

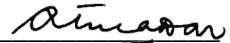
Modelling the Dynamic Yield Curve for the Canadian Bond Market

by
Sayan Maity

A thesis submitted to
Saint Mary's University, Halifax, Nova Scotia
in Partial Fulfillment of the requirements for
the Degree of Masters in Applied Economics

August 2020, Halifax, Nova Scotia

© [Sayan Maity, 2020]

Approved By : 
Dr. Atul Dar
Supervisor

Date: 13/08/02020

Acknowledgement

I would like to extend my gratitude towards my thesis supervisor Dr. Atul Dar. for helping me to carry out this study. His constant guidance and insights shaped this project to where it is. I would also like to thank our program director Dr. Ather Akbari for his wholehearted support and cooperation on various aspects of my project and my academic life.

I would like to give a special thanks to Dr. Yigit Aydede for introducing me to R programming language and machine learning, which helped me widen my horizon of economic analysis.

I would like to thank Armaan Hossain, Vaishnavi M. Godse and Rachel Adekemi Siyanpeter, for giving me their helping hand to complete my dissertation and help me in every aspect of my life in this new country.

I would also like to thank all of my classmates for being the most supportive people I could expect.

Moreover, I like to thank Baoliang Chan for being such a great roommate and letting me work in any situation.

I end by thanking my parents for always standing by me and my brothers and sister for inspiring me.

Contents

Abstract	4
1. Introduction	5
2. Theory	7
3. Data	16
4. Estimating Models	20
4.1 In Sample Modelling	20
4.1.1 Two-Step Modelling	20
4.1.2 One Step Modelling	23
4.2 Out of Sample Forecasting	25
5. Results	27
5.1 Two-Step Model	27
5.2 One Step Modelling	36
5.3 Forecast Performance	38
5.4 Structural Break: Regime Switching Model	41
6. Conclusion	43
References	43
Appendix 1: Limit of the yield curve	45
Appendix 2: Kalman Filter for the One-step DNS model	46

Modelling Dynamic Yield Curve for Canadian Bond Market

By Sayan Maity

Abstract

The yield curve or the term structure of the interest has been one of the key leading macroeconomic indicators and forecasting the yield curve could provide vital information about future macroeconomic performance. Following the seminal works of Diebold & Li (2006a) and Diebold et. Al. (2006b) Many papers tried to evolve different techniques that can model the inner dynamics of the yield curve and forecast for future periods efficiently. This paper also followed the same path of Diebold & Li (2006) to model the inner dynamics of the Canadian zero-coupon yield curve and added some new structure following different stylized facts obtained from the factors. Primarily this paper estimated the model in a two-step way with the factor dynamics being VAR, VECM, EGARCH and DCC-EGARCH along with the one-step model using Kalman filter. A grid search is performed to calibrate the yield curve factors and using that all the models are compared using the root mean square forecast error. VECM and EGARCH turned out to be the best models revealing the different short term and long-term dynamics and the parsimonious nature of the model. Furthermore, a regime-switching model is also estimated to find the changing volatility structure of the yield curve as indicated in the previous models.

Keywords: term structure of the interest, dynamic Nelson Siegel, short-term and long-term dynamics

[JEL Classification : C4, C6, E43, E44, E47, G23]

Submitted on 13th August 2020

1. Introduction

Monetary authorities around the world rely on research-based evidence to inform policymaking. To do so effectively, they need to have access to reliable statistical indicators of financial markets, including those that reflect expectations of investors and how those expectations respond to policy. Different monetary institutions around the world rely on different indicators from different market as a leading indicator to predict future movements in the economy. One such leading indicators are the zero-coupon yield curve. A zero-coupon yield curve is a specific type of yield curve that plots interest rates/ yield to maturity on zero-coupon bonds to different maturities. These curves enable to price arbitrary cash flows, fixed-income instruments, and derivatives¹.

The use of the yield curve or specifically, the slope of the yield curve as a leading indicator is not new. Many economists argue that the yield curve could be a better measure for predicting future inflation, future economic growth and recession (Bernard & Gerlach, 1998) than other leading economic indicators like the unemployment rate, consumer confidence and GDP. Wright (2006) argued that the term spread measures the difference between real short term and expected short term interest rate. Thus, the higher the term spread, the greater the likelihood of a recession in the upcoming period.

The link between the term structure of interest rate and the zero-coupon yield curve is very evident. The yield to maturity is just the constant interest rate at which discounted cash flow of a bond equals the price. This serves as the direct link between the zero-coupon

¹ (*Actuarialtoolkit.soa.org*, actuarialtoolkit.soa.org/tool/glossary/zero-coupon-yield-curve.)

yield and the interest rate as the term structure of interest can be modelled by the zero-coupon yield curve.

It is well recognized that there is a feedback relationship between the macroeconomic and financial factors. However, this feedback relation is hard to model by both macro and financial economist due to their different outlook. This is very evident in the case yield curve and term structure modelling. Macroeconomists, financial economists, and market participants all have attempted to build useful models of the yield curve. But the resulting models are very different in form and fit, which reflects the modelling demands of researchers from various fields and their motives for modelling the yield curve. The yield curve models developed by macroeconomists focuses on the role of expectations of inflation and future real economic activity in the determination of the yields. On the other hand, a financial economist focuses on specific determinants that arise from the model itself that could determine the movement in the yield curve. Though this paper will not explicitly study the effect macroeconomics factor, there is certain usage of macroeconomics concept of a long and short run.

In this paper, we build a parsimonious model of the yield curve that can capture both the inner dynamics of the yield curve and can forecast the yield efficiently, without explicitly modelling the feedback between financial and macroeconomic factors. The basis of this model comes from Diebold & Li (2006a) and Diebold et. Al. (2006b). These two papers suggested a three-factor dynamic factor model following the famous Nelson and Siegel (1984) model of the yield curve. The main underlying assumption was the yield curve has some temporal dimension in the way it propagates through time. Thus, modelling the yield curve dynamically could reveal different aspects of the term structure which is not evident from the static model.

In this paper, the yield curve dynamics is modelled in different ways. Using the monthly data set from 1986 to 2020, a series of different approaches are adopted to estimate the perfect factor dynamics that are present in the term structure. These models will be then compared against each other based on their forecast performance. The model that has better forecast performance should have fit the model better and can also explain different aspects of factor dynamics both in the long and short run.

In the next section, the evolution of the literature from the static Nelson and Siegel (1984) model to dynamic Diebold & Li (2006a) and Diebold et. Al. (2006b) models will be discussed (Section 2). Along with that, the factor structure of the yield curve and how that shapes the yield curve will also be discussed. In subsequent sections, data issues (Section 3), the estimating models along with the prediction strategy (Section 4), and the results (Section 5) are presented. The paper concludes by drawing implications and points to directions for future research (Section 6).

2. Theory

To provide the background to the model, we first discuss a few key concepts of the bond market: the discount curve, the forward rate curve, and the yield curve.

Let us assume, $P(\tau)$ denote the price of τ -period discount bond or the present value of \$1 receivable after τ period. The rate of return or yield for the bond for the period τ is denoted by $y(\tau)$. Now, if we assume that the yield is compounding continuously then we can write

$$P(\tau) = e^{-\tau y(\tau)} \quad (2.1)$$

The yield or the rate of return can be replaced by the market interest rate for any financial asset. So, modelling of bond yield can also be somewhat related to the modelling of the

interest rate. Thus, like interest rate, we can define a forward rate $f(\tau)$, which can be defined as

$$f(\tau) = -\frac{P'(\tau)}{P(\tau)} \quad (2.2)$$

So, the knowledge of the discount curve helps to find both spot and forward rate. From these two equations, a relationship between spot and forward rate can be defined

$$y(\tau) = \frac{1}{\tau} \int_0^\tau f(u) du \quad (2.3)$$

in other words, spot rates are the equally weighted average of forward rates.

In theory, one can work with anyone of these factors to model the nature of bond yield or the term structure of interest rate. But in reality, none of these factors is directly observable, and they are generally derived from the observed bond price. To estimate this, different mathematical models have been established.

The first approach to model the yield curve using the discount curve was performed by McCulloch (1971). He used polynomial splines to model the discount curve as

$$y(m) = a_0 + \sum_{j=1}^k a_j f_j(m) \quad (2.4)$$

In essence, the discount rate is presented as a weighted sum of forward rates. He defined a different polynomial structure to define the forward rate. In the latter paper, McCulloch (1975), he adjusted his estimations for taxes. One of the key problems of this model is that due to the polynomial structure, the fitted discount curve diverges in the long run.

This is solved by the usage of exponential splines, which is proposed by Vasicek & Fong (1982). Their model is closely related to the initial assumption, as represented in (2.1) above.

$$P(\tau) = e^{-\tau y(\tau)} \quad (2.5)$$

One of the critical improvements of this model is the usage of the negative transformation of maturity τ . For this the model yields –

$$\lim_{\tau \rightarrow \infty} f(\tau) = \alpha \quad (2.6)$$

where, α is a fixed value. This means the forward rate and simultaneously, the zero-coupon rate both converge to a fixed value as maturity increases. This helped to model the yield curve in the long end accurately. But one of the overall problems of this discount rate modelling is that it does not ensure a positive forward rate, a problem that can be solved by not modelling the discount curve but modelling the forward rate itself, as shown by Fama & Bliss (1987). Using this we can get the famous Fama Regression which is the multiperiod forecast of the short rate

$$P_{t+\tau}(n) - y_t(\tau) = \alpha + \beta(f_t(n) - y_t(\tau)) + \varepsilon_{t+\tau}(n) \quad (2.7)$$

All the values in the above regression are in logarithm, and so, this excess return of the long term over short term bonds can be defined as a function of forward rate. The forward rate constructed using this assumption of time variability of the excess return of long-term bonds is called "unsmoothed Fama Bliss" forward rates, which are further transferred into unsmoothed Fama Bliss yields through weighted averaging. This unsmoothed Fama Bliss yields accurately price the included bonds and are often used as raw yields to fit any model of the yield curve. Unsmoothed Fama-Bliss yields are one of the significant innovations in yield curve modelling and are still used by various central banks to estimate the term structure. Based on Fama Bliss yields, many researchers fit different empirical yield curves. Such fitting smooths the unsmoothed Fama Bliss yields and gives us a smoothed yield curve.

Given the ever-changing scenario of the market, the yield curve moves a lot, and so it can have many different shapes. Yields can be increased at an increasing or decreasing rate, decreasing at an increasing or decreasing rate, flat, U-Shaped, S-shaped, humped and many more. To fit all those shapes, one needs a very simple and parsimonious model which is flexible enough to incorporate all those shapes.

Nelson-Siegel class of models which comes from Nelson and Siegel (1987), are the prime examples of such parsimonious model. According to Nelson and Siegel (1987), the key to parsimonious modelling lies in the solution of difference or differential equation of spot rates, which is forward rate. For example, if we have instantaneous forward rate $f(\tau)$ for maturity τ which can be described as a solution to a second-order differential equation with real and unequal roots, $y(\tau)$ can be described as:

$$f(\tau) = \beta_0 + \beta_1 e^{-\tau\lambda_1} + \beta_2 e^{-\tau\lambda_2} \quad (2.8)$$

Where λ_1 & λ_2 are constants and β_0, β_1 & β_2 are derived from the initial condition. The spot rate $y(\tau)$ will then be the weighted average of forward rates as it is discussed earlier.

$$y(\tau) = \frac{1}{\tau} \int_0^\tau f(u) du \quad (2.9)$$

Here, the key to achieving the different shape of the yield curve are values of the parameters β_0, β_1 and β_2 . But Nelson and Siegel (1987) found that given various values of constants like λ_1 and λ_2 the model is overparameterized. So, they come up with a more parsimonious model with just one λ .

$$f(\tau) = \beta_0 + \beta_1 e^{-\tau\lambda} + \beta_2 \tau \lambda e^{-\tau\lambda} \quad (2.10)$$

Solving this, the famous Nelson-Siegel yield curve can be shown to be:

$$y(\tau) = \beta_0 + \beta_1 \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} \right) + \beta_2 \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right) \quad (2.11)$$

Certain properties² of this functional form make it very useful for yield curve estimation.

