Modeling Swedish government yields with the Dynamic Nelson Siegel and the Dynamic Nelson Siegel Svensson Model

Malick Senghore
Masters Thesis,Lund University, Sweden.

CONTENTS

ABSTRACT **1**

1 INTRODUCTION **2**

2 INTEREST RATES THEORY **4**

2.1 The Stochastic discount factor and the Spot interest rates 4
2.2 Zero Coupon bonds 4
2.2.1 Zero-Bond Curve 5
2.3 The Day-Count Convention and The Compounding Types 6
2.3.1 Day-Count Convention (year fraction) 6
2.3.2 Compounding Types 6
2.3.3 Continuously compounded interest rates 6
2.3.4 Simply compounded rate/The LIBOR 7
2.4 Yield Curve/The Zero-coupon Curve 8
2.5 Coupon Bonds 8
2.5.1 Fixed Coupon Bonds 9
2.5.2 Floating rate Bonds 9
2.6 Forward rates 10
2.7 Arbitrage Theory 11
2.7.1 Risk Neutral Valuation Formula 16
2.7.2 Numeraires 17
2.7.3 Change of Numeraire 17

3 DYNAMIC MODELS **19**

3.1 Vector Autoregressive Process (VAR) 19
3.2 State Space Modeling 19
3.2.1 The Kalman Filter 20

4 SHORT RATES **22**
4.1 Term Structure 22
4.1.1 Affine Term Structure (ATS) 23
4.2 Short interest rate models 24
4.2.1 Term Structure Equation (TSE) 25
4.2.2 The Merton model 27
4.2.3 The Vasicek model 28
4.2.4 The Cox, Ingersoll and Ross (CIR) model 29
4.2.5 The Dothan model 30

5 THE NELSON-SIEGEL MODELS **32**
5.1 Nelson-Siegel Model (NS) 32
5.1.1 Dynamic Nelson-Siegel Model (DNS) 35
5.1.2 Arbitrage Free Nelson-Siegel model (AFNS) 36
5.1.3 Dynamic Nelson-Siegel Svensson model (DNSS) 41

6 APPLICATION OF THE NELSON SIEGEL MODELS **43**
6.1 Data 43
6.2 Estimation Procedure Overview 46
6.2.1 Estimation Procedure DNS 48
6.2.2 Estimation Procedure DNSS 51
6.2.3 Estimation Procedure AFNS 52
6.2.4 Forecasting Procedure 53
6.3 Estimation Results and Analysis 55
6.3.1 DNS Estimation Results 55
6.3.2 AFNS Estimation Results 61
6.3.3 DNSS Estimation Results 66
6.4 Forecasting 72
6.5 Conclusion 74
6.6 Extension 74

BIBLIOGRAPHY **75**

ABSTRACT

The purpose of this thesis is to model and forecast Swedish government yields by using three classes of the *Nelson Siegel Model Family*. The three models considered are the *Dynamic Nelson Siegel Model*, *Arbitrage-Free Nelson Siegel Model* and *Dynamic Nelson Siegel Svensson Model*. A brief introduction to interest rate theory is given with emphasis on coupon bonds and yield curves. To introduce the concepts needed for the arbitrage-free model, arbitrage theory is introduced. The modeling framework used in this thesis implements the Kalman Filter, thereby necessitating introduction of State Space modeling and the derivation of the Kalman Filter.

The Nelson Siegel model classes under study are introduced and an estimation procedure for each model is detailed. In general, all model parameters are estimated by both cross-sectional and time-series optimization. The method of estimation employed ensures that we have stable and meaningful estimates. Our modeling procedure, shows that indeed the independent three factor Dynamic Nelson Siegel model do represent well Swedish government bonds both in-sample fit and out-of-sample forecast.

Chapter 1

INTRODUCTION

In my bachelors thesis, I gave a detailed account of the history of the bond market in Sweden and the theory behind *fixed income securities* in general. Swedish government bonds were then fitted with the *Diebold and Li (2006)* interpretation of the *Original Nelson Siegel*, by applying the *two-step-approach* described in their paper.

In this thesis, I intend to extend on the results of my bachelors thesis and investigate, amongst the various *Nelson Siegel* model families, the one that best estimates and forecast the Swedish government yields. I will consider two main classes of the *Nelson Siegel family*, i.e. the *Dynamic Nelson Siegel (DNS)* and the *Dynamic Nelson Siegel Svensson (DNSS)*; with both independent- and correlated factor models will be studied. For the empirically based *DNS* models, I will also study its arbitrage-free theoritically based models, namely the *Arbitrage Free Nelson Siegel (AFNS)*, by following the work of *CDR (2008)* and model the dynamics of the its factors under the Q-measure as well as add a yield-adjustment term to the yields measurement equation.

Various authors have taken this road before and there are ample literatures on the models I intend to work with. However, the estimation approach that I will implement in this thesis makes my contribution unique. I will employ both cross-section and time-series optimization to obtain estimates of the parameters of interest, in such a way that the factor estimates will be stable, statistically and economically meaningful.

Basically, I represent the estimation problem in *state-space* form, which enable me to use the Kalman Filter to obtain optimal parameter estimates by using the prediction error decomposition of the the likelihood of the parameters given the data, and estimate all the model parameters simultaneously. This is in line with the *one-step approach* proposed by *Diebold and Li (2006)*.

Instead of using the parameter estimates from the Kalman Filter Maximum Likelihood, as described above, and build the model factor dynamics, I used only the estimate(s) of the decay parameter(s) and fixed the information matrix in our models and thereafter compute the model factors by using *Ordinary least Squares (OLS)* method. This approach ensures that the factor estimates are stable and ready to use for forecasting.

I will exhaustively search for optimal initial parameters. This step is necessitated by the fact that the modeling framework that I adopted in this thesis is very sensitive to initial values. For each model, based on its structure, I described and applied the best procedure to obtain these initial parameters.

The thesis proceeds as follows. In the Chapter 2, I gave a not so short introduction to interest rates theory with emphasis on *coupon bonds* and *arbitrage theory*. In Chapter 3, I introduced the Kalman Filter and work through its derivation. In Chapter 4, *Affine Processes* and some well known short rates models are introduced. In Chapter 5, I introduced the Nelson Siegel family that will be relevant for this thesis. In the last part of the thesis, I applied the Nelson Siegel models on Swedish government yields and analyzed their in-sample- and out-of-sample-fit.

Chapter 2

INTEREST RATES THEORY

As mentioned in my introduction, the theory of interest rates was detailed in my bachelors thesis. For completeness, I will briefly discuss the interest rate theory needed to understand the methods and terminologies related to this thesis. I therefore start by introducing the *stochastic discount factor* and the *spot rates*.

Our goal in interest rate theory is to study and understand the variability of the underlying assets, i.e the interest rates. It is therefore very important to allow the interest rates to be stochastic and not assumed fixed, as is usually the case in other markets. For example, when studying stock options, the interest rates are assumed to be fixed and the stocks, which are the underlying assets, are allowed to vary.

