEN01 was greater than that of SVEN30. Their corresponding instantaneous yield curves would be downwards sloping, as illustrated below.

For example, for the term structure on the day '2000-08-01' where that happened, the shape of the yield curve would indeed be downwards sloping as in the plot below. Whereas if the spot rate of SVEN01 is smaller than that of SVEN30, the yield curves would result in an upward sloping yield curve, as in the case of the yield curve on '1986-01-02'.



However, the shapes of the yield curves are difficult to explain and thus require the use of the Nelson-Siegel and Nelson-Siegel-Svensson Models that captures the shape of the yield curve with a set of parameters.

3.2 Comparing the Nelson-Siegel and Nelson-Siegel-Svensson Models

The Nelson-Siegel (NS) and Nelson-Siegel-Svensson (NSS) Models can be estimated using a non-linear regression model by minimizing the sum of squared errors given a set of spot rates *r(t)* at various maturities *t*

The NS model is given in equation 5:

$$r(t) = \theta_0 + \left(\theta_1 + \frac{\theta_2}{\theta_3}\right)\left(\frac{1-e^{-\theta_3 t}}{\theta_3 t}\right) - \frac{\theta_2}{\theta_3} e^{-\theta_3 t},\ for\ \theta_0, \theta_3 > 0 \quad (5)$$

And the NSS model adds an addition 2 parameters, given in equation 6:

$$r(t) = \theta_0 + \left(\theta_1 + \frac{\theta_2}{\theta_3}\right)\left(\frac{1-e^{-\theta_3 t}}{\theta_3 t}\right) - \frac{\theta_2}{\theta_3} e^{-\theta_3 t} + \frac{\theta_4}{\theta_5}\left(\frac{1-e^{-\theta_5 t}}{\theta_5 t}\right) for\ \theta_0, \theta_3, \theta_5 > 0 \quad (6)$$

To find a set of theta values that minimizes the sum of squared errors, we first write a function to return the sum of squared errors based on the equations 5 and 6. An example for the NS model is given below:

We will apply the transformation: $\theta_i = (\beta_i)^2$ to the parameters θ_0 and θ_3 so that the parameter θ_i, remains positive.

```
ns_sse <- function(par, maturity) {
  sum((yields-((par[1])^2 + (par[2] + par[3]/(par[4])^2)*(1-exp(-
(par[4])^2*maturity))/((par[4])^2*maturity) - (par[3]*exp(-
(par[4])^2*maturity)/(par[4])^2)))^2)
}
```

Hence in this example, the optimization will return the parameter β_0 and β_3 instead. To obtain θ_0 and θ_3, we then take the square of the parameter.
For each day of the dataset, we will then optimize the parameters over the 30 different bond maturities and store the parameters, sum of squared errors in a dataframe.

Since we have 2 models, we would like to find which model is a better fit for the data set that we have.

We will use the following methods to compare between the 2 models:

- Sum of Squared Errors (SSE)
- Akaike's Information Criteria (AIC) and Bayesian's Information Criteria (BIC)

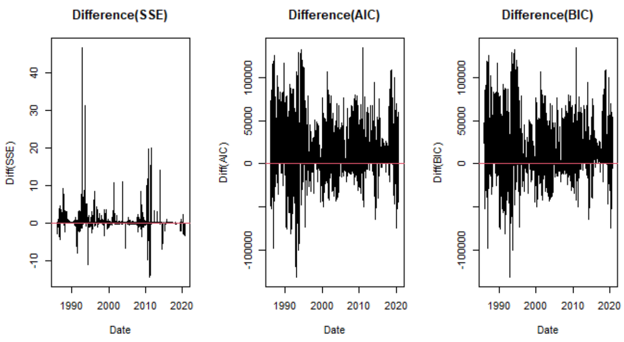
We find the difference between the SSE of the NS model and the NSS model by subtracting one from the other and plot it against time. If the difference were to be positive, then the NSS model gives a better fit as the SSE is smaller which is indeed the case.
To find the AIC and BIC, we first have to calculate them using their respective formulas:

$$AIC = n \log(deviance) + 2K;\ BIC = n \log(deviance) + K \log n$$

where n = number of observations and K = number of parameters

Similarly, a positive difference between the AIC and BIC of the 2 models indicate that the NSS would be better as the AIC and BIC would be smaller.

Plotting all 3 criteria together across time, we get the following plot:



Here we clearly see that the NSS model gives a better fit most of the time. This is again due to the fact that we have more parameters and hence can be more accurate in our predictions.

3.3 Impact of Parameters of the Model

Finally, we want to see how exactly each parameter of the model affects the shape of the yield curve.