Specifically,

$$\lim_{\tau \rightarrow 0} y(\tau) = f(0) = r \quad (2.12)$$

$$\lim_{\tau \rightarrow \infty} y(\tau) = \beta_0 \quad (2.13)$$

This comes along with the parsimony and the flexibility of the curve given by its factor structure. Thus, for a long time Nelson–Siegel yield curves have been very popular among policymakers to estimate the yield curve. This idea of parsimonious modelling is followed by several other researchers and has led, through various extensions or restrictions, to different, and more flexible, variations of the same yield curve. Litterman and Scheinkman (1991) models yield curve by only considering the first three components of principal component analysis. Svensson (1994) introduces additional parameters that allow the yield curve to have an additional hump. These two models are very popular among the central banks to construct the zero-coupon yield curve. Apart from that Bjork and Christensen (1999) proposed a similar model to Svensson by introducing a second slope factor to the three factors NS model. On the other hand, Bliss (1997) proposed an alternative to make NS model more flexible by relaxing the restriction that the slope and curvature should be governed by the same decay component and allowing for two decay components which were present in the initial idea of Nelson & Siegel (1984).

One of the things can be argued that the yield curve evolves dynamically, which means that it should contain both a cross-sectional and a temporal dimension. Although these Nelson–Siegel class models capture the individual yield curves, they don't explain how they are linked can evolve over time. But with the presence of factor structure, a time dimension

² For proof check Appendix 1

can be introduced to explain how these different factors are interacting over the period resulting in the changes in the yield curve.

Many studies introduce this time dimension in different ways. Thus, Diebold, Piazzesi, and Rudebusch (2005) proposed a two-factor dynamic model based on Litterman and Scheinkman (1991) variation of the Nelson-Siegel model. They argued that since the first two principal components explain nearly all variation in interest rates, a two-factor model may suffice to forecast the term structure. However, two factors will not be enough to fit the entire yield curve. The first three-factor dynamic variation of the Nelson-Siegel model was proposed in Diebold & Li (2002). This model is based on the original Nelson-Siegel model and has substantial flexibility to match the changing shapes of the yield curve.

The models used in this paper come from the seminal papers by Diebold & Li (2006) and Diebold, Rudebusch and Aruoba, (2006b), and will henceforth be referred to as the Dynamic Nelson-Siegel (DNS) model. Much of the subsequent models in the literature are different variations of the model proposed in these papers.

The basic model of both Diebold & Li (2006a) and Diebold et al. (2006b) is just the original Nelson Siegel Model with coefficients having an added time dimension. So, the extended forward rate curve is

$$f_t(\tau) = \beta_{0t} + \beta_{1t}e^{-\tau\lambda_t} + \beta_{2t}\tau\lambda_t e^{-\tau\lambda_t} \quad (2.14)$$

And the corresponding yield curve is :

$$y_t(\tau) = \beta_{0t} + \beta_{1t} \left(\frac{1-e^{-\lambda_t\tau}}{\lambda_t\tau} \right) + \beta_{2t} \left(\frac{1-e^{-\lambda_t\tau}}{\lambda_t\tau} - e^{-\lambda_t\tau} \right) \quad (2.15)$$

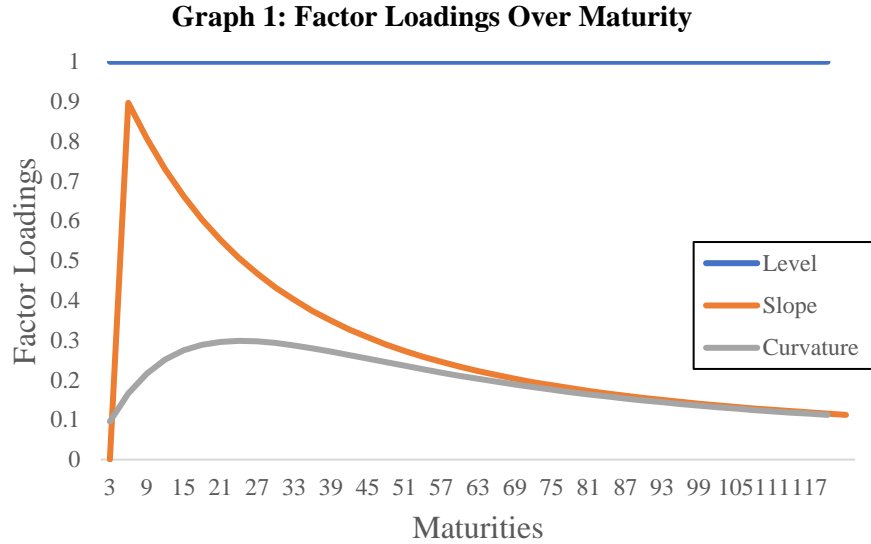
This equation is basically the original NS model with dynamics added to the parameter. Before turning to the DNS model, we explore the factor structure of the NS and DNS class of models.

These parameters with their temporal dimension can be seen as dynamic or latent factors which can explain the dynamics of yield at any given maturity. These factors remain unobserved in the yield curve and so they are called latent factors. Here the high dimension observed factors like many yields across maturities are driven by only three latent factors. This directly follows from the basic assumption of the latent factor model where a high dimensional set of the variable can be explained by a lower dimension set of parameters. This idea was also hinted in Litterman and Scheinkman (1991) where the yield curve was modelled by only considering the first three components of principal component analysis as they intuitively explain the level, slope, and curvature factor of the yield curve.

The coefficients of these factors or the factor loadings determine the cross-section of yield. The parameter λ_t , which explain the exponential decay rate of the yield curve, determine the factor loadings of the yield curve. These factor loadings are nothing but the explanatory variables $1, \left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right), \left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right)$ with parameters $\beta_{0t}, \beta_{1t}, \beta_{2t}$. Though we are considering only about the dynamics of these factors, we must analyze the behaviour of the factor loadings to understand these factors and what determines the overall shape of the yield curve.

First look at the factor loading of β_{1t} i.e. $\frac{1-e^{-\lambda\tau}}{\lambda\tau}$. It begins at one but monotonically decays down to zero. This is often called "short term factor" and it mostly affects the short-term yields. Next, we have the factor loading of β_{2t} i.e. $\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}$. It begins at zero, increases, and then again decay down to zero. This is called "medium-term factor" and it affects the medium-term yields. Lastly, we have β_{0t} which is constant at one and unlike the other two factors, does not decay down to zero. It is called "long term factor" as it affects

the long yields. If we fixed the value λ_t to a specific value for a cross-section, these features are visible from the graph below.



Also, these factors determine the shape of the yield curve. Such as β_{0t} determine the level of the yield curve: a shift in β_{0t} shifts the yield curve parallelly. Similarly, β_{1t} determines the slope of the yield curve: an increase in β_{1t} increase the short-run yield but the long-run yield remains unchanged. Finally, β_{2t} determines the curvature of the yield curve: an increase in β_{2t} increase medium yield but not the short and long-run yields.

To emphasize the "level, slope & curvature" factor DNS model can be rewritten as

$$y_t(\tau) = l_t + s_t \left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} \right) + c_t \left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right) \quad (2.16)$$

where $t = 1, \dots, T$; $\tau = 1, \dots, N$. Diebold (2006b) provides a state-space representation of the model to work with.

$$y_t = \Lambda f_t + \varepsilon_t \quad (2.17)$$

Where the variables are

$$y_t = \begin{pmatrix} y_t(\tau_1) \\ y_t(\tau_2) \\ \vdots \\ y_t(\tau_N) \end{pmatrix}, \quad f_t = \begin{pmatrix} l_t \\ s_t \\ c_t \end{pmatrix}, \quad \varepsilon_t = \begin{pmatrix} \varepsilon_t(\tau_1) \\ \varepsilon_t(\tau_2) \\ \vdots \\ \varepsilon_t(\tau_N) \end{pmatrix}$$

& the parameter matrix Λ

$$\Lambda = \begin{pmatrix} 1 & \frac{1 - e^{-\lambda\tau_1}}{\lambda\tau_1} & \frac{1 - e^{-\lambda\tau_1}}{\lambda\tau_1} - e^{-\lambda\tau_1} \\ 1 & \frac{1 - e^{-\lambda\tau_2}}{\lambda\tau_2} & \frac{1 - e^{-\lambda\tau_2}}{\lambda\tau_2} - e^{-\lambda\tau_2} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1 - e^{-\lambda\tau_N}}{\lambda\tau_N} & \frac{1 - e^{-\lambda\tau_N}}{\lambda\tau_N} - e^{-\lambda\tau_N} \end{pmatrix}$$

where, $t = 1, \dots, T$. The stochastic errors $\varepsilon_t(\tau)$ are idiosyncratic or depends on the maturity τ . Hence each yield is driven by both common and idiosyncratic factors.

Next, the transition equation is used to describe common factor dynamics. The following VAR (1) model is used for this purpose:

$$(f_t - \mu) = A(f_{t-1} - \mu) + \eta_t \quad (2.18)$$

where the variables are

$$f_t = \begin{pmatrix} l_t \\ s_t \\ c_t \end{pmatrix}, \quad \eta_t = \begin{pmatrix} \eta_t^l \\ \eta_t^s \\ \eta_t^c \end{pmatrix}$$

and the parameter vector & matrices are

$$\mu_t = \begin{pmatrix} \mu^l \\ \mu^s \\ \mu^c \end{pmatrix}, \quad A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

Here μ is the factor mean and A governs factor dynamics.

The standard assumptions are made in this connection are that the white noise transition and factor disturbances are orthogonal to each other and the initial state:

$$\begin{pmatrix} \eta_t \\ \varepsilon_t \end{pmatrix} \sim WN \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} Q & 0 \\ 0 & H \end{pmatrix} \right)$$

and the initials states are:

$$E(f_0 \eta_t) = 0$$

$$E(f_0 \varepsilon_t) = 0$$

Moreover, the Diebold-Li model is formulated such that the state equation factor disturbances η_t are correlated, and therefore the corresponding covariance matrix Q is non-diagonal. However, the model imposes diagonality on the covariance matrix H of the observation equation disturbances ε_t such that deviations of observed yields at various maturities are uncorrelated.

The dynamic factor Nelson Siegel model has much flexibility, and many researchers have tried to improve upon this using different factor dynamics or different estimation techniques altogether. Hautsch & Yang (2012) used a stochastic volatility Bayesian inference model to compute the factor dynamics. Different factor dynamics like VECM and GARCH was introduced in Tsui & Wu (2013). Also, the regime-switching model is applied by Xiang & Zhu (2013) which is a very frequently observed phenomenon in macroeconomic parameters. In this study, many of these techniques are incorporated to achieve a parsimonious model fit and a good forecast performance.

3. Data

The Bank of Canada zero-coupon yield curve data will be used for all these models. The zero-coupon yields are constructed using the methodology of Bolder, Johnson & Metzler (2004). In that paper, they sought to assess which model would be preferable for the estimation of the zero-coupon yield of the Canadian bond market. Bolder and Gusba (2002)

compared several estimation algorithms using Canadian government bond data. All these algorithms can be classified into three broad categories: spline-based and functional-based. Bolder and Gusba (2002) concluded that the Merrill Lynch exponential spline (MLES) model, as described by Li et al. (2001), is the most desirable term-structure estimate model when evaluated against the criteria of goodness of fit, the composition of pricing errors, and computational efficiency. Thus, this MLES method is used to build this database.

The data for government bond yields is monthly, running from 1986 January to 2020 June, and is obtained from the Bank of Canada. Also, all the daily yields are calculated for the maturity of 300 months or 25 years and sometimes up to 30 years. Working with this data directly has some advantages and disadvantages. First, using the daily yield with all 300 months of maturities could provide more data point, with increasing variability the noise in the data will be pretty high and the general trend of the factors will remain unseen. This could be a huge challenge. Also, if all 300 months of maturity are used, this large volume of data raises significant difficulties in devising feasible regression strategies. To get rid of any confusion the structure of Diebold et. al. (2006) is followed and only 120 months of maturities are taken. Also, mainly for the ease of operation and feasibility of regression all the daily data are converted into a monthly average.

At the other end of the maturity spectrum, any maturity under three months is not considered for estimation. the volatility of yields might be significantly higher for bonds with a residual maturity of only a few months. It has been claimed that these bonds do not display normal behaviour vis-à-vis market interest rates and would be a source of bias in yield curve estimation. Thus, they should be excluded from the analysis.

Before we turn to estimation, we provide some summary statistics of this data in Table 1 below for all given maturities, i.e. from three months to 120 months.