2.1 The Stochastic discount factor and the Spot interest rates

The Stochastic discount factor D(t,T) is basically used to relate amounts of money at two different time points. The stochastic discount factor D(t,T) between two time instant t and T is the amount at time t that is equivalent to one unit of currency payable at time T, and is given by

$$D(t,T) = \frac{B(t)}{B(T)} = \exp\left(\int_t^T r_s ds\right)$$

The stochastic discount factor leads us to the simplest forms of loans in the *money market*, namely the *zero-coupon bonds.*

2.2 Zero Coupon bonds

Zero-coupon bonds are the simplest form of loan in the money market. They have only one payment stream under their whole lifetime, that is, the face value of the bond payable to the bond holder at maturity. Zero coupon bonds are in practice not directly observable in the market. Long maturities ZCB are not traded at all, they can however be obtained by *bootstrapping* coupon bonds.

The zero coupon bonds are also sometimes referred to as fixed income securities as they provide the owner with a deterministic amount, that is known when the bond is issued.

Formally, a zero coupon bond that matures at time T is defined as:

Definition 2.1. *A T-maturity zero coupon bond is a contract that guarantees its holder the payment of one unit of currency at time T, with no intermediate payments. The contract value, i.e the price of a zero coupon bond at time $t < T$ is denoted by P(t,T) and P(T,T) = 1 for all maturity times T and is equal to the present value of the nominal amount which can be written as:*

$$P(t, T) = \frac{P(T, T)}{1 + r \times \frac{d}{360}}$$

where r is a deterministic interest rate and d refers to number of days remaining for the zero coupon bond to mature.

In this thesis, r is allowed to be stochastic which implies that the zero coupon bond price is also stochastic and its graph is referred to as the *zero coupon curve.*

2.2.1 Zero-Bond Curve

The bond price is a stochastic process with two random variables, t and T. If t is fixed, then $P(t, T)$ is a smooth function of T. This function provides the prices, at the fixed time t, for the bonds of all possible maturities.

The graph of this function is called "the bond price curve at t" or the term structure at t of the discount factor. Formally we can define the zero-bond curve as:

Definition 2.2. *The zero-bond curve at time t, also known as the term structure of the discount factors, is the graph of the function*

$$T \mapsto P(t, T), T > t$$

which because of the positivity of interest rates, is a T-decreasing function starting from $P(t, t) = 1$.

If on the other hand, T is fixed, the price $P(t, T)$ will be a scalar stochastic process. It gives the prices at different times of a bond with fixed maturity T and its trajectory is very irregular.

As mentioned earlier, T-bond are not directly observable in the market, we therefore need two fundamental features of the interest rates to be able to relate zero coupon bond prices to interest rates. These features are the *Day-Count convention* to be applied in the rate definition and the *Compounding-Type.*

It is important to note that some assumptions must be made to guarantee the existence of a market that is sufficiently rich and regular where these bonds are traded. That is, we have to assumed that there exists a frictionless market for zero-coupon bonds for every maturity time T and that the price of a zero-coupon bond at time t equals one, i.e $P(t, t) = 1$ for all t.

The assumption that $P(t,t) = 1$ is necessary to ensure that we avoid arbitrage pricing. We also have to assumed that for each fixed time $t < T$, the price of a zero-coupon bond that matures at time T is differentiable with respect to the time of maturity T.

2.3 The Day-Count Convention and The Compounding Types

The compounding types and the day-count convention are the two fundamental properties of interest rates that are needed to enable us to used zero-coupon bonds to price interest rates. These two properties are described below.

2.3.1 Day-Count Convention (year fraction)

We denote by $\tau(t,T)$ the chosen time measure between t and T, which is usually referred to as year fraction between the dates t and T. When t and T are less than one-day distant, $\tau(t,T)$ is to be interpreted as the time difference T-t (in years). The day-count convention helps us to compute the interest payable at the end of an interest- or loan period.

The numerator represents the number of day in the interest- or loan period and the denominator represents the number of days in the reference period, which in Sweden is 360 days. Observed that this is one of many definitions for the day-count and will suffice for our purpose.

2.3.2 Compounding Types

The compounding type refers to how the interest rate is computed based on both the initial principal amount invested and the accumulated interest generated in the earlier periods.

Basically, the compounding types can be classified into four main categories, i.e. *continuously-compounded rates*, *simply-compounded rates*, *k-times-per-year compounded rates* and *annually-compounded rates*. Of these four compounding types, which can be expressed both as forward rates or spot rates, the simply-compounding type, also called the **LIBOR rates**, is the most commonly used both in theory and in practice. In this thesis however, continuously compounded interest rates will be used. Below we described the continuously compounded interest rates and the simply compounded rates.

2.3.3 Continuously compounded interest rates

Basically, continuously compounding type is the constant rate prevailing on an investment on a zero-coupon bond at time t for maturity at a future time interval $[S, T]$ that yields a unit of currency at time of maturity.

If the contracting date coincides with the start of the interval, then we have a continuously-compounded spot rate, otherwise we have a continuously-compounded forward rate. The continuously compounded spot rates and the continuously compounded forward rates are describe below:

The continuously compounded forward rate contracted at time t for the period $[S,T]$ is defined as

$$R(t;S,T) = -\frac{\log P(t,T) - \log P(t,S)}{\tau(T,S)}$$

The continuously compounded spot rate contracted at time S for the period $[S,T]$ is defined as:

$$R(S;S,T) = -\frac{\log P(S,T)}{\tau(T,S)}$$

The continuously compounded spot rate is a constant rate, from which we can derive the price of a zero coupon bond as

$$R(S;S,T)\tau(S,T) = -\log P(S,T) \tag{2.3.1}$$

$$\Rightarrow P(S,T) = \exp(R(S,T)\tau(S,T)) \tag{2.3.2}$$

2.3.4 Simply compounded rate/The LIBOR

When accruing occurs proportionally to the time of the investment then we have a simply compounded spot rate $L(t,T)$. The simply compounded rate is also referred to as the LIBOR rates. LIBOR stands for *London Inter-Bank Offer Rates* and it is the rate used most commonly in the market. By definition:

Definition 2.3. ***The LIBOR interest rates***
The LIBOR rate prevailing at time t for the maturity T is the constant rate at which an investment has to be made to produce an amount of one unit of currency at maturity, starting from P(t,T) units of currency at time t, when accruing occurs proportional to the investment time.

The LIBOR rates are forward rates that can either be quoted as continuously compounded rates or simple rates. The simple rates notation is the one most commonly used in the markets, whereas the continuously compounded notation is used for theoretical purposes. The simple LIBOR rate can be quoted as forward rates or spot rates as follows:

The simple forward rate contracted at time t for the period $[S,T]$ is called the LIBOR forward rate and is defined as

$$L(t;S,T) = -\frac{P(t,T) - P(t,S)}{\tau(T,S)P(t,T)}$$

The simple spot rate contracted at time S for the period $[S,T]$ is called the LIBOR spot rate and it is defined as:

$$L(S;S,T) = -\frac{P(S,T) - P(S,S)}{\tau(T,S)P(S,T)}$$

Note that the LIBOR rates are simply compounded spot rates. Since $L(S; S, T)$ is also a constant, this implies that it is also consistent with the price of zero coupon bonds, i.e.

$$L(S; S, T) = \frac{1 - P(S,T)}{\tau(S,T)P(S,T)} \tag{2.3.3}$$

$$\Rightarrow 1 + P(S,T) = L(S,T)[\tau(S,T)P(S,T)] \tag{2.3.4}$$

$$\Rightarrow P(S,T) = \frac{1}{1 + L(S,T)\tau(S,T)} \tag{2.3.5}$$

2.4 Yield Curve/The Zero-coupon Curve

A fundamental curve that can be obtained from the market data of interest rates is the zero-coupon curve at a given date t, also referred to as the *Yield curve*.