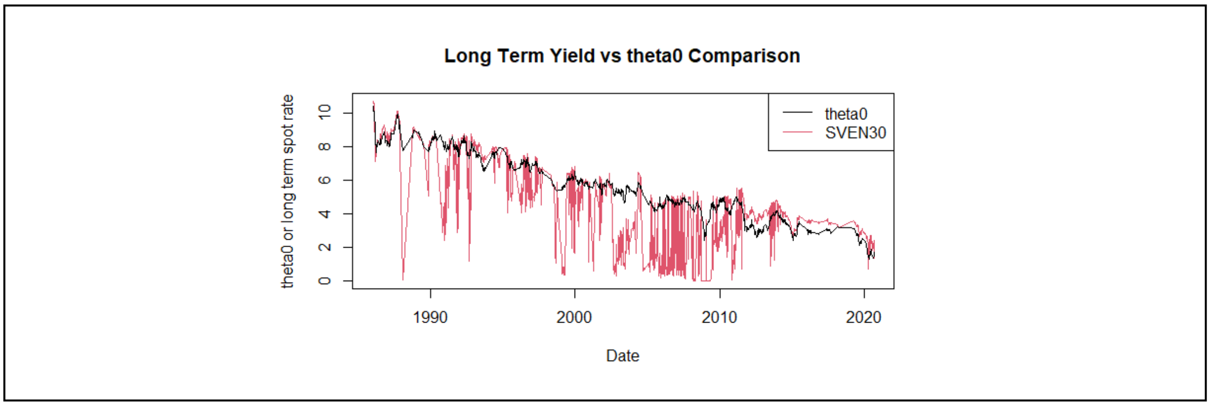
According to Gilli et al. (2010), each parameter of the model can affect an aspect of the shape:

- The first term θ_0 is independent of time and thus is seen as the long-term spot rate.
- The second term $(\theta_1)\left(\frac{1-e^{-\theta_3 t}}{\theta_3 t}\right)$ controlled by θ_1 will decay to zero as t increases and affects thus affects the slope of the yield curve in the short term
- The last term $\left(\frac{\theta_2}{\theta_3}\right)\left(\frac{1-e^{-\theta_3 t}}{\theta_3 t} - e^{-\theta_3 t}\right)$ controlled by $\frac{\theta_2}{\theta_3}$ will increase before decaying to zero, resulting in a hump, giving the yield curve a U-shape or a concave shape.

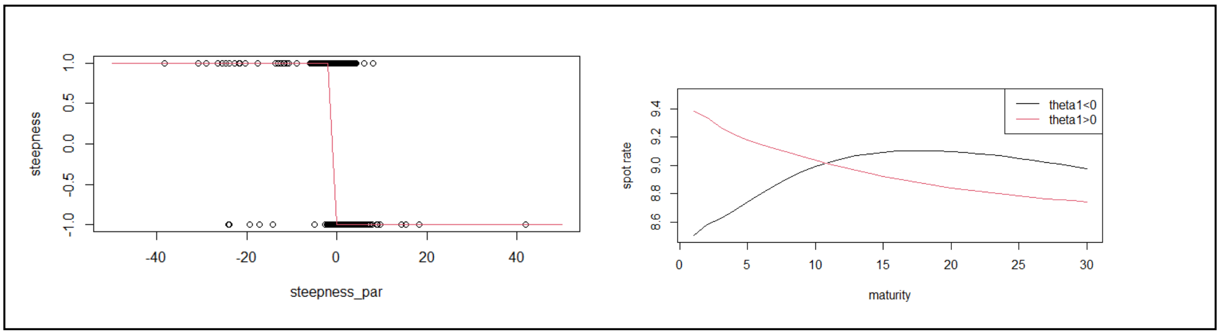
The additional parameters in the NSS model will account for a second hump in the yield curve that gives it an S-shape when a second hump is present.

However, Gilli et al. (2010) also notes that the humps in the yield curve are difficult to locate when the yield curve is not stable and creates a local maximum and local minimum. Hence, we will restrict our analysis for this portion to just the data points in which the NS model fits better where we will only expect 1 hump.

Plotting θ_0 over time, we see that the value of θ_0 follows a similar downwards trend of the long-term bond spot rates over time. However, the θ_0 is a lot more volatile than the long-term bond spot rates.



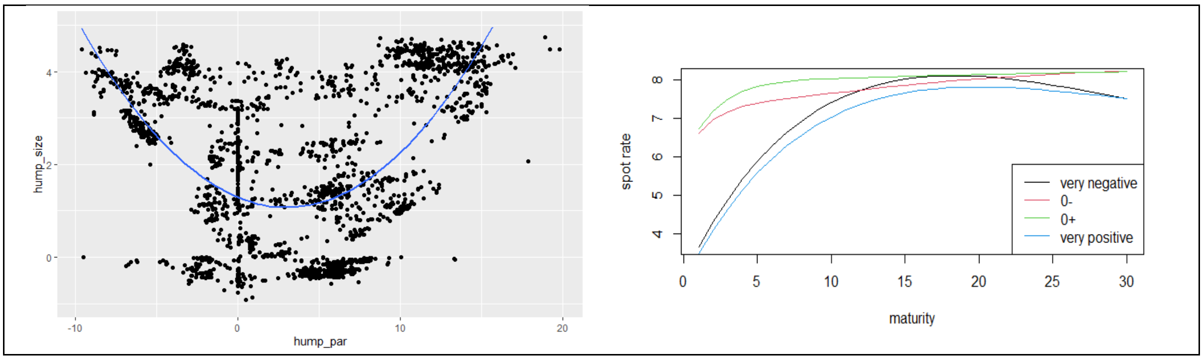
Next, we test if the sign of θ_1 affects the slope of the yield curve in the short term by plotting the sign of SVEN02-SVEN01 against θ_1. We observe the pattern as described in the literature where positive θ_1 causes the yield curve to slope downwards.