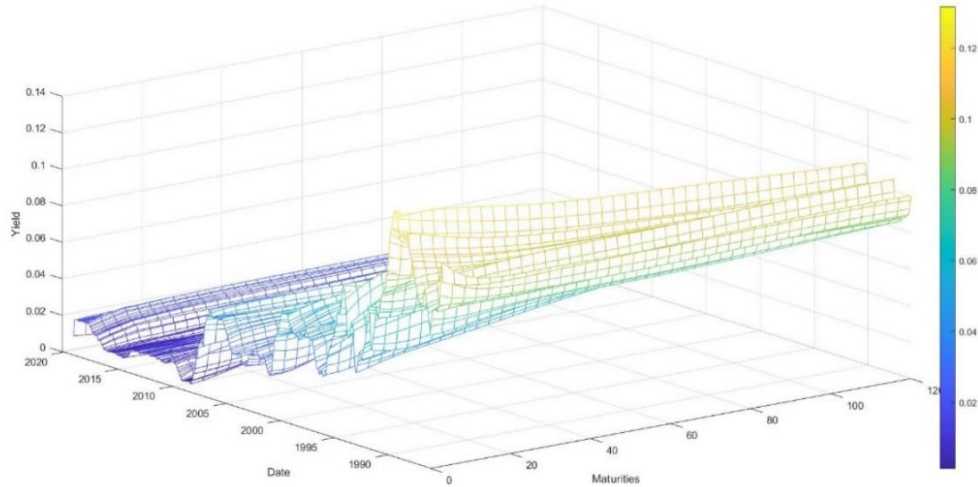
Table 1: Summary Statistics of Yield Curve

Maturities	Mean	SD	Min	Max	ACF (1)	PACF (1)
3	0.0412	0.0334	0.00136	0.132	0.9913	0.9937
6	0.0417	0.0327	0.00210	0.129	0.9913	0.9941
9	0.0422	0.0323	0.00333	0.129	0.9912	0.9942
12	0.0427	0.0319	0.00413	0.129	0.991	0.9942
15	0.0432	0.0316	0.00408	0.128	0.991	0.9942
18	0.0437	0.0313	0.00390	0.128	0.9909	0.9942
21	0.0441	0.0311	0.00379	0.127	0.9908	0.9942
24	0.0446	0.0309	0.00374	0.126	0.9907	0.9942
27	0.0450	0.0307	0.00376	0.125	0.9906	0.9943
30	0.0454	0.0306	0.00386	0.124	0.9906	0.9944
33	0.0458	0.0304	0.00401	0.123	0.9906	0.9945
36	0.0462	0.0303	0.00422	0.122	0.9906	0.9946
39	0.0466	0.0302	0.00445	0.121	0.9906	0.9947
42	0.0469	0.0301	0.00462	0.121	0.9906	0.9948
45	0.0473	0.0300	0.00482	0.120	0.9906	0.9949
48	0.0476	0.0299	0.00504	0.119	0.9906	0.995
51	0.0480	0.0298	0.00528	0.119	0.9906	0.9951
54	0.0483	0.0297	0.00553	0.119	0.9906	0.9952
57	0.0486	0.0296	0.00580	0.118	0.9906	0.9953
60	0.0489	0.0295	0.00609	0.118	0.9906	0.9953
63	0.0492	0.0294	0.00638	0.117	0.9906	0.9954
66	0.0495	0.0294	0.00668	0.117	0.9906	0.9955
69	0.0498	0.0293	0.00685	0.117	0.9906	0.9956
72	0.0500	0.0292	0.00703	0.116	0.9906	0.9956
75	0.0503	0.0291	0.00722	0.116	0.9905	0.9957
78	0.0505	0.0291	0.00742	0.116	0.9905	0.9957
81	0.0508	0.0290	0.00763	0.115	0.9905	0.9958
84	0.0510	0.0289	0.00786	0.115	0.9905	0.9959
87	0.0512	0.0288	0.00800	0.114	0.9905	0.996
90	0.0514	0.0288	0.00806	0.114	0.9905	0.996
93	0.0517	0.0287	0.00813	0.114	0.9905	0.9961
96	0.0519	0.0286	0.00820	0.113	0.9905	0.9961
99	0.0521	0.0286	0.00828	0.113	0.9905	0.9962
102	0.0522	0.0285	0.00835	0.112	0.9905	0.9963
105	0.0524	0.0284	0.00843	0.112	0.9905	0.9963
108	0.0526	0.0284	0.00851	0.112	0.9905	0.9964
111	0.0528	0.0283	0.00859	0.111	0.9905	0.9965
114	0.0530	0.0282	0.00868	0.111	0.9906	0.9966
117	0.0531	0.0282	0.00877	0.111	0.9906	0.9966
120	0.0533	0.0281	0.00885	0.110	0.9906	0.9967

In this table, the main summary statistics like mean, standard deviation maximum and the minimum value are reported. Overall, we can see the mean yield is increasing with maturity. Also, the yield is less volatile as maturity increases, as we can from decreasing standard deviation. A similar story can be told from maximum and minimum values, as the gaps between the two narrows as maturity rises. Finally, autocorrelation and partial autocorrelation function for one lag for all the maturities are also reported. We can see a decreasing autocorrelation and increasing partial autocorrelation for the maturities.

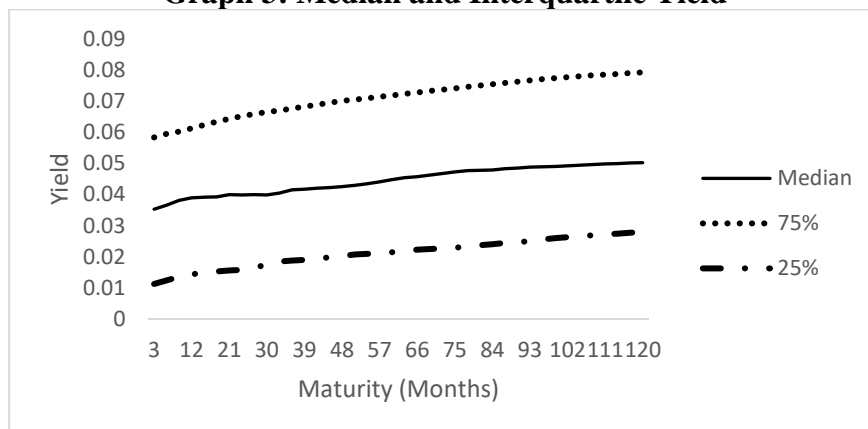
Furthermore, all these yields are represented using a mesh plot of yield which is presented below.

Graph 2: Mesh Plot of Yields



A very interesting feature can be observed from this graph. During the 90's we can see a high expectation for the short-term yield from the presence of an inverted yield curve. This is a clear representation of the high volatility in the yield in the early decade, although this volatility decreases thereafter. Though the direct increase in the yield is not visible here, this can be seen in the median and interquartile plot of the yield for every maturity in the following chart. This indicates a still almost flat but positive hump increasing yield curve

Graph 3: Median and Interquartile Yield



4. Estimating Models

The main two purposes of the modelling are in-sample modelling and out of sample forecast.

4.1 In Sample Modelling

There are two main different ways that this model can be fitted to the yield curve. these models are broadly classified as one-step and two-step models.

4.1.1 Two-Step Modelling

The main idea of the fitting yield curve of two-step modelling is first, fit the individual cross-sectional daily NS yield curve using OLS. This gives a 3-dimensional time series of estimated coefficients $\{\hat{l}_t, \hat{s}_t, \hat{c}_t\}_{t=1}^T$ and corresponding n-dimensional series of residual pricing errors. $\{\hat{\epsilon}_t(\tau_1), \hat{\epsilon}_t(\tau_2), \dots, \hat{\epsilon}_t(\tau_N)\}_{t=1}^T$. Thus, according to the standard state-space model, an N-dimensional time series of yields boils down to a 3-dimensional time series of yield factors.

The next step is to build dynamics from the estimated coefficients. For this Diebold & Li (2006) used both univariate AR(1) and Vector Auto-regression (1) specification, where in the equation

$$\hat{y}_{t+1}(\tau) = \hat{\beta}_{0,t+1} + \hat{\beta}_{1,t+1} \left(\frac{1-e^{-\lambda_t \tau}}{\lambda_t \tau} \right) + \hat{\beta}_{2,t+1} \left(\frac{1-e^{-\lambda_t \tau}}{\lambda_t \tau} - e^{-\lambda_t \tau} \right) \quad (4.1.1)$$

the individual factors are of the following form in the AR(1) and VAR(1) specifications respectively:

$$\hat{\beta}_{i,t+1} = \hat{c}_i + \hat{\gamma}_i \hat{\beta}_{it} \quad (4.1.2)$$

$$\hat{\beta}_{t+1} = \hat{c} + \hat{\Gamma} \hat{\beta}_t \quad (4.1.3)$$

Both specifications are used in the analysis, although, VAR tends to produce poor forecasts due to the potential for sample overfitting in the presence of interaction between the factors. Apart from these two specifications, other specifications, such as VECM, E-GARCH, DCC-EGARCH and the Markov Switching Vector Autoregression Model, can be used to test the factor dynamics. The purpose of using different specification is to check whether the sample forecast can be improved by retaining the interactions between the factors and minimizing the in-sample overfitting generated by VAR.

The VECM model for lag one can be specified like this

$$\Delta\hat{\beta}_{t+1} = \hat{c} + \hat{\Gamma}\Delta\hat{\beta}_t + \hat{\gamma}ECT_t \quad (4.1.4)$$

Where, $\Delta\hat{\beta}_t = \hat{\beta}_t - \hat{\beta}_{t-1}$ and ECT_t is the error correction term of this model. The usefulness of an error correction specification is that allows us to model and test the existence of long-run relationships between variables, but also to examine short-run dynamics, which represent the adjustment to the long run equilibrium relationships. In this model the coefficient of $\Delta\hat{\beta}_t$ indicates the short-run dynamics between the factors. The ECT_t represents the long-term cointegrating relation. $\hat{\gamma}$ is the speed at which the variables return to their long-run values from the short-run deviation. Thus, VECM could capture the short and long-run dynamics of the yield curve. This can be very crucial for both macroeconomic and term structure modelling. In this paper, the Johansen MLE method (Johansen (1988)) is used to estimate the VECM model.

The next two models belong to the GARCH class of models. There are three main reasons for using such models of time series processes: volatility clustering, fat-tails of distributions and volatility mean reversion, which represent the so-called stylized facts frequently visible in the high volume of financial data. Although two-step DNS in itself

reduces the dimension of the data with the factor structure employed, some of these features can still be observed. A popular specification is GARCH (1,1), since it provides a good fit to many financial time series, given its ability to model both volatility persistence and the momentum within the system. Also, this model is very parsimonious as it can capture long lags of shocks with very few parameters. For this paper, a slightly different structure of GARCH is used to model every factor individually. Instead of using a linear specification, the conditional variance is modelled exponentially, as initially proposed by Nelson (1991), and also used by Christofi (1998) for fitting the NS yield factors.

$$\text{For individual coefficients, } \hat{\beta}_{i,t+1} = \hat{c}_i + \hat{\gamma}_i \hat{\beta}_{it} \quad \forall i = 1(1)3 \quad (4.1.5)$$

And, for the conditional variance:

$$\ln(\sigma_{i,t}^2) = \omega + \alpha_i Z_{i,t-1} + \delta_i \{ |Z_{i,t-1}| - E(|Z_{i,t-1}|) \} + \eta_i \ln(\sigma_{i,t-1}^2) \quad \forall i = 1(1)3 \quad (4.1.6)$$

$$\sigma_{ij,t} = \rho_{ij} \sigma_{i,t} \sigma_{j,t} \quad \forall i, j = 1(1)3, i \neq j \quad \& \quad Z_{i,t} = \frac{\varepsilon_{i,t}}{\sigma_{i,t}}$$

This exponential GARCH or EGARCH model was proposed by Nelson (1991). This is an extension of the general GARCH model. This type of EGARCH models is really helpful to model the asymmetries and shocks in the financial market. The δ_i represents the ARCH effect and the η_i represents the standard GARCH effect. The α_i , which represents the asymmetric effect, is one of the most factors of interest. If $\alpha_i = 0$ then the model is symmetric but when it is negative it suggests higher volatility through negative shock. Given the nature of yield factors, such a structure of conditional variance can be useful for modelling the varying volatility associated with different factors. The exponential GARCH or EGARCH model was used in Christofi(1998) for fitting on the NS yield factors.