The yield curve describes the relationship between the returns on bonds (i.e the yield) with the same *credit-risk* but with different maturities. Formally we can define the yield curve at time t as:

Definition 2.4. *The yield curve at time t, is the graph of the function:*

$$T \mapsto P(t,T), \ \textit{where } T > t \tag{2.4.1}$$

which decreases with maturity T.

In the Swedish bond market, bonds with maturities less than a year are quoted as *simple rates* while those with maturities more than a year are quoted as *effective rates*.
To construct a yield curve for the whole tenor structure, we therefore need to convert the simple rates to effective rates in order to ensure that our yield curve is representing bonds with the same credit-risk. Theoretically, only effective rates for zero-coupon bonds guarantee a risk-free return if the bond is exercise at maturity.

Note that the term yield curve is also used to denote other curves that are deduced from quotes from the interest rate market. Thus there exist other representations of the yield curve different from the one given in Def: 2.4. In this thesis, we will use the above definition to represent our yields curves.

The slope of the yield curve varies over time and assumes different shapes and forms, which is a reflexion of the values on the short- and long yields. The yield-curve slope is sometimes used as a future GDP and inflation development indicator. Depending on its slope, we can classify the yield-curve as *normal*, *inverted* or *flat*. A more detailed discussion on the yield curve can be found in my bachelors thesis.

2.5 Coupon Bonds

As mentioned earlier, the zero coupon bond are not very well traded in the markets and markets that do trade in them, trade in zero coupon bonds with very short maturity. What is traded in the markets are the coupon bonds.

Given fixed points in time $T_0, T_1, \ldots, T_n$, *(known as a tenor structure)*, where T_0 is the issuing date of the bond and $T_1, \ldots, T_n$ are the coupon dates, coupon bonds can be classify into two main groups namely, fixed coupon bonds and floating rate bonds.

2.5.1 Fixed Coupon Bonds

This is the simplest coupon bond. It is a bond which for some intermediate points in time will provide the holder with predetermined payments called coupons.

Formally, fixed coupon bonds can be defined as:

Definition 2.5. *For fixed points in time $T_0, T_1, \ldots, T_n$, where T_0 is the issuing date of the bond, the owner of the fixed coupon bond receives the deterministic coupon C_i, for $i = 1, 2, \ldots, n-1$. At time T_n the owner receives the face value A.*

The coupon bonds can be replicated by holding a portfolio of zero coupon bonds with maturities T_i, for $i = 1, 2, \ldots, n$. That is we pay C_i zero coupon bonds of maturities T_i for $i = 1, 2, \ldots, n-1$ and $A + C_n$ bonds with maturity T_n. With this portfolio, for a time $t < T_1$, we can price the coupon bonds as:

$$p(t) = A \cdot P(t, T_n) + \sum_{i=1}^{n} C_i \cdot P(t, T_i)$$

The coupon bonds are quoted mostly in term of returns on the face value A over a given period $[T_{i-1}, T_i]$ and not in monetary terms. For example, given that the ith coupon has a return equal to r_i, this implies

$$c_i = r_i \cdot (T_i - T_{i-1}) \cdot A$$

If the interval lengths are equal (i.e. $T_i = T_0 + i\delta$) and the coupon rates for each interval is equal to a common rate r, then we have a **standardize coupon bond**. The price $p(t)$, for $t < T_1$, of a standardized coupon bond is given by:

$$p(t) = A \cdot \left(P(t, T_n) + r\delta \sum_{i=1}^{n} P(t, T_i) \right)$$

2.5.2 Floating rate Bonds

These are various coupon bonds for which the value of the coupon is not fixed at the time the bond is issued, but rather reset for every coupon period. Mostly, but not always, the resetting is determined by some financial benchmark. One of the simplest floating rate bonds is where the rate r_i is set to the **spot LIBOR** rate $L[T_i, T_{i-1}]$.

The floating rate bonds can be replicated by using a self-financial bond strategy, with an initial cost $P(t, T_{i-1})$ at time t and reinvesting the amount receive at time T_{i-1} in bonds that matures at time T_i. Thus the price p(t) for the floating rate bonds, given that the coupon dates are equally spaced (i.e. $T_i = T_0 + i\delta$) and assuming that the face value equals to one, for time $t < T_1$ is given by:

$$p(t) = P(t, T_n) + \sum_{i=1}^{n} [P(t, T_{i-1}) - P(t, T_i)] = P(t, T_0).$$

In particular, if $t = T_0$, then $p(T_0) = 1$.

2.6 Forward rates

The basic construction for interest rates is the forward rate and it is used to adjust for interest rate risk. Forward rates are interest rates that can be locked in today for an investment in a future time period, and are set consistently with the current yield of discount factors.

Generally speaking, the holder of a forward contract has the obligation to buy or sell a certain product at a future date for a given price. It is a bilateral contract that is traded over the counter and can be constructed in any way to suit the parties involved.

The forward contract is characterized by three time instants, namely $t < S < T$, where t is the contract date, S the date the contract is effective and T the exercise date.

A forward rate can be defined as

Definition 2.6. *Given three fixed time points $t < S < T$, a contract at time t which allows an investment of a unit amount of currency at time S, and gives a risk less deterministic rate of interest over the future interval $[S, T]$ is called **The forward rate**.*

Observe that the spot rates are forward rates where the time of contracting coincides with the start of the interval over which the interest rate is effective (i.e. $t = S$).

Another way to define the forward rates is through a *Forward Rate Agreement (FRA)*. By demanding that the FRA be priced fairly, we obtain the forward rates.

2.7 Arbitrage Theory

Basically, arbitrage in a given market implies the possibility to invest no amount of money at time t and receive a non negative amount at time $T > t$ with a positive probability. In other words, the law of one price for the same or identical commodities must hold if no arbitrage opportunity should exist in the market. This argument can be formalized by using probability theory reasoning to help us understand the economic interpretations of absence of arbitrage in a financial market. This connection will then lead us to **the martingale approach** of pricing financial derivatives and to the first- and second fundamental theories of mathematical finance which, for completeness will summarized at the end of this section.
First we need to define some basic economic concepts that will aid in defining mathematically the concept of arbitrage in a financial market.

To properly define the needed basic economic concepts, we need to specify a market in which the assets we are interested in pricing are traded. At the very basic level, we assume a probability space $(\Omega, \mathcal{F}_t, \mathcal{F}, \mathbb{P})$, where $\mathcal{F}_t$ is all the information gain from the past and just before time t (also called the filtration at time t), $\mathbb{P}$ is the historic probability measure (i.e. the probability taken on the observed data) and $\mathcal{F}$ is the total information set of the whole process.

Given a market on the above probability space, time $T > 0$, and $n + 1$ dividend paying assets, denoted by $S(t) = [S_0(t), S_1(t), \dots, S_n(t)]$, that are traded continuously from time zero to time T and that $S(t)$ are modeled as stochastic processes

$$dS_t = r_t S_t dt$$

adapted to the filtration $\mathcal{F}_t$, where $S_0(t) = 1$ and r_t is the instantaneous short-rate described earlier.
The adaptiveness of this filtration means that the information generated by the underlying assets are contained in the total information set of the process (i.e. $\mathcal{F}_t \in \mathcal{F}$). It is assumed that $S_0(t) > 0$ for all time $t > 0$.
Two important observations are to be noted here. Firstly, the dividend paying assets are assumed to be **semi-martingales** which is essential but not necessary to ensure that the normalized contingent claim is integrable for a fixed martingale measure and secondly, the underlying asset index by zero i.e. $S_0(t)$ is the risk free account which implies that the discounting factor $D(t, T)$ is given by $\frac{1}{S_0(t)}$.