We can verify this by plotting the yield curves of various θ_1 values and indeed we see that it is the case.

Lastly, the size of the hump is measured by subtracting the short term yield from the maximum point of the yield curve if the yield curve has a concave shape and subtracting the minimum point of the yield curve from the short term yield if the yield curve has an inverted shape.

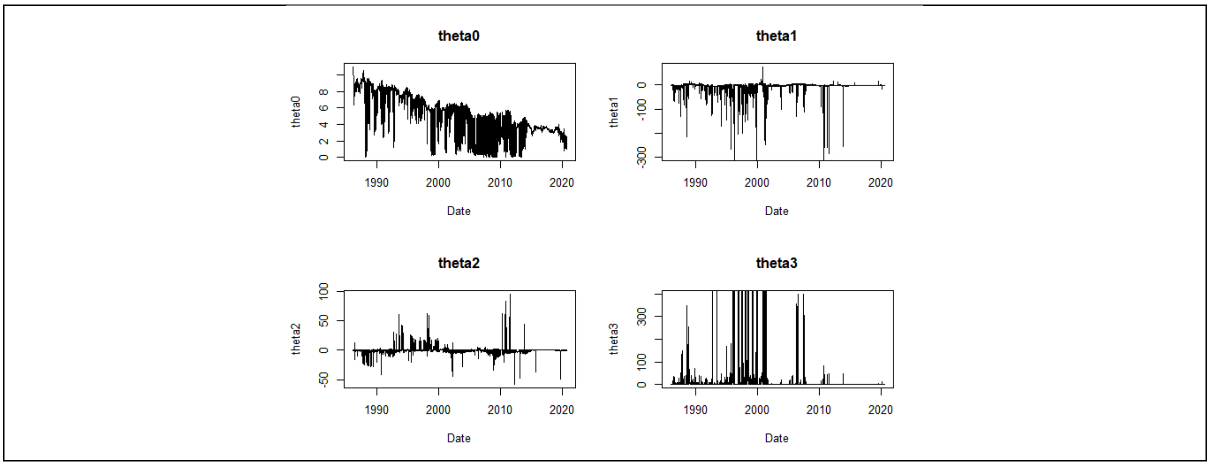
Plotting the size of the hump against $\frac{\theta_2}{\theta_3}$ we observe a weak quadratic relationship that makes sense that as this parameter approaches zero there would be no hump and the size of the hump increases as the parameter becomes extremely positive or negative.



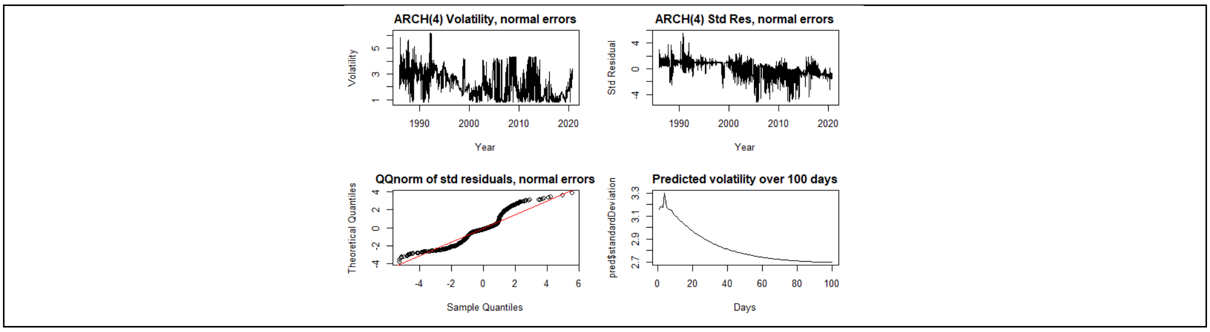
We can verify this by plotting the yield curves of various $\frac{\theta_2}{\theta_3}$ values and indeed we see it to be the case.

4. Conclusion

Our analysis of the parameters uses only the parameters of the NS model even though we have shown that NSS model parameters are better at fitting the yield curves. At the same time, our analysis of 1 parameter may not hold if the other parameters are very volatile as can be seen in when we plot the values of θ over time.



One possible further study we could do in future would be to identify a period where the parameters are quite stable to create a predictive model for forecasting future yields or modelling volatility of the parameters for that purpose using ARCH, GARCH and APARCH models. For example, using the ARCH (4) model to θ_0, we can predict the volatility of θ_0 as shown below:



5. References

McWhinney, J. (2020, August 28). The impact of an inverted yield curve. Retrieved from https://www.investopedia.com/articles/basics/06/invertedyieldcurve.asp

Gilli, et al. (2010, March 30) Calibrating the Nelson–Siegel–Svensson Model. Retrieved from https://comisef.eu/files/wps031.pdf