This model can be extended to a multivariable context to capture the inner dynamics between the factors [see Christofi (1998) for this type of VAR-EGARCH process]. For this

paper, the dynamic control correlation or DCC-GARCH model will be used to fit the multivariable model following the methods of Engel and Sheppard (2001). In this model, for β_t with variance S_t , we can write:

$$S_t = D_t R_t D_t \quad (4.1.7)$$

Here, R_t is the matrix of the varying correlations and D_t is the diagonal matrix of the time-varying standard deviation from the E-GARCH model. This is such that-

$$\sigma_{i,t} = \sqrt{\exp[\omega + \alpha_i Z_{i,t-1} + \delta_i \{|Z_{i,t-1}| - E(|Z_{i,t-1}|)\}] + \eta_i \ln(\sigma_{i,t-1}^2)} \quad (4.1.8)$$

All the GARCH parameters are obtained by minimizing log-likelihood function which can be written as

$$L(\Theta) = \frac{1}{2} [NT \ln(2\pi) - \sum_{t=1}^T (\ln|S_t| + \varepsilon_t' S_t^{-1} \varepsilon_t)] \quad (4.1.9)$$

All these models are compared against each other in terms of model fitting and forecast ability, with the model with the lowest forecast error being preferred over the others.

4.1.2 One Step Modelling

The one-step algorithm to estimate DNS yield curve in Diebold et al (2006b) was an improvement on Diebold and Li (2006). In this method, the state space structure of the DNS is utilized. This helps to estimate all the parameters simultaneously. This model is explained in the equation from (2.16) to (2.18).

Now, the error covariance at time t for the factors in f_t can be written as-

$$P_t = E(e_t e_t') = E \left[(f_t - \hat{f}_t)(f_t - \hat{f}_t)' \right] \quad (4.1.10)$$

where, $e_t = (f_t - \hat{f}_t)$ and \hat{f}_t is the estimated value of f_t . Let us assume the prior estimate of \hat{f}_t is \hat{f}_t^p . The idea is that every time the algorithm runs, this estimation about the factors has been updated. This is done by combining old estimates with the measurement data –

$$\hat{f}_t = \hat{f}_t^p + K_t(y_t - \Lambda \hat{f}_t^p) \quad (4.1.11)$$

Here K_t = Kalman Gain and $(y_t - \Lambda \hat{f}_t^p)$ is called the innovation or the measurement residual.

This equation can be expanded such as –

$$\begin{aligned} \hat{f}_t &= \hat{f}_t^p + K_t(\Lambda f_t + \varepsilon_t - \Lambda \hat{f}_t^p) \\ \hat{f}_t &= \hat{f}_t^p + K_t\Lambda(f_t - \hat{f}_t^p) - K_t\varepsilon_t \end{aligned} \quad (4.1.12)$$

Now the error term for f_t will be

$$\begin{aligned} e_t &= f_t - \hat{f}_t = f_t - \hat{f}_t^p - K_t\Lambda(f_t - \hat{f}_t^p) + K_t\varepsilon_t \\ e_t &= (I - K_t\Lambda)(f_t - \hat{f}_t^p) + K_t\varepsilon_t \end{aligned} \quad (4.1.13)$$

and the error covariance would be

$$P_t = E(e_t e_t') = (I - K_t\Lambda)P_t^p(I - K_t\Lambda)' + K_t H_t K_t' \quad (4.1.14)$$

which can be expanded as

$$P_t = P_t^p - K_t\Lambda P_t^p - P_t^p \Lambda' K_t' + K_t(\Lambda P_t^p \Lambda' + H_t)K_t' \quad (4.1.15)$$

Now, all these matrices are known except K_t . The K_t needs to be such the error variance is minimized. If \mathcal{J}_t is the error variance matrix such that

$$\mathcal{J}_t = E[|e_t|^2] = E[e_t' e_t] = E[Tr(e_t e_t')] = Tr(P_t) \quad (4.1.16)$$

Then the minimization problem can be written as -

$$\min_{K_t} \mathcal{J}_t = \min_{K_t} Tr(P_t) \quad (4.1.17)$$

Which gives the updated K_t as

$$K_t = P_t^p \Lambda' (\Lambda P_t^p \Lambda' + H_t) \quad (4.1.18)$$

Now using this the covariance update equation becomes-

$$P_t = (I - K_t\Lambda)P_t^p \quad (4.1.19)$$

The priori projection in the next period is achieved using –

$$\begin{aligned} (f_{t+1} - \mu) &= A(f_t - \mu) \\ \hat{f}_{t+1} &= \mu + A(\hat{f}_t - \mu) \end{aligned} \quad (4.1.20)$$

And the priori error would be –

$$\begin{aligned} e_{t+1}^p &= f_{t+1} - \hat{f}_{t+1} \\ &= [\mu + A(f_t - \mu) + \eta_t] - \mu + A(\hat{f}_t - \mu) \\ &= Ae_t + \eta_t \end{aligned} \quad (4.1.21)$$

And subsequently, the priori error covariance would be

$$P_{t+1}^p = E[e_{t+1}^p e_{t+1}^{p'}] = AP_t A' + Q \quad (4.1.22)$$

This completes the algorithmic loop ³

4.2 Out of Sample Forecasting

In assessing a model's forecast ability, we need to evaluate not only its within-sample performance also its ability to forecast beyond the sample period. In this paper, forecast accuracy is measured using the root mean square error of forecasts based on a comparison of test sample and the model's forecasts of yield. The test sample includes the yields from May 2019 to May 2020. To calculate the yields the subsequent factors are forecasted for each model. Using those factors, the forecasted yields are calculated. These forecasted yields are then compared with the test sample yields by calculating the root mean square error between original and forecasted yields. The lower the root mean square error the better will be the accuracy.

³ For detail steps of the proof check Appendix 2

The basic assumption was that there is a degree of randomness in the way the yield curve moves along the time. So, any model outperforming the random walk model would be a valid response to model the yield curve.

Three primary random walk models are used as the base model these are.

$$\text{Random walk of yield (with a drift): } y_{t+1} = y_t + d + \varepsilon_t \quad (4.2.1)$$

$$\text{Random walk of factors(without a drift): } \beta_{i,t+1} = \beta_{i,t} + \eta_t \quad (4.2.2)$$

$$\text{Random walk of factors(with a drift): } \beta_{i,t+1} = \beta_{i,t} + d_i + \eta_t \quad (4.2.3)$$

These are the models that model the yield curve. each of this model is expected to outperform random walk to some degree.

AR(1): Given the AR(1) model as: $\hat{\beta}_{i,t+1} = \hat{c}_i + \hat{\gamma}_i \hat{\beta}_{i,t}$ the forecast would be –

$$\hat{\beta}_{i,t+h} = \hat{c}_i (\sum_{j=0}^{h-1} \hat{\gamma}_i^j) + \hat{\gamma}_i^h \hat{\beta}_{i,t} \quad (4.2.4)$$

where $\hat{\beta}_{i,t}$ is the estimated value at time t

VAR(1): Given the VAR(1) specification $\hat{\beta}_{t+1} = \hat{c} + \hat{\Gamma} \hat{\beta}_t$, the forecast would be –

$$\hat{\beta}_{t+h} = \hat{c} \left(\sum_{j=0}^{h-1} \hat{\Gamma}^j \right) + \hat{\Gamma}^h \hat{\beta}_t \quad (4.2.5)$$

VECM(1): Given the VECM(1) specification, $\Delta \hat{\beta}_{t+1} = \hat{c} + \hat{\Gamma} \Delta \hat{\beta}_t + \hat{\gamma} ECT_t$ the forecast would be $\hat{\beta}_{t+h} = \hat{\beta}_t + \sum_{i=0}^{h-1} \Delta \hat{\beta}_{t+i}$ such that -

$$\Delta \hat{\beta}_{t+i} = \hat{c} (\sum_{j=0}^{i-1} \hat{\Gamma}^j) + \hat{\Gamma}^i \Delta \hat{\beta}_t + \hat{\gamma} (\sum_{k=0}^{i-1} \hat{\Gamma}^k ECT_{t+i-1-k}) \quad (4.2.6)$$

E-GARCH(1,1): The conditional variance: $\ln(\sigma_{i,t}^2) = a_{t-1} + \eta_i \ln(\sigma_{i,t-1}^2)$ where,

$a_{t-1} = \omega + \alpha_i Z_{i,t-1} + \delta_i \{ |Z_{i,t-1}| - E(|Z_{i,t-1}|) \}$. The h period ahead forecast is calculated

$$\text{as } \ln(\sigma_{i,t+h}^2) = \sum_{m=0}^{h-1} a_{h-1-m} \eta_i^m + \eta_i^h \ln(\sigma_{i,t}^2) \quad (4.2.7)$$

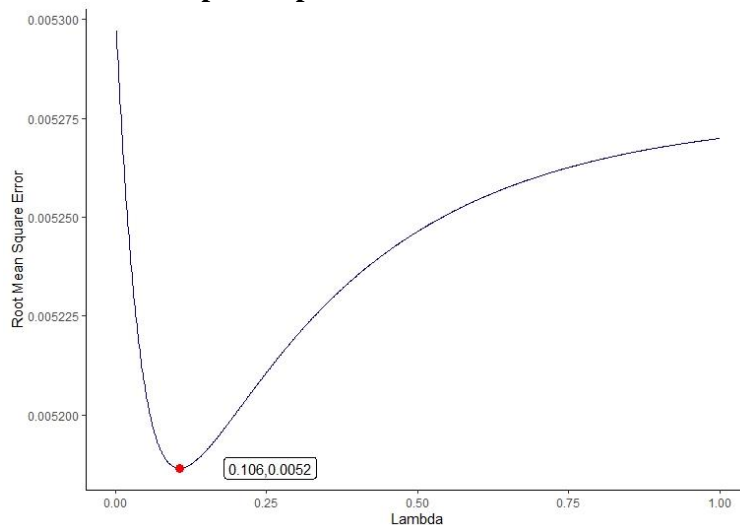
5. Results

The main goal of this paper is to compare different models to see which one is better in terms of in-sample modelling and how much it can improve on out of sample forecast

5.1 Two-Step Model

For two-step modelling, the first task is to fix the value of λ . In the original paper, Diebold & Li (2006) fixed this value to 0.0609 which was estimated from Diebold et.al. (2002). But that estimation was based on the US market and so that may not be the best value to use here. To find a feasible value for λ , a grid search is performed using the whole Two-step with VAR model discussed below. The data from the first 400 dates are used for the estimation and called train set. The prediction is performed on the rest of the data which is called test set. The model is trained for the whole train set for each value of λ . A prediction is generated for the next 13 periods, which will be compared with the test data. The goal is to find the λ that minimizes the root mean squared prediction error when compared to the test set. Our results from the algorithm indicate that the mean squared prediction error is minimized at $\lambda = 0.106$. This is shown in the graph below

Graph 4: Optimal value for Lambda



Using this value of λ , the cross-sectional yield regressions can be performed. This would give out the estimates of the three coefficients in the DNS model, β_1 , β_2 , and β_3 . The summary statistics below gives a good idea about some basic nature of these factors. Apart from the standard measures like mean and standard deviation, the auto-correlation function for one month, 12 months and 30 months and the partial autocorrelation are reported. Also, augmented Dicky-Fuller statistics are reported for all three coefficients. Given the ADF values, the null hypothesis that the factors are non-stationary can be rejected.

Table 2: Summary Statistics of the factors

Factor	Mean	Std. Dev	Min	Max	ACF (1)	ACF (12)	ACF (30)	PACF(1)	ADF
β_1	0.05616	0.0276	0.0106	0.1101	0.9905	0.9046	0.7775	0.9882	-4.104***
β_2	-0.01165	0.0160	-0.0461	0.0402	0.9707	0.4419	0.0402	0.9467	-3.438**
β_3	0.02431	0.0218	-0.0836	0.0447	0.9322	0.4030	0.1802	0.9244	-5.080***

*** p<0.01, ** p<0.05, * p<0.1

Now, the next step of the two-step DNS model is to figure out the dynamics between the factors, using the VAR model, which is the primary model for factor dynamics. Diebold and Li (2006a) used a VAR(1) specification for the sake of parsimony and transparency. Though lag selection measures are indicating differently, we choose the VAR (1) model. Also, the factor dynamics without the intercept is found to perform better in terms of out of sample forecast as we see later.