With this market, we can now define a *portfolio*, also called a *trading strategy*, in continuous time as

Definition 2.7. *A portfolio is a collection of underlying assets in the market. This is denoted by $\boldsymbol{h}(t) = [h_0(t), \dots, h_n(t)]$, whose components $h_i(t)$ for $i = 0, \dots, n$ are locally bounded and predictable. The components of the trading strategy $h_i(t)$ represents the number of the underlying asset we hold at time t of asset **i**.*

Note that it is quite acceptable for the portfolio weights, i.e.$h_i's$, to attain positive, zero or negative values. If the portfolio weights are positive, then we have a **long position** on the underlying asset, whereas if the portfolio weights are negative, we say that we have a **short position** on the underlying. Also, the portfolio weights are not fixed constants but time varying and may change value over time. Hence the value of a portfolio at any given time instance t, is given by **the value process** of the trading strategy.

Definition 2.8. *The value process $V_{t\geq 0}$ associated with the portfolio $h(t)$ is defined as*

$$V^h(t) = \sum_{i=0}^{n} h_i(t)S_i(t) = h_i(t) \cdot S_i(t) \text{ for } 0 < t < T$$

Definition 2.9. *The gain process G_t associated with the portfolio $h(t)$ is given by*

$$G^h(t) = \int_0^t h_i(u)dS_i(u) = \sum_{i=0}^{n} \int_0^t h_i(u)dS_i(u)$$

If no additional cash flow is allowed after the trading strategy has been created at time t and all the changes in the dynamics of the portfolio comes about through changes in the assets already hold in the portfolio, then the trading strategy is said to be **self-financing**. A self-financing portfolio can be defined in three different ways as below:

Definition 2.10. *A portfolio is called a self-financing portfolio if its value changes only due to changes in the asset prices. This can be express as*

$$dV^h(t) = \sum_{i=0}^{n} h_i(t)dS_i(t) = h_i(t) \cdot dS_i(t) \text{ for } 0 < t < T$$

A self-financing portfolio can also be expressed using the gain process associated with the portfolio as

Definition 2.11. *A portfolio $h(t)$ is self-financing if the value process $V(t) \geq 0$ are such that*

$$V^h(t) = V^h(0) + G^h(t) \text{ for } 0 < t < T$$

A relation similar to the self-financing portfolio expressed in terms of the gain process also holds when asset prices are all expressed in terms of the bank account value. Thus

Definition 2.12. *Let $h(t)$ be a portfolio, then $h(t)$ is self-financing if and only if*

$$D(0,t)V^h(t) = V^h(0) + \int_0^t h_u d(D(0,u)S_u)$$

The relation between the self-financing portfolio expressed in terms of the risk-free asset was proved by Harrison and Pliska (1981).

Since the trading strategy $h(t)$ is given in terms of the absolute weights of each of the underlying assets in the portfolio, **relative weights** are sometimes necessary to ease computations. The relative value $u_i(t)$ of the assets $S_i(t)$, for $i = 0, \ldots, n$, is the portfolio w.r.t the total value of the absolute portfolio $V(t)$. The value process of a self financing portfolio can be used to obtained the relative portfolio. The relative portfolio weights are defined as

Definition 2.13. *For a given trading strategy, the relative portfolio weights u_i, which in general can be both $u_i \leq 0$ and $u_i \geq 0$, is given by the fraction of the total value from asset i. This can be expressed as*

$$u_i(t) = \frac{h_i(t) \cdot S_i(t)}{V^h(t)}, \text{ where } i = 0, \ldots, n \text{ and } \sum_{i=0}^{n} u_i(t) = 1 \tag{2.7.1}$$

Using the relative portfolio, the dynamics of the self-financing portfolio can be re-written as

$$\begin{aligned} dV^h(t) &= \sum_{i=0}^{n} h_i(t) dS_i(t) \\ &= V^h(t) \sum_{i=0}^{n} \frac{h_i(t) S_i(t)}{V^h(t)} \cdot \frac{dS_i(t)}{S_i(t)} \end{aligned}$$

Thus

$$dV^h(t) = V^h(t) \sum_{i=0}^{n} u_i(t) \cdot \frac{dS_i(t)}{S_i(t)}$$

As mentioned earlier, an arbitrage possibility on a financial market refers to the possibility of making a positive amount of money without taking any investment risk. Formally, an arbitrage opportunity can be defined as:

Definition 2.14. *An arbitrage possibility on a financial market, for every $t > 0$, is a self-financing portfolio $h(t)$ such that:*

$$V^h(0) = 0$$

$$P(V^h(t) \geq 0) = 1$$

$$P(V^h(t) \geq 0) > 0.$$

A market where arbitrage possibility occurs is called an inefficient market which goes against our main assumption of a financial market, that is **efficient**. In other to ascertain efficiency in a financial market, it is required that the value process of the dynamics of a portfolio to be **locally risk-free** and have a return equal to the return on the bank account with probability one. In other wards in an arbitrage free market there can exist only one short rate of interest.

locally risk-free implies that at any given time there exist not stochasticity in the value process.

The above argument can be expressed as follows

Definition 2.15. *For every (locally) risk-free self-financed trading strategy $h(t)$ of the form*

$$dV^h(t) = k(t)V^h(t)dt \qquad \text{where } k(t) \text{ is any } \mathcal{F}_t\text{-adapted process} \tag{2.7.2}$$

it must hold that the probability of the adapted process $k(t)$ is equals to the risk-free interest rate r(t) is equals to one for almost all times t to avoid arbitrage opportunity. That is

$$P(k(t) = r(t) \text{ for almost all times } t\ ) = 1 \text{ to attain market efficiency.} \tag{2.7.3}$$

Market efficiency can also be stated by using the connection between the economic concept of absence of arbitrage and the mathematical property of existence of a probability measure (i.e. $\mu : \mu(\Omega) = 1$) called the equivalent martingale measure (EMM). The existence of an EMM in a market tells us that the market is arbitrage free. This is referred to as the **first fundamental theory in mathematical finance**. An equivalent martingale measure is defined below:

Definition 2.16. *An equivalent martingale measure denoted by $\mathbb{Q}$ is a probability measure on the space $(\Omega, \mathcal{F})$ such that:*

1. *The measures $\mathbb{P}$ and $\mathbb{Q}$ are equivalent if and only if $\mathbb{P}(A) = 0$ is equivalent to $\mathbb{Q}(A) = 0$, for every $A \in \mathcal{F}$.*
2. *The discounted price processes $\frac{S_i}{S_0}$ are $\mathbb{Q}$-martingales, for all $i \in (0, \dots, n)$. That is*
$$\mathbb{E}^{\mathbb{Q}}[(D(0,t)S_i(t)|\mathcal{F}_u] = D(0,u)S_i(u)$$
for all $i = 1, \dots, n$ and all $0 \le u \le t \le T$, where $\mathbb{E}^{\mathbb{Q}}$ denote the expectation under $\mathbb{Q}$.
3. *The Randon-Nikodym derivative $\frac{d\mathbb{Q}}{d\mathbb{P}}$ is square integrable with respect to $\mathbb{P}$*

If the martingale measure defined above is unique then we say that the market is **complete**. This statement is often referred to as the second fundamental theory of mathematical probability. Note that the Randon-Nikodym derivative is also referred to as the **likelihood ratio (L)** between the equivalent measure and the historic measure and it is generally used to enable movement to and from these two measures. That the Randon-Nikodym derivative $L = \frac{d\mathbb{Q}}{d\mathbb{P}}$ is square integrable implies that

$$\int_{-\infty}^{\infty} L d\mathbb{P} \text{ is finite}$$

Now that we have formalized the concept of absence of arbitrage and have define a market that is efficient, we now need to consider the assets that are traded in this market. These traded assets are referred to as financial derivative or **contingent claims**. Our objective is to determine the prices of these derivatives. Contingent claims are like insurance contracts on which it is stated that the holder of the contract will receive a deterministic amount of money, called the **payoff**, at a specified date in the future called the exercise date. Note that the payoff can be both positive, negative or zero. Contingent claims are completely defined in terms of the underlying assets, which in our case are the interest rates but they can also be written on other underlying such as stocks, bond, or other financial assets.

Definition 2.17. *Given the financial market defined above with vector price process* $S = [S_0, \ldots, S_n]^T$. *A contingent claim with exercise date* ***T*** *is any stochastic variable* $\mathcal{X}$ *such that* $\mathcal{X} \in \mathcal{F}_T^S$. *In this context,* $\mathcal{X} \in \mathcal{F}_T^S$ *means that it is possible to derive the price of the contingent claim at time T. Below is a formal definition of a contingent claim.*

A contingent claim, also referred to as a T-claim , is called a **simple claim** if it is of the form

$\mathcal{X} = \Phi(S(T))$. The function Φ is called the **contract function**

A contingent claim $\mathcal{X}$ is said to be **attainable** if there is exists a self-financing portfolio **h** such that the value of the portfolio at the time of maturity **T** is equals to that of the contingent claim. In other words, a claim is attainable if

$$V^h(t) = \mathcal{X}.$$

Note that attainable claims are also referred to as **hedgable** claims. If a contingent claim is attainable and there exists a self-financing portfolio h, then it implies that the claim can be traded in the financial markets.

If all contingent claims $\mathcal{X}$ on a financial market are attainable, then the market is called **complete**. Recall that completeness was defined above as equivalent to the existence of a unique equivalent measure.

Thus if we assumed that the market is efficient, the price of the contingent claim at time t, denoted by $\Pi(t, \mathcal{X})$ can be determine in two different ways.
Firstly we could demand consistency of the price of the underlying with the price of the contingent claim. In other words, to avoid arbitrage opportunities, the extended market given by $(\Pi(;\mathcal{X}), S_0, \ldots, S_n)$ must be arbitrage free.
Secondly, if given an attainable claim with a self-financing portfolio then the value process at time t generated by **h** for $0 < t < T$ must be equal to the price of the contingent claim at time t, if arbitrage opportunity is to be avoided, i.e. to attain an efficient financial market, then

$$\Pi(t, \mathcal{X}) = V^h(t) \text{ for } 0 < t < T$$

Determining the price of a contingent claim by demanding price consistency of the claim with that of the underlying implies the existence of a martingale measure

$\mathbb{Q}$ for the extended market $(\Pi(;\mathcal{X}), S_0, \ldots, S_n)$. This leads us to the martingale approach for pricing derivative, which is a very powerful and effective pricing mechanism. Thus applying the definition of a martingale measure under $\mathbb{Q}$ we obtain the **general pricing formula** for $\mathbb{Q}$. The general pricing formula is define below

Definition 2.18. *The arbitrage free price process for the T-claim $\mathcal{X}$ is given by*

$$\Pi(t, \mathcal{X}) = S_0(t)\mathbb{E}^{\mathbb{Q}}\left[\frac{\mathcal{X}}{S_0(T)}|\mathcal{F}_t\right] \tag{2.7.4}$$

where $\mathbb{Q}$ is the martingale measure for the market $S_0, \ldots, S_n$, with S_0 as the numeraire. Note that the martingale measure is not unique and different choices of $\mathbb{Q}$ will give raise to different price processes.

If it is assumed that $S_0(t)$ in the general pricing formula is the risk-less bank account, then $S_0(t)$ can be re-written as

$$S_0(t) = S_0(0) \cdot \exp^{\int_0^t r(s)ds} \qquad \text{where r is the short rate} \tag{2.7.5}$$

This leads us to the well celebrated formula in financial mathematics called the risk neutral valuation formula **RNVF**. Note that if we even go about determining the price of a contingent claim by the second method mentioned above, we will come to the same conclusion i.e. the RNVF.

2.7.1 Risk Neutral Valuation Formula

The RNVF is a martingale approach to derivative pricing, where the price of a claim is determine by taking the expectation, under the $\mathbb{Q}$-martingale measure, of the discounted T-claim with the money account as a numeraire given an adapted filtration, i.e.

$$\Pi(t, \mathcal{X}) = \mathbb{E}^{\mathbb{Q}}\left[\frac{\mathcal{X}B(t)}{B(T)}|\mathcal{F}_t\right] = \mathbb{E}^{\mathbb{Q}}\left[\exp^{\int_t^T r(s)ds} \cdot \mathcal{X}|\mathcal{F}_t\right]$$

The above result implies that the discounted price process is a $\mathbb{Q}$-martingale measure and the existence of a $\mathbb{Q}$-martingale measure implies absence of arbitrage in our market. This again shows that the martingale property can be interpreted in economic terms as been the same as the existence of a self-financing portfolio.
Note that in the above risk neutral valuation formula, the bank account was used to normalized the underlying assets and thereby relating the price of a contingent claim for different time points. We thus say that the bank account is a **numeraire**. Though the elegance of the RNVF is undoubted, the expectation under the $\mathbb{Q}$-martingale measure is not always easily solved. This is indeed the case when the underlying assets of interest are stochastic, for example interest rates.

Therefore, given a stochastic underlying, the bank account is no longer suitable as a discounting factor because the joint distribution $(B(T), \Phi(S))$, which is needed to solve the expectation under the $\mathbb{Q}$-measure of the RNVF is laborious to compute since it will involve the computation of a double integral. To solved this problem, we need to change the numeraire. Firstly, we properly define a numeraire and its characteristics.

2.7.2 Numeraires

A numeraire can be seen as a price-unit over time, because it relates prices at different time points. Choosing a numeraire implies that the relative prices are considered instead of the securities themselves. Thus it must be strictly positive, which allows it to give some value to each time point. This implies that any other non positive non-dividend-paying asset (that is not necessarily the bank account) can also be used as a numeraire. Formally it is defined as follows:

Definition 2.19. *A Numeraire N_t for the market $\boldsymbol{S} = (S_0, \ldots, S_n)$ is any positive process that is adapted to the filtration generated by the assets in the market (i.e. ${\mathcal{F}_t}^S$), which is of the form*

$$N_t = N_0 + \int_0^t \alpha_u d\boldsymbol{S}(u) = \alpha_t^T \boldsymbol{S}(t) \tag{2.7.6}$$

where α_t^T is a self-financing strategy that is predictable given the filtration $\mathcal{F}_t$.