Table 3: Estimated VAR coefficients

Variables	β_1	β_2	β_3
L. β_1	0.9980*** (0.0023)	0.0056 (0.0079)	0.0035 (0.0052)
L. β_2	0.0031 (0.0037)	0.9459*** (0.0125)	0.0374*** (0.0082)
L. β_3	-0.126*** (0.0438)	0.123*** (0.0436)	1.156*** (0.0592)

Standard errors in parentheses
 *** p<0.01, ** p<0.05, * p<0.1

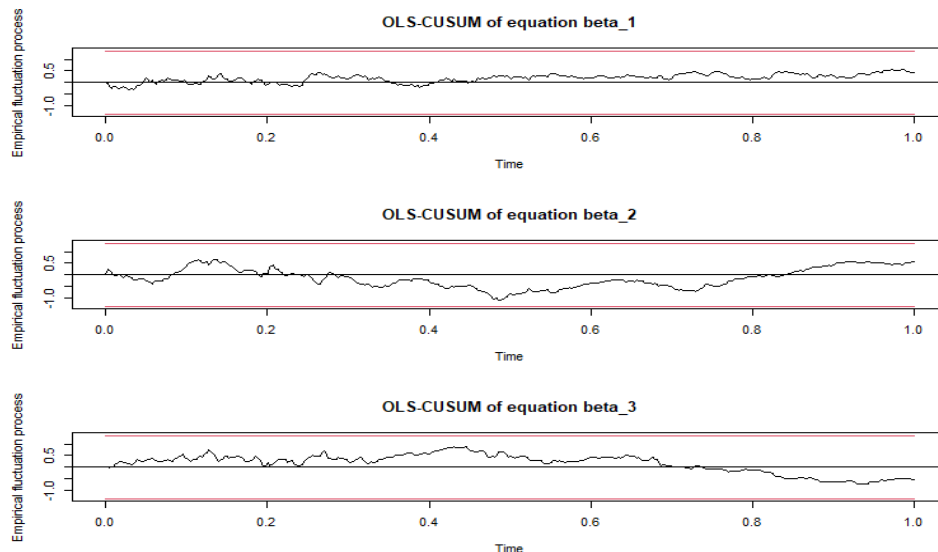
Also, the VAR(1) model, is the starting point for all two-step and one step estimations models. So, the validity of the model is very important. One of the basic assumptions of

the DNS is the normality assumption for the factors. Much of the estimation approaches are built upon the assumption of normality of the error of the factor dynamics, so we test that assumption using the Jarque-Bera test, as well as tests for skewness and kurtosis. The null hypothesis is that the VAR disturbance term is normally distributed. Unfortunately, all three tests reject the null hypothesis. This can be dealt with by working with alternative factor dynamic specifications. Apart from this, the CUSUM test is also performed to check the eigenvalue stability of the VAR. The null hypothesis is that a factor coefficient is stable, and it is only rejected when the coefficient exceeds the unit value. For individual coefficients, we find that all of them lie inside the unit circle, and so the null hypothesis cannot be rejected.

Table 4: VAR Normality Test

Jarque-Bera test	chi2	1892.9
	df	6
	Prob > chi2	0
Skewness test	chi2	79.827
	df	3
	Prob > chi2	0
Kurtosis test	chi2	1813
	df	3
	Prob > chi2	0

Graph 5: VAR Stability Test



Overall, we can be said that the VAR specification for the yield factors is stable and can be used for in sample modelling to some degree. But another goal of this study is to improve the out of sample forecast. The first alternative model to the VAR(1) is the VECM(1) model.

Table 5: Estimated VECM Coefficient

Variables	$\Delta\beta_1$	$\Delta\beta_2$	$\Delta\beta_3$
L. $\Delta\beta_1$	0.1885** (0.0579)	0.2498** (0.0854)	0.2033 (0.1931)
L. $\Delta\beta_2$	-0.0121 (0.0364)	0.3950*** (0.0538)	-0.0571 (0.1215)
L. $\Delta\beta_3$	0.0298 (0.0152)	0.0271 (0.0225)	0.3184*** (0.0507)
ECT	0.0004 (0.0009)	-0.0068*** (0.0014)	0.0161*** (0.0031)
Intercept	-0.0002 (0.0002)	0.0010*** (0.0003)	-0.0022*** (0.0006)

Standard errors in parentheses
 *** p<0.01, ** p<0.05, * p<0.1

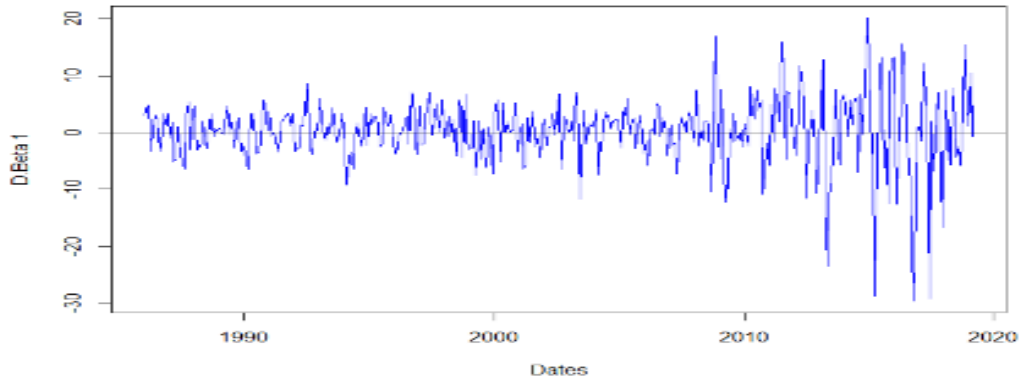
A vital feature of the VECM model is that it captures both long-run and short-run dynamics. The short-run factor dynamics are captured by the coefficient matrix of $\Delta\beta_t$. In line with expectations, we find that the individual $\Delta\beta_t$ have a significant relationship with their lagged counterparts. On the other hand, cross dynamics between the factors appear to be mostly absent, which somewhat solves the problem of overparameterization and gives a very parsimonious dynamic structure.

The error correction term shows the long-term Granger Causality such that when it is significant it shows a granger causality from dependent to the lagged independent factors, which indicates a long-run factor dynamics. As discussed in the section 4.1.1, the coefficient of error correction term measures the speed of convergence in the long run. So,

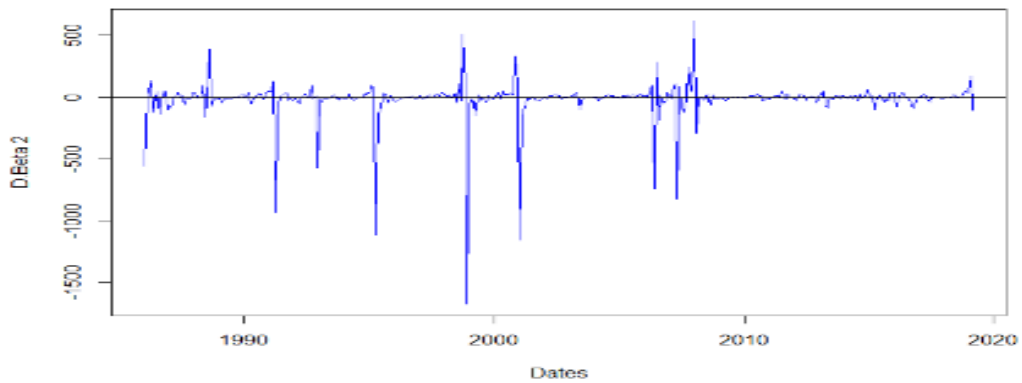
if the coefficient is zero (insignificant), there is no convergence or, in other words, no long-run factor dynamics. For a long run convergence to exist, the coefficient should be less than zero and greater than -1. In any other case, the relationship becomes explosive and never converges to a long-run value. The level factor does not have any long-run factor dynamics as the coefficient of the error correction term is not significant. It remains overall stable in the long run and only changes with the change in the intercept. On the other hand, both short and medium run factors seem to have some factor dynamics as the coefficient of the error correction term is significant. This is because they represent the slope and the curvature of the yield curve, respectively, which change in short and medium run resp. Given the negative sign, the short-term slope factor is converging, but the positive sign for the medium-term curvature factor implies divergence. The short-term factor is converging at a speed of 0.68 %, where the medium-term factor is diverging at 1.61%. Overall, the short-term factor has a long-run relationship, which is converging in nature, and the medium-term factor has a diverging long-run relationship.

The next two models of consideration are the GARCH class of models. From the VAR(1) model, is it evident that there are many rooms for improvement in terms of the variance structure of the DNS yield factors? There is enough evidence that would support the presence of a GARCH structure in the variance structure of yields. As mentioned earlier, GARCH models are useful in the presence of volatility clustering, which can be seen from Graph 2 as a small change of volatility is followed by a small change and large change of volatility is followed by another large change.

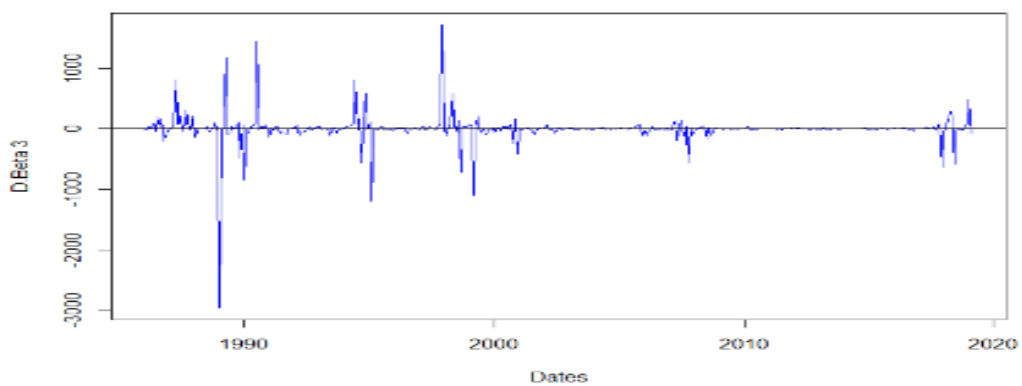
Graph 6: Volatility clustering of individual DNS factors



Beta 1



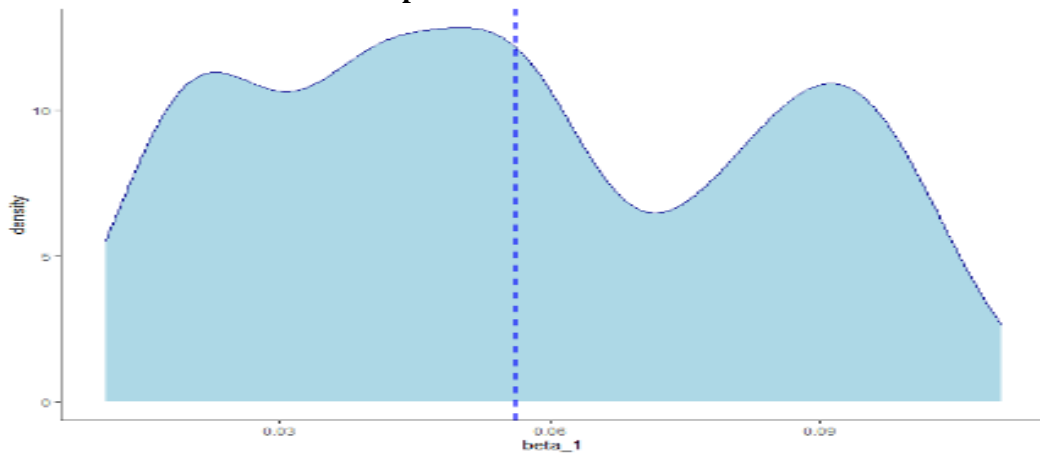
Beta 2



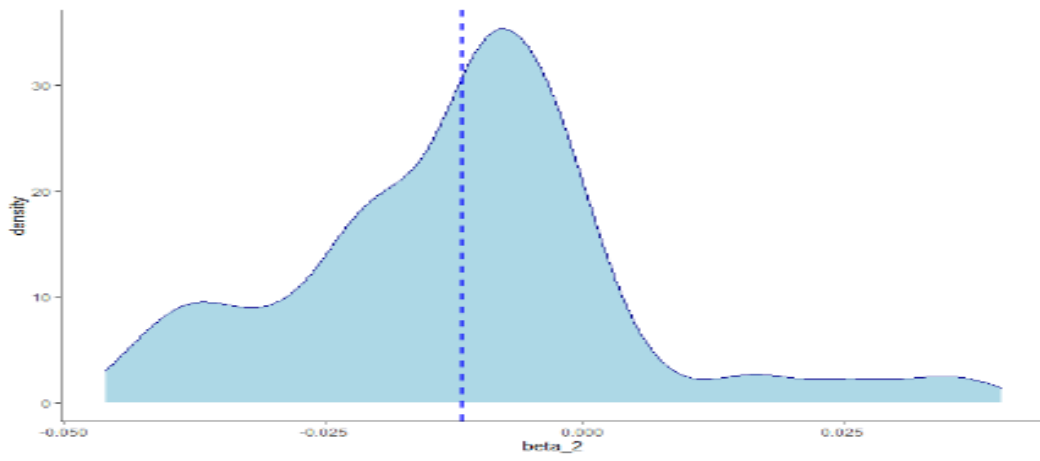
Beta 3

Another supporting piece of the GARCH structure is the presence of fatter tails, which is evident from every kleptocratic nature of the distribution of the DNS factors