2.7.3 Change of Numeraire

As mentioned earlier, the equivalent martingale measure $\mathbb{Q}$ gives us all the necessary information we need about our market in terms of arbitrage and completeness. The conditional expectation of a claim under the $\mathbb{Q}$-equivalent measure also determines all its arbitrage free prices. Therefore what is needed to price contingent claims is the $\mathbb{Q}$-martingale measure. A fundamental question is how do we change to the $\mathbb{Q}$-measure so as to attained an arbitrage free price of a general claim? The answer to this fundamental question is the **likelihood ratios**.

The likelihood ratio

The likelihood ratio, also referred to as the Randon-Nikodym derivative gives us a relation between the $\mathbb{P}$-measure and the $\mathbb{Q}$-measure and it is formally defined as follows:

Definition 2.20. *Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and an absolutely continuous probability measure $\mathbb{Q}$. Then there is a stochastic variable $\boldsymbol{L}$ with*

$$L \geq 0 \text{ (almost surely) and } \mathbb{E}^{\mathbb{P}}[L] = 1 \text{ such that} \tag{2.7.7}$$

$$\mathbb{Q}(A) = \mathbb{E}^{\mathbb{P}}[I_A L] \text{ for all } A \in \mathcal{F} \tag{2.7.8}$$

$$\mathbb{E}^{\mathbb{Q}}[X] = \mathbb{E}^{\mathbb{P}}[XL] \text{ for all } \mathcal{F}\text{-measurable } X. \tag{2.7.9}$$

L is called a Likelihood Ratio (LR) for the change of measure. Furthermore, if $\mathbb{P}$ and $\mathbb{Q}$ are equivalent, then the LR L above fulfills

$$L \geq 0 \text{ (almost surely) and} \tag{2.7.10}$$

$$\mathbb{P}(A) = \mathbb{E}^{\mathbb{Q}}[I_A L^{-1}] \text{ for all } A \in \mathcal{F}. \tag{2.7.11}$$

The following proposition by Geman et al. (1995) provides a fundamental tool for pricing contingent claims for models with stochastic rates, and a generalization of the risk neutral valuation formula to any numeraire.

proposition 2.1. *Assume there exists a numeraire $\boldsymbol{N}$ and a probability measure $\mathbb{Q}^N$, equivalent to the initial measure $\mathbb{Q}_{\nvdash}$, such that the price of any traded asset $\mathbb{X}$ (without intermediate payments) relative to $\boldsymbol{N}$ is a martingale under $\mathbb{Q}^N$, i.e.*

$$\frac{X_t}{N_t} = E^N\left[\frac{X_T}{N_T}|\mathcal{F}_t\right] \; 0 \leq t \leq T. \tag{2.7.12}$$

Let $\mathbb{U}$ be an arbitrary numeraire. Then there exists a probability measure $\mathbb{Q}^U$, equivalent to the initial $\mathbb{Q}_0$, such that the price of any attainable claim $\mathbb{Y}$ normalized by $\mathbb{U}$ is a martingale under $\mathbb{Q}^U$, i.e.

$$\frac{Y_t}{U_t} = E^U\left[\frac{Y_T}{U_T}|\mathcal{F}_t\right] \; 0 \leq t \leq T. \tag{2.7.13}$$

Moreover, the Radon-Nikodym derivative defining the measure $\mathbb{Q}^U$ is given by

$$\frac{\partial Q^U}{\partial Q^N} = \frac{U_T N_0}{U_0 N_T}. \tag{2.7.14}$$

Chapter 3

DYNAMIC MODELS

3.1 Vector Autoregressive Process (VAR)

We begin this chapter by discussing an extension of the univariate *AR-process* called the **Vector Autoregressive process (VAR)** that is used for modeling univariate time-series data, The VAR model is one of the most popular and easy to used model for describing dynamic behaviors, as well as, providing meaningful forecast of economic and financial time series. Formally, a VAR-process is defined as

Definition 3.1. *Let* $X_t = (x_{1t}, x_{2t}, \ldots, x_{nt})'$ *be a set of* $(n \times 1)$ *observations. Then the process*

$$X_t + \Phi_1 X_{t-1} + \Phi_2 X_{t-2} \ldots + \Phi_p X_{t-p} = \epsilon_t \tag{3.1.1}$$

where Φ_i are $(n \times n)$ coefficient matrices and ϵ_t is an $(n \times 1)$ zero mean white noise vector process with constant covariance matrix Σ_ϵ, is called an stationary Vector Autoregressive Process of order p, i.e. **VAR(p)**.

It is important to observed that the $VAR(p)$ process can sometimes be too restrictive to properly represent the main characteristics of the data. Therefore, additional deterministic terms (such as linear trends) might be needed to represent the data. Moreover, external variables may also be added to the $VAR(\cdot)$ process for the data representation to be proper. Then equation 3.1.1 is generalized as

$$X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \ldots + \Phi_p X_{t-p} + \Pi D_t + G Y_t + \epsilon_t \tag{3.1.2}$$

where D_t is an $(l \times 1)$ deterministic matrix, Y_t represents the $(m \times 1)$ matrix of external variables and Π and G are parameter matrices.

3.2 State Space Modeling

As in most modeling procedure, the main aim is to find an appropriate way to model the relationship between input and output signals of a system. These types of models focuses only on the *external description of the system.* To gain some insight on the internal state of the system under study, we resort to state space modeling. This is obtained by defining a state vector such that the dynamics of the system can be described by a *Markov process.*

A state space model is formulated in discrete time by using a (multivariate) difference equation or a (multivariate) differential equation in continuous time, describing the dynamics of the state vector X_t, and a static relation between the state vector and the (multivariate) observation Y_t. Thus a linear state space model consist of two sets of equations, *the system equation*

$$X_t = A_t X_{t-1} + B_t u_{t-1} + e_{1,t} \tag{3.2.1}$$

and the *observation equation*

$$Y_t = C X_t + e_{2,t} \tag{3.2.2}$$

where X_t is the N-dimensional random state vector that is not directly observable. u_t is a deterministic input vector and Y_t is a vector of observable stochastic output, and A_t, B_t, and C_t are deterministic matrices in which the parameters are embedded. Finally the processes $e_{1,t}$ and $e_{2,t}$ are uncorrelated white noise processes. Thus the system equation equation describes the evolution of the system states whereas the observation equation describes what can be directly measured.

For linear time systems, in which the system noise $e_{1,t}$ and the measurement noise $e_{2,t}$ are taken to be Gaussian with zero mean, the *Kalman filter* is used to estimate the hidden state vector and also for providing predictions. The Kalman filter is described below.

3.2.1 The Kalman Filter

For linear dynamic systems, the Kalman filter provides the optimal prediction and reconstruction of the latent state vector. The foundation of the Kalman filter is based of the *linear projection theorem* that is stated below:

Theorem 3.1. *Let $\boldsymbol{Y} = (Y_1, \ldots, Y_m)^T$ and $\boldsymbol{X} = (X_1, \ldots, X_m)^T$ be random vectors, and let the $(m+n)$-dimensional vector $(\boldsymbol{Y}, \boldsymbol{X})^T$ have the mean*

$$\begin{pmatrix} \mu_Y \\ \mu_X \end{pmatrix} \text{ and covariance } \begin{pmatrix} \Sigma_{YY} & \Sigma_{YX} \\ \Sigma_{XY} & \Sigma_{XX} \end{pmatrix}$$

Define the linear projection of $\boldsymbol{Y}$ on $\boldsymbol{X}$

$$E[\boldsymbol{Y}|\boldsymbol{X}] = \boldsymbol{a} + \boldsymbol{B}\boldsymbol{X} \tag{3.2.3}$$

Then the projection and the variance of the projection error is given by

$$E(\boldsymbol{Y}|\boldsymbol{X}) = \mu_Y + \Sigma_{YX}\Sigma_{XX}^{-1}(X - \mu_X) \tag{3.2.4}$$

$$Var(E(\boldsymbol{Y}|\boldsymbol{X})) = \Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{YX}^T \tag{3.2.5}$$

Finally, the projection error, $\boldsymbol{Y} - E(\boldsymbol{Y}|\boldsymbol{X})$, and $\boldsymbol{X}$ are uncorrelated, i.e.

$$C(\boldsymbol{Y} - E(\boldsymbol{Y}|\boldsymbol{X}), \boldsymbol{X}) = 0 \tag{3.2.6}$$

From the above proposition, the Kalman filter equations for reconstructing, updating and predicting the latent states are generated. Since $e_{1,t}$ and $e_{2,t}$ in the state space model are assumed to be normally distributed, then $X_t|Y_t$ is also normally distributed and are thus completely characterized by its mean

$$\hat{X}_{t|t} = E(X_t|Y_t) \tag{3.2.7}$$

and variance

$$\Sigma^{xx}_{t|t} = Var(X_t|Y_t) \tag{3.2.8}$$

The optimal linear reconstruction of the states, which in linear time invariant systems is given by the Kalman filter is obtained from

$$\hat{X}_{t|t} = \hat{X}_{t|t-1} + K_t(Y_t - C\hat{X}_{t|t-1}) \tag{3.2.9}$$

and the variance of the reconstruction is given by

$$\Sigma^{xx}_{t|t} = \Sigma^{xx}_{t|t-1} - K_t \Sigma^{yy}_{t|t-1} K_t^T \tag{3.2.10}$$

$$= \Sigma^{xx}_{t|t-1} - K_t C \Sigma^{xx}_{t|t-1} \tag{3.2.11}$$

where the **Kalman gain** at time t K_t, is given by

$$K_t = \Sigma^{xx}_{t|t-1} C_t^T [C_t \Sigma^{xx}_{t|t-1} C_t^T + \Sigma_{2,t}]^{-1} \tag{3.2.12}$$

Now the one-step predictions is given by:

$$X_{t+1|t} = A_t \hat{X}_{t|t} + B_t u_t \tag{3.2.13}$$

$$\Sigma^{xx}_{t+1|t} = A_t \Sigma^{xx}_{t|t} A_t^T + \Sigma_{1,t} \tag{3.2.14}$$

$$Y_{t+1|t} = C_t \hat{X}_{t+1|t} \tag{3.2.15}$$

As can be observed, the Kalman filter is a recursive filter and thus initial conditions

$$\hat{X}_{1|0} = E(X) = \mu_0 \tag{3.2.16}$$

$$\Sigma^{xx}_{1|0} = Var(X_t) = \Sigma_0 \tag{3.2.17}$$

are needed. Finally, the innovation (i.e. the measurement error) is given by

$$\hat{Y}_{t+1|t} = Y_{t+1} - \hat{Y}_{t+1|t} \tag{3.2.18}$$

and its variance R_{t+1} is computed as

$$R_{t+1} = Var(\hat{Y}_{t+1|t}) \tag{3.2.19}$$

$$= \Sigma^{yy}_{t+1|t} \tag{3.2.20}$$

$$= C_t \Sigma^{xx}_{t+1|t} C_t^T + \Sigma_{2,t} \tag{3.2.21}$$

If the assumptions of normality and linearity are no longer valid, then the Kalman filter will no longer be valid for parameter estimations and state prediction. We therefore have to resort to non-linear state space modeling, which in a sense are approximative filters. The *Extended Kalman Filter* **EKF** and *Particle filter* are some of these non-linear filters that can be used for parameter estimations.

Chapter 4

SHORT RATES

4.1 Term Structure

The term structure of interest rates describes the relationship between interest rates and time to maturity. The standard way of measuring the term structure of interest rates is by means of the spot rate curve, or the yield curve, on zero coupon bonds. Note that the yield to maturity and the spot rate on a zero coupon bond are the same. The zero coupon bonds are used here to eliminate the *'coupon- effect'*, which refers to the situation were two identical bonds bear different coupon-rates and have different yield to maturity.

On the other hand, using the zero-coupon yield limits us to bonds with maturities of at most twelve months. Bonds with maturities less than twelve months are referred to as treasury bills). Thus longer maturity zero-coupon bonds need to be derived from coupon-bearing bonds. In practice, the entire term structure of interest rates is not directly observed and therefore need to be estimated.

Generally speaking, the term structure estimation methods are designed for the purpose of approximating one of three equivalent representations of the yield curve i.e *the forward rate curve, the discount curve* and *the spot rate curve*. If one of these representations are obtained, the others can be derived from it.

An aspect of yields is that they are not normally distributed. As mentioned earlier, bonds with different maturities are traded at the same time. Bonds with long maturities are risky when held over short period and the holder (a risk-averse investor) will demands compensation for bearing such risk. Thus the markets in which these bonds are traded will not be free from arbitrage unless long yields are risk-adjusted expectation of average future short rates.

One way to take into account the risk on long term maturity bonds is to model the continuously compounded spot rate $R(t, T)$ as an affine function in the short rate $r(t)$ as

$$R(t, T) = a(t, T) + b(t, T)r(t) \tag{4.1.1}$$

where a and b are deterministic functions of time. A model satisfying the relationship above is called an *affine term structure model*

4.1.1 Affine Term Structure (ATS)

Affine term structure captures the risk for long maturity bonds in an arbitrage free financial market. Affine term structure models are arbitrage free models in which bond prices are affine in mature, i.e. they can be expressed as a constant plus a linear term. ATS is define as

Definition 4.1. *If the term structure*

$$\{P(t,T) : T \geq 0\}$$

is given by a function $P(t,T) = F^T(t, r(t))$ *where*

$$F^T(t, r(t)) = e^{(A(t,T)-B(t,T)r(t))}$$

and A *and* B *are deterministic functions, (independent of* $r(t)$*), the term structure* $\{P(t,T) : T \geq 0\}$ *is said to be affine.*

Observe that the ZCB price $P(t,T)$ is expressed as

$$P(t,T) = e^{A(t,T)-B(t,T)r(t)} \tag{4.1.2}$$

if the relationship that models the continuously compounded spot rate $R(t,T)$ as an affine function in the short rate $r(t)$ (i.e equation 4.1.1) holds. Therefore, we have that

$$a(t,T) = \frac{-(lnA(t,T))}{(T-t)} \tag{4.1.3}$$

$$b(t,T) = \frac{B(t,T)}{(T-t)} \tag{4.1.4}$$

The above definition is true for all models of the short rate $r(t)$ with dynamics given by:

$$dr(t) = [\alpha(t)r(t) + \beta(t)]dt + \sqrt{\gamma(t)r(t) + \delta t}dW(t)$$

From which we identify the deterministic functions in the affine term structure for the ZCB-price as:

- $A_t(t,T) = \beta(t)B(t,T) - \frac{\gamma(t)}{2}B^2(t,T); A(T,T) = 0$
- $B_t(t,T) = -\alpha(t)B(t,T) + \frac{\gamma(t)}{2}B^2(t,T) - 1; B(T,T) = 0$

Observed that if $\gamma = 0$, it is then possible to explicitly solve the equations, thereby giving us

$$B(t,T) = \int_t^T e^{\int_t^S \alpha(u)du} dS \tag{4.1.5}$$

$$A(t,T) = \int_t^T \frac{\delta(S)B^2(S,T)}{2} - \int_t^T \beta(S)B(S,T)dS \tag{4.1.6}$$

The affine term representation will be very crucial when price short-rates.