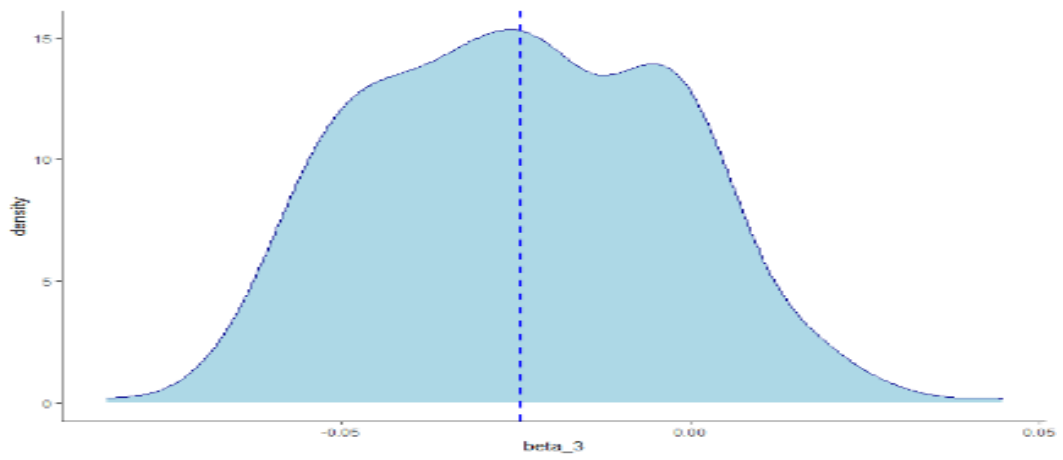
Graph 7: Distribution of the factors



Beta 1



Beta 2



Beta 3

The final evidence for GARCH is through the presence of mean reversion. Any mean-reverting time series can be represented by an Ornstein-Uhlenbeck stochastic differential equation –

$$d\beta_t = \theta(\mu - \beta_t)dt + \sigma dW_t \quad (5.1)$$

Here, θ is the rate of reversion and W_t is a Weiner process. This is a very common feature in financial time series specifically term structure of interest rate models. Many famous interest rate models such as Vasicek and CIR models uses a similar structure. So, the assumption of mean reversion is not out of context. This is also supported by the augmented dicky fuller statistics as all of them are significant rejecting the null hypothesis of no mean reversion. Given all these stylized facts the usefulness of the GARCH class of models become more evident. To maintain the parsimonious nature of DNS, the first model is the EGARCH model with AR(1) specification.

Table 6: Estimated EGARCH(1,1) Coefficient

Coefficients	β_1	β_2	β_3
μ_i	0.1011*** (0.0013)	-0.0007 (0.0022)	-0.0219*** (0.0037)
γ_i	1*** (0.0037)	0.9798*** (0.0134)	0.9665*** (0.0181)
ω_i	-3.3075 (5.4537)	-0.0721*** (0.0070)	-0.1701 (0.2556)
α_i	0.06845 (0.0602)	-0.0298*** (0.0079)	-0.0337 (0.0902)
η_i	0.7283 (0.4467)	0.9937*** (0.0005)	0.9822*** (0.0259)
δ_i	0.3148 (0.2936)	0.1464*** (0.0476)	0.2315*** (0.0533)

Standard errors in parentheses
 *** p<0.01, ** p<0.05, * p<0.1

For both β_1 (level) and β_3 (curvature) the model is symmetric as for both the factors the asymmetric effects or the α_i is insignificant or not different from zero. But in case of β_2 (slope) or the short-term factors this coefficient is negative and significant. This implies

that the negative shock can create higher volatility in the short run.

Next specification is the DCC-EGARCH specification which is just the multivariate representation of the EGARCH model. For better prediction accuracy, an ARMA (2,1) specification is used with this model.

Table 7: Estimated DCC-EGARCH Coefficient

	β_1	β_2	β_3	Joint
μ_i	0.1016*** (0.0018)	0.0108** (0.0050)	-0.0238** (0.0008)	
γ_1	0.6485*** (0.0026)	0.9651*** (0.0201)	0.9851*** (0.0254)	
γ_2	0.3535*** (0.0036)	0.0252** (0.3635)	-0.1015*** (0.0015)	
ς_i	0.6074*** (0.0329)	0.3663*** (0.0572)	0.2640*** (3.9304)	
ω_i	-0.0830*** (0.0030)	-0.0092 (0.0398)	0.0471*** (0.0029)	
α_i	0.0139 (0.0311)	-0.0521 (0.0389)	0.0810*** (0.0006)	
η_i	0.9933*** (0.000005)	0.9993*** (0.0035)	0.9956*** (0.2085)	
δ_i	0.0394*** (0.00034)	0.1253*** (0.0455)	-0.0480 (0.0296)	
dcca1				0.0543 (0.9015)
dccb1				0.9334** (0.3946)

Standard errors in parentheses
 *** p<0.01, ** p<0.05, * p<0.1

One of the interesting outcomes of this specification is the significance of the ARMA(2,1) coefficients. This gives a clear picture of the innovation or the dynamics of the error term. This is also evident from the stylized fact of the yield factors. A difference from the previous model is that in this specification there is no asymmetric evidence for slope factor.

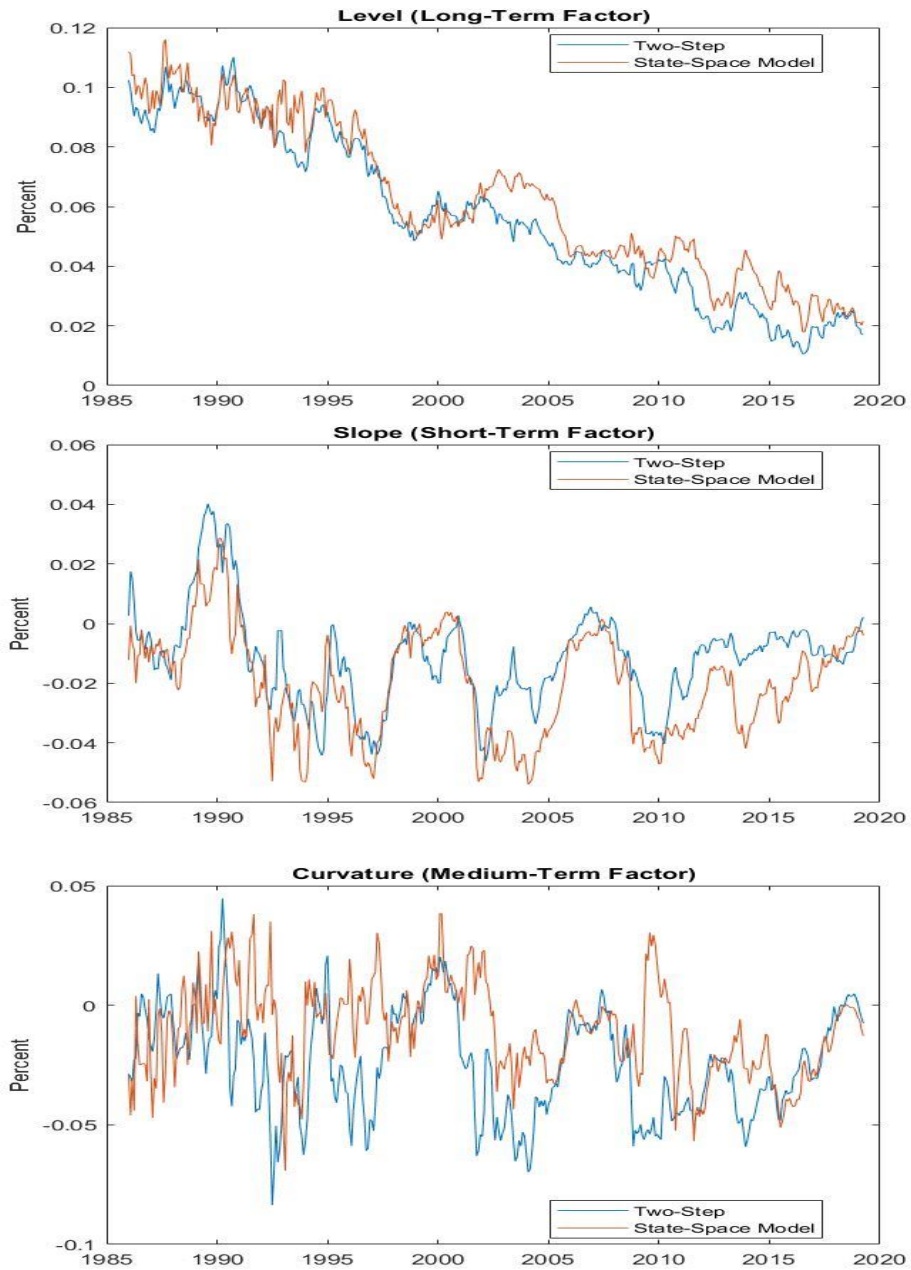
5.2 One Step Modelling

The one-step models use the state-space structure of the DNS to estimate parameter dynamics through Kalman filter. To begin the loop for Kalman filter, initial parameter dynamics, factor mean, and the decay factor are set at the level of the two-step VAR model. The initial disturbance loading is calculated from the initial VAR disturbance covariance matrix. The mean square error variance is minimized through maximizing the likelihood function of the DNS factors using Marquart & Berndt-Hall-Hall-Hausman algorithm. The algorithm loops for 25000 times which helps for faster calculation while maintaining the smoothness of the estimators. The estimated factor dynamics and disturbance loading is presented below.

	L_t	S_t	C_t
Parameter Estimates			
L_{t-1}	0.988797	0.01507	0.030347
S_{t-1}	0.008715	0.961461	-0.01494
C_{t-1}	0.017485	0.026293	0.895154
μ	0.061574	-0.01609	-0.01056
Estimated B Matrix (Disturbance Loading)			
L_t	0.00355		
S_t	-0.0018	0.004306	
C_t	-0.00675	-0.00317	0.007359

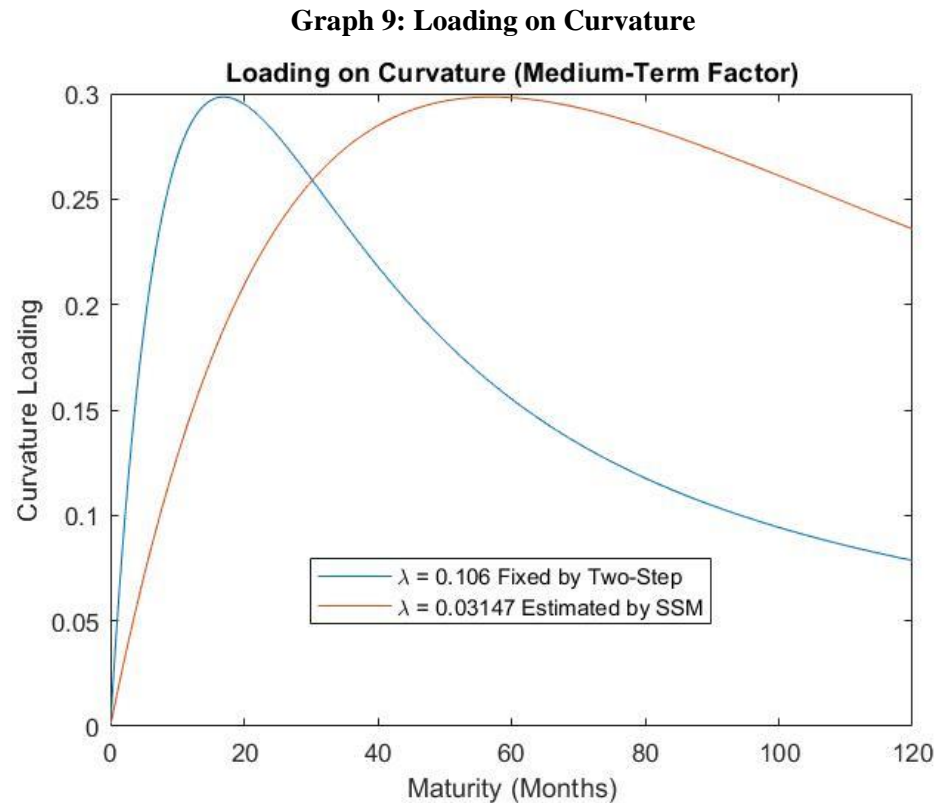
As expected from all the previous models, the own dynamics of the factors are 0.99, 0.96 and 0.89. The cross dynamics are small and not that important. Factor means are somewhat close but not similar to the two-step factor means. The disturbance loading seems increasing from level to slope to curvature factor. Overall, the two-step and the one-step factors seem to move very closely as seen in the following graphs

Graph 8: DNS Factors- Two-step vs State Space



The estimated decay rate from the state-space model is $\lambda = 0.03147$ which translates to nearly 57 months of maturity and is very different from the two-step fixed level of 0.106. The parameter λ determines at which point the curvature of the yield curve is maximized. From the two-step model, the curvature is maximized at the maturity of almost 17 months.

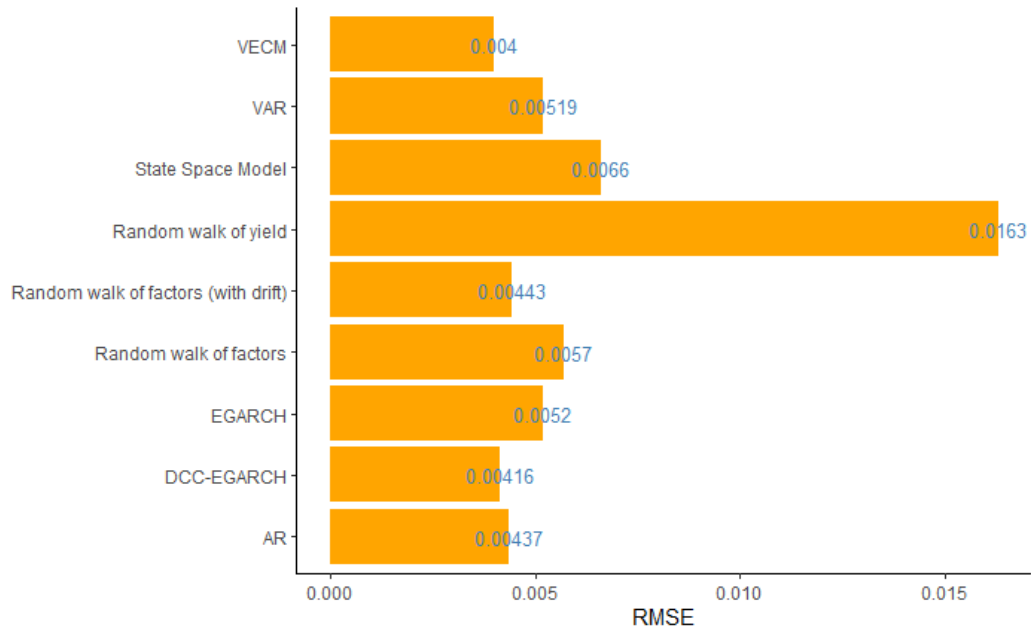
This indicates that the short-term yield (maturing in 1 to 4 year) is more valuable than the long-term (20+ year since maturity) yield. Graph 9 below plots the curvature factor loading obtained for the values of the decay parameter, against maturity.



5.3 Forecast Performance

All these models have some advantages in terms of parsimony, and volatility structure serving as a good explanation of the underlying dynamics. But a model of such kind can only be useful if it has better predictive powers. As mention earlier, the sample is divided for train and test set and the model is built upon the train set which is for 400 periods starting from January 1986 and ends at April 2019. The test set for the next 13 months which ends in June 2020. The goal is to choose the model which has a minimum mean square error. Also, another relevant issue is how any model compares to the random walk models. These calculated root-mean-square-errors are presented in a bar diagram.

Graph 10: Root Mean Square Error for different models



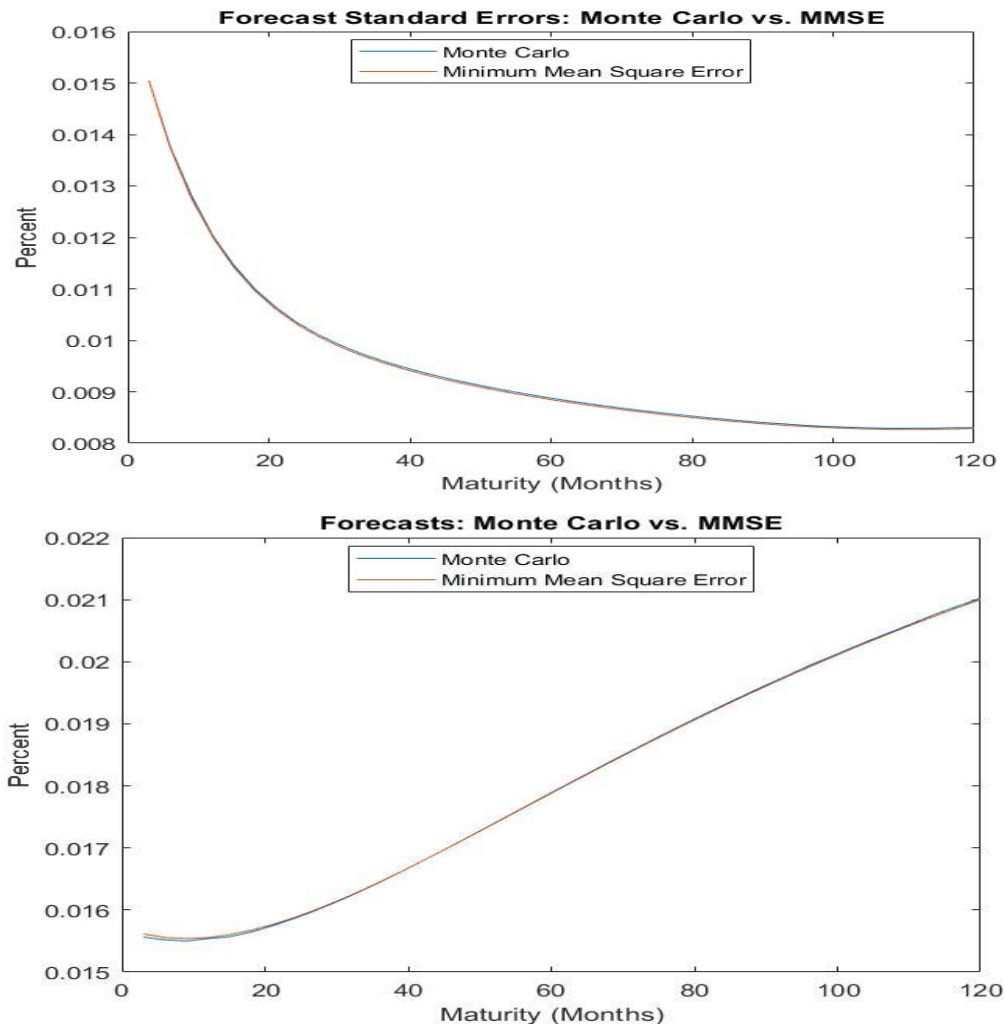
All the models have successfully outperformed the pure random walk of yield, which is expected. This is also a piece of very strong evidence in favour of the structural model or DNS. The root-mean-square errors of two other random walks of factors (with and without drift) are also presented as a base case. In the case of the no drift assumption EGARCH, AR, DCC-EGARCH and VECM all outperform the corresponding random walk, and AR, DCC-EGARCH and VECM still outperform the random walk with the drift assumption. Overall, hence, VECM and DCC-EGARCH appear to be the best models.

Two interesting facts can be derived from this. Firstly, the univariate dynamics generally performs better than multivariate dynamics. Only the multivariate model which has better theoretical justification for the inner dynamics can perform better than the univariate model. This overall shows the parsimonious nature of yield curve dynamics. Secondly, unlike all the two-step specifications, the state space models performed very poorly in terms of predictions. This may be caused by the initial values obtained from VAR which has its own

problems, such as the violation of assumptions (for example, the presence of non-normality).

Though the one-step model cannot outperform other models, its state space structure allows us to do simulations which would be very useful and will be accurate if the initial values are derived from the strong models like VECM or DCC-EGARCH. For this paper, the basic VAR estimations are used to simulate the yield. In the following two graphs, these simulations are presented along with the standard error using Monte Carlo and minimum MSE simulations. Both of these methods are showing an upward rising yield with similar error structure.

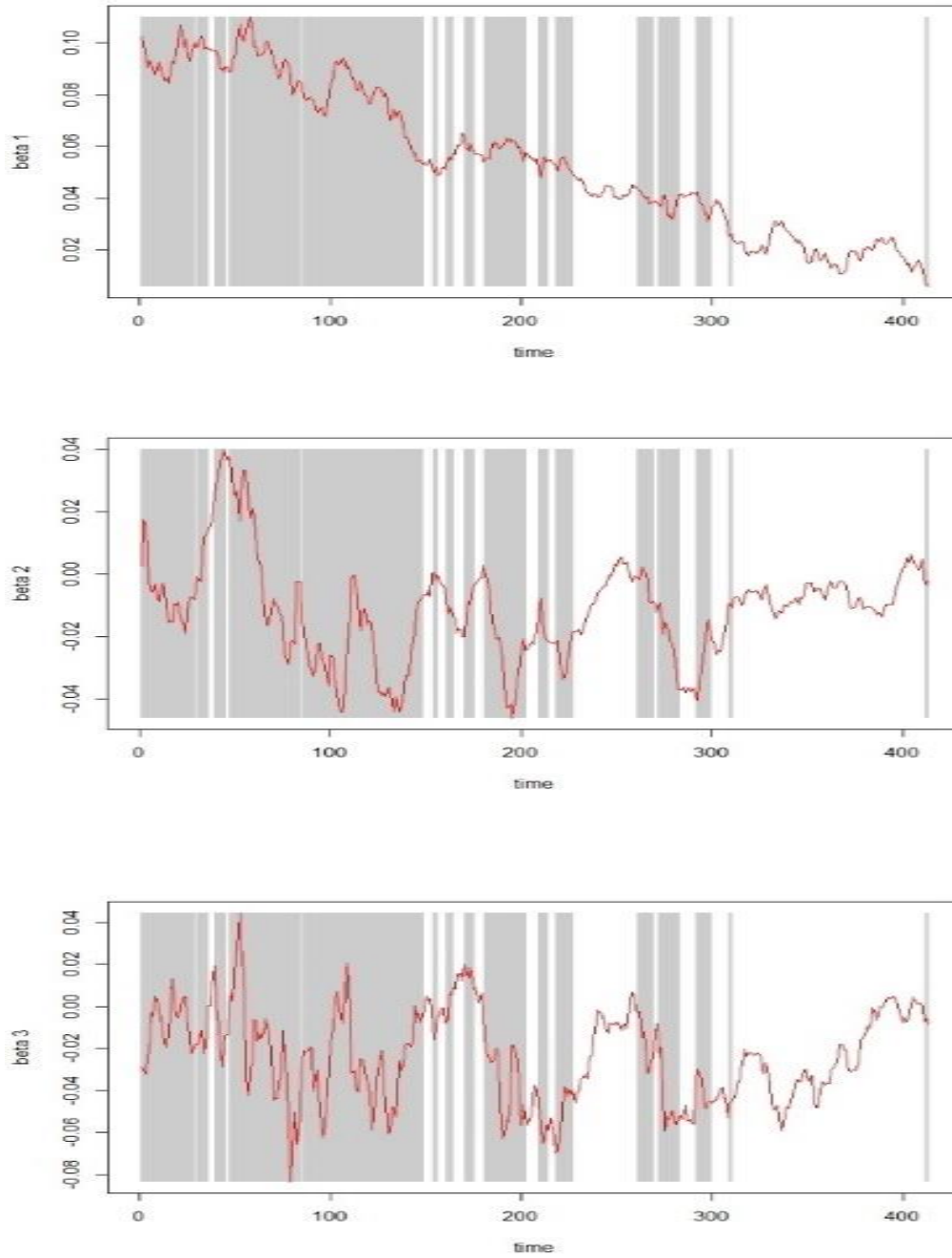
Graph 6: Yield Curve Simulation through MCMC and RMSE



5.4 Structural Break: Regime Switching Model

Though the var stability test indicates a stable model over the full period of observation, the mean reversion and the volatility clustering of the factors indicate otherwise. Given all the evidence, some structural breaks can be expected in the overall yield curve dynamics as well as in its factor structure. As argued by Xiang & Zhu (2013), the no-arbitrage restriction disadvantage for DNS as it cannot capture the effect of the different regime. They proposed an alternative Regime Switching DNS using the reversible jump Markov chain Monte Carlo (RJCMC) process. The model is mostly the same as normal VAR with regime dependent intercept and normally distributed disturbance term. Switches between regimes are governed by a discrete Markov chain with transition probability $p_{ij} = P(\text{regime} = j | \text{regime} = i) (\forall i \neq j)$. For this paper, the number of regimes kept at two which can be labelled as high and low volatility regimes respectively. This is not a bad assumption as in most of the financial literature maximum three volatility regimes are observed. This paper limits the study only for the modelling of the factors and finding evidence for the presence of any regime like structures in the factor. The factors are modelled using the estimated regime and the following graphs show the presence of difference regimes. These regimes indicate the changing Volatility of the yield curve. This was also indicated through the volatility clustering and mean reversal nature of the yield curve. In the graphs we can see that for a long time the yield curve factors are volatile but for last couple of years they became less volatile than before.

Graph 7: Different Regimes from regime-switching model Beta



6. Conclusion

The primary goal for this paper was to model the yield curve, with both good in-sample modelling as well as out of sample forecasting. As postulated earlier, the models which have better theoretical backing, like VECM or EGARCH, would be expected to have a better forecast performance. The yield curve or term structure of interest rates generally shows both long term and short-term dynamics. Both models incorporate these elements very well. This somehow connects back to the feedback loop between macro and financial parameters and how they can be modelled. This type of model turns out to be very parsimonious and thus can reveal many aspects of the macroeconomy through its factor structure and error variance.

As explained earlier, there are factor structures other than NS factors that can incorporate some of the temporal behaviour of the yield curve. In Cochrane and Piazzesi (2008), they integrate different factor structures different from NS or DNS. The affine model of yield curve using the bond structure premia used by Cochrane and Piazzesi (2008) can still outperform DNS in terms of forecasting. However, in terms of macro-finance models, these are the best specifications allowing for expansion.

References

- Aljinović, Zdravka, Tea Poklepović, and Kristina Katalinić. "Best fit model for yield curve estimation." *Croatian Operational Research Review* 3.1 (2012): 28-40.
- Ayouche, S., Ellaia, R., & Aboulaich, R. (2016). A Hybrid Particle Swarm Optimization-Nelder-Mead Algorithm (PSO-NM) for Nelson-Siegel-Svensson Calibration. *International Journal of Economics and Management Engineering*, 10(4), 1365-1369.
- Bandara, Wachi, and Richard Munclinger. "Term structure modeling with structural breaks: a simple arbitrage-free approach." *Available at SSRN 1974033* (2011).

- Bernard, Henri, and Stefan Gerlach. "Does the term structure predict recessions? The international evidence." *International Journal of Finance & Economics* 3.3 (1998): 195-215.
- Björk, Tomas, and Bent Jesper Christensen. "Interest rate dynamics and consistent forward rate curves." *Mathematical Finance* 9.4 (1999): 323-348.
- Bliss, Robert Russell. "Testing term structure estimation methods." *Advances in Futures and Options research* 9 (1997): 197-232.
- Bolder, D. J., Johnson, G., & Metzler, A. (2004). An empirical analysis of the canadian term structure of zero-coupon interest rates. Available at SSRN 1082833.
- Caldeira, João F., et al. "Forecasting the yield curve with the arbitrage-free dynamic Nelson–Siegel model: Brazilian evidence." *Economia* 17.2 (2016): 221-237.v
- Chauvet, M., & Potter, S. (2005). Forecasting recessions using the yield curve. *Journal of forecasting*, 24(2), 77-103.
- Choudhry, Moorad, and Michele Lizzio. *Advanced fixed income analysis*. Elsevier, 2015
- Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica: Journal of the Econometric Society*, 591-605.
- Christensen, Jens HE, Francis X. Diebold, and Glenn D. Rudebusch. "The affine arbitrage-free class of Nelson–Siegel term structure models." *Journal of Econometrics* 164.1 (2011): 4-20.
- Cochrane, J. H., & Piazzesi, M. (2009). Decomposing the yield curve. In *AFA 2010 Atlanta Meetings Paper*.
- Diebold, F. X. "Li., C. (2002), "Forecasting the Term Structure of Government Bond Yields,"." *Manuscript, Department of Economics, University of Pennsylvania*.
- Diebold, Francis X., and Canlin Li. "Forecasting the term structure of government bond yields." *Journal of econometrics* 130.2 (2006): 337-364.
- Diebold, Francis X., and Glenn D. Rudebusch. *Yield curve modeling and forecasting: the dynamic Nelson-Siegel approach*. Princeton University Press, 2013.
- Diebold, Francis X., Canlin Li, and Vivian Z. Yue. "Global yield curve dynamics and interactions: a dynamic Nelson–Siegel approach." *Journal of Econometrics* 146.2 (2008): 351-363.
- Diebold, Francis X., Glenn D. Rudebusch, and S. Boragan Aruoba. "The macroeconomy and the yield curve: a dynamic latent factor approach." *Journal of econometrics* 131.1-2 (2006): 309-338.
- Diebold, Francis X., Lei Ji, and Canlin Li. "A three-factor yield curve model: non-affine structure, systematic risk sources, and generalized duration." *Macroeconomics, Finance and Econometrics: Essays in Memory of Albert Ando* 1 (2006): 240-274.
- Diebold, Francis X., Monika Piazzesi, and Glenn D. Rudebusch. "Modeling bond yields in finance and macroeconomics." *American Economic Review* 95.2 (2005): 415-420.
- Fama, Eugene F., and Robert R. Bliss. "The information in long-maturity forward rates." *The American Economic Review* (1987): 680-692.

- Filipović, Damir. "A note on the Nelson–Siegel family." *Mathematical finance* 9.4 (1999): 349-359
- Geyer, Charles J. "Practical markov chain monte carlo." *Statistical science* (1992): 473-483.
- Gimeno, R., & Nave, J. M. (2006). Genetic algorithm estimation of interest rate term structure. *Banco de Espana Research Paper No. WP-0634*.
- Hautsch, N., & Yang, F. (2012). Bayesian inference in a stochastic volatility Nelson–Siegel model. *Computational Statistics & Data Analysis*, 56(11), 3774-3792.
- Levant, J., & Ma, J. (2017). A dynamic Nelson-Siegel yield curve model with Markov switching. *Economic Modelling*, 67, 73-87.
- Levant, Jared, and Jun Ma. "Investigating United Kingdom's monetary policy with Macro-Factor Augmented Dynamic Nelson–Siegel models." *Journal of Empirical Finance* 37 (2016): 117-127.
- Litterman, Robert, and Jose Scheinkman. "Common factors affecting bond returns." *Journal of fixed income* 1.1 (1991): 54-61.
- McCulloch, J. Huston. "An estimate of the liquidity premium." *Journal of Political Economy* 83.1 (1975): 95-119..
- Parker, E. (2017). The entropic linkage between equity and bond market dynamics. *Entropy*, 19(6), 292.
- Svensson, Lars EO. *Estimating and interpreting forward interest rates: Sweden 1992-1994*. No. w4871. National bureau of economic research, 1994.
- Tsui, A., J. X. Wu, and Zhaoyong Zhang. "Modelling the term structure of Japanese bond yields with the Nelson-Siegel model." (2013).
- Vasicek, Oldrich A., and H. Gifford Fong. "Term structure modeling using exponential splines." *The Journal of Finance* 37.2 (1982): 339-348.
- Xiang, J., & Zhu, X. (2013). A regime-switching Nelson–Siegel term structure model and interest rate forecasts. *Journal of Financial Econometrics*, 11(3), 522-555

Appendix 1: Limit of the yield curve

$$\begin{aligned}
 \lim_{\tau \rightarrow 0} y(\tau) &= \lim_{\tau \rightarrow 0} \left[\beta_0 + \beta_1 \left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} \right) + \beta_2 \left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right) \right] \\
 &= \beta_0 + \beta_1 \lim_{\tau \rightarrow 0} \left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} \right) + \beta_2 \lim_{\tau \rightarrow 0} \left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right) \quad (\text{betas are constant}) \\
 &= \beta_0 + \beta_1 \lim_{\tau \rightarrow 0} \left(\frac{\lambda e^{-\lambda\tau}}{\lambda} \right) + \beta_2 \lim_{\tau \rightarrow 0} \left(\frac{\lambda e^{-\lambda\tau}}{\lambda} \right) - \beta_2 \lim_{\tau \rightarrow 0} e^{-\lambda\tau} \quad (\text{l' hospital rule}) \\
 &= \beta_0 + \beta_1 \\
 &= f(0) = r \quad (\text{instantaneous short rate})
 \end{aligned}$$

$$\lim_{n \rightarrow \infty} y(\tau) = \lim_{n \rightarrow \infty} \left[\beta_0 + \beta_1 \left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} \right) + \beta_2 \left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right) \right]$$

$$\begin{aligned}
&= \beta_0 + \beta_1 \lim_{n \rightarrow \infty} \left(\frac{1 - e^{-\lambda \tau}}{\lambda \tau} \right) + \beta_2 \lim_{n \rightarrow \infty} \left(\frac{1 - e^{-\lambda \tau}}{\lambda \tau} - e^{-\lambda \tau} \right) \quad (\text{betas are constant}) \\
&= \beta_0 + \beta_1 \lim_{n \rightarrow \infty} \left(\frac{\lambda e^{-\lambda \tau}}{\lambda} \right) + \beta_2 \lim_{n \rightarrow \infty} \left(\frac{\lambda e^{-\lambda \tau}}{\lambda} \right) - \beta_2 \lim_{n \rightarrow \infty} e^{-\lambda \tau} \quad (\text{l' hospital rule}) \\
&= \beta_0 = \text{Constant}
\end{aligned}$$

Appendix 2: Kalman Filter for the One-step DNS model

$$P_t = E(e_t e_t') = E \left[(f_t - \hat{f}_t)(f_t - \hat{f}_t)' \right]$$

where, $e_t = (f_t - \hat{f}_t)$ and \hat{f}_t is the estimated value of f_t . Now, let's assume the prior estimate of \hat{f}_t is \hat{f}_t^p . The idea is that every time the algorithm runs, this estimation about the factors has been updated. This is done by combining old estimates with the measurement data –

$$\hat{f}_t = \hat{f}_t^p + K_t(y_t - \Lambda \hat{f}_t^p)$$

Here K_t = Kalman Gain and $(y_t - \Lambda \hat{f}_t^p)$ is called the innovation or the measurement residual.

This equation can be expanded such as –

$$\hat{f}_t = \hat{f}_t^p + K_t(\Lambda f_t + \varepsilon_t - \Lambda \hat{f}_t^p)$$

$$\hat{f}_t = \hat{f}_t^p + K_t \Lambda (f_t - \hat{f}_t^p) - K_t \varepsilon_t$$

$$P_t = E(e_t e_t') = E \left[\left\{ (I - K_t \Lambda)(f_t - \hat{f}_t^p) + K_t \varepsilon_t \right\} \left\{ (I - K_t \Lambda)(f_t - \hat{f}_t^p) + K_t \varepsilon_t \right\}' \right]$$

$$P_t = E \left[(I - K_t \Lambda)(f_t - \hat{f}_t^p)(f_t - \hat{f}_t^p)' (I - K_t \Lambda)' + (I - K_t \Lambda)(f_t - \hat{f}_t^p) \varepsilon_t' K_t' + \right.$$

$$\left. K_t \varepsilon_t (I - K_t \Lambda)(f_t - \hat{f}_t^p) + K_t \varepsilon_t \varepsilon_t' K_t' \right]$$

$$P_t = (I - K_t \Lambda) E \left[(f_t - \hat{f}_t^p)(f_t - \hat{f}_t^p)' \right] (I - K_t \Lambda)' + (I - K_t \Lambda) E \left[(f_t - \hat{f}_t^p) \varepsilon_t' K_t' \right] +$$

$$E \left[K_t \varepsilon_t (f_t - \hat{f}_t^p)' (I - K_t \Lambda)' \right] + K_t E(\varepsilon_t \varepsilon_t') K_t'$$

$$P_t = (I - K_t \Lambda) P_t^p (I - K_t \Lambda)' + K_t H_t K_t'$$

$$P_t = P_t^p - K_t \Lambda P_t^p (I - K_t \Lambda)' + K_t H_t K_t'$$

$$P_t = P_t^p - K_t \Lambda P_t^p - P_t^p \Lambda' K_t' + K_t \Lambda P_t^p \Lambda' K_t' + K_t H_t K_t'$$

$$P_t = P_t^p - K_t \Lambda P_t^p - P_t^p \Lambda' K_t' + K_t (\Lambda P_t^p \Lambda' + H_t) K_t'$$