4.2 Short interest rate models

Interest rates model theory originally assumes a specific one dimensional dynamics for the instantaneous spot rate process $r(t)$. These types of dynamics are also referred to as *one factor short rate model.* The one dimensional dynamics ensures that all the interest rate derivatives traded can easily be priced by using the arbitrage theory reasoning discussed earlier. That is to say, given a risk-neutral measure, the arbitrage free price of a simple claim with payoff X_T at time t is given by the $RNVF$, as the expectation of a function of the process $r(t)$. As mentioned earlier, if a risk neutral measure exist, then the arbitrage-free price at time t of a simple claim with payoff X_T is given by

$$X_t = E_t\left[e^{(-\int_t^T r(u)du)} X_T\right] \tag{4.2.1}$$

Again, the zero coupon-bond price at time t for the maturity T (i.e. P(t,T)) is characterized by a unit amount of currency available at time T obtained, under the risk neutral measure, by

$$P(t,T) = E_t\left[e^{(-\int_t^T r(u)du)}\right] \tag{4.2.2}$$

Indeed, if the distribution of $\left\{e^{-\int_t^T r(s)ds}\right\}$ conditional on the filtration F_t, is attainable in terms of a chosen dynamics for short-rate, then bond prices can be computed. Since bond prices, especially ZCB, are the basic building blocks in almost all traded assets with interest rates as underlying, knowing how to attain their prices implies that we also know the rates for all the other underlying assets. Thus, the whole zero-coupon curve is characterized in terms of the distributional properties of $r(t)$.

In the early attempts to model interest rates (e.g. Vasicek (1977) which will be describe later), it was very natural to model the short rate by assuming a stochastic process on the short rate $r(t)$, instead of the fixed rate r. This was done to mimic the construction of a locally risk-free portfolio, as in the **Black-Scholes** formula for pricing European options. In its general from, the short-rate is model as

$$dr(t) = \mu(t, r(t))dt + \sigma(t, r(t))dW^{\mathbb{P}}(t) \tag{4.2.3}$$

where μ and σ are deterministic functions of t and $r(t)$ and $W^{\mathbb{P}}$ is a Brownian motion on the objective measure $\mathbb{P}$. Using the short rates dynamics in its general form as underlying assets of a contingent claim, even though the short rate in not directly observed in the market, it is possible to apply the same procedure as when pricing simple claims in the *Black-Scholes* market, to obtain formulas for pricing the short-rate. Using the *Black-Scholes* approach leads us to the *Partial Differential Equation* PDF formula and the *Risk Neutral Valuation Formula* RNVF for the short rates, which are defined below.

Firstly, **the market price of risk**, which is a fundamental equation in the pricing of short-rates, will be stated as a proposition

proposition 4.1. *If the zero-coupon bond (ZCB) market is free of arbitrage then there exist a stochastic process $\lambda(t)$ such that if*

$$dP(t,T) = \mu^T(t,r(t))dt + \sigma^T(t,r(t))dW^{\mathbb{P}}(t),$$

then the relation

$$\frac{\mu^T(t,r(t)) - r(t)P(t,T)}{\sigma^T(t,r(t))} = \lambda(t)$$

holds for any finite maturity T. λ is called ***the market price of risk***

It is important to observed that the *market price of risk* does not refer to price as used in everyday language, but tells us how much we gain by investing in the risky assets instead of investing only in the risk-free account. Thus λ is sometimes referred to as *the excess return* with respect to a risk-free investment per unit of risk.

Another important observation is that the *market price of risk*, which does not depend on the maturity date T, is the quantity that connects the objective measure $\mathbb{P}$ and the risk-neutral measure $\mathbb{Q}$. This connection is obtained by applying the fact that there exist a measure $\mathbb{Q} \sim \mathbb{P}$ and is defined by the *Randon-Nikodym Derivative*

$$\frac{\partial Q}{\partial P}\Big|_{\mathcal{F}_t} = \exp\big(-\frac{1}{2}\int_0^t \lambda^2(u)du - \int_0^t \lambda(u)dW^{\mathbb{P}}(u)\big)$$

where $\mathcal{F}_s$ is the information generated by r up to time t. Thus applying this change of measure to the short rate dynamics given in the objective measures $\mathbb{P}$(as in equation[4.2.3]), generates the short-rate dynamics, in the risk-neutral measure $\mathbb{Q}$, as

$$dr(t) = [\mu(t,r(t)) - \lambda(t)\sigma(t,r(t))]dt + \sigma(t,r(t))dW(t) \tag{4.2.4}$$

where $W(t) = W^{\mathbb{P}}(t) + \int_0^t \lambda(u)du$ is a Brownian motion in the risk neutral measure $\mathbb{Q}$.

Note that our main objective is to attain pricing formulas for interest rates and we are not directly interested in knowing the dynamics of the short rate in the objective measures $\mathbb{P}$. Therefore we can directly model short rate dynamics in the risk-neutral measures $\mathbb{Q}$, given the market price of risk is specified. It is the parameter values of the short rate models, in the risk neutral measure, that are desired to determine derivative prices. Hence all the short rate models that are considered from now on will be priced under the risk neutral measure $\mathbb{Q}$.

Two very common short rate models that are widely used are the **Vasicek model**, the **Cox, Ingersoll and Ross (CIR) model**. These models are also referred to *endogenous term structure models* because of the fact that the rates are outputs rather than an inputs to the model.

4.2.1 Term Structure Equation (TSE)

If $F^T(t,r(t))$ is the price of a ZCB, with dynamics

$$dF^T[t,r(t)] = \mu^T[t,r(t)]F^T[t,r(t)]dt + \sigma^T[t,r(t)]F^T[t,r(t)]dW_t$$

where the processes μ^T and σ^T (which are the same processes used to compute the market price of risk) are functions of the ZCB-price and are also the same parameters defining the short rate model. Solving the equation for the market price of risk leads us to the *term structure equation* which is describe below.

Theorem 4.1. *In an arbitrage free market, the price of the simple claim is given by $P(t,T) = F^T(t, r(t))$, where the function $F^T(t, r(t))$ is twice differentiable and satisfies the partial differential equation PDE*

$$\begin{cases} F_t^T + (\mu - \lambda\sigma)F_r^T + \sigma^2/2F_{rr}^T - rF^T = 0 \\ F_t^T(T, r) = 1. \end{cases} \tag{4.2.5}$$

For a general short rate derivative with payoff function $X(r(T))$, the partial differential equation that gives the price of a simple claim is given by the theorem below:

Theorem 4.2. *In an arbitrage free ZCB market a derivative given by the payoff function $X(r(T))$ has a pricing formula F(t,r) that satisfies the PDE*

$$\begin{cases} F_t + (\mu - \lambda\sigma)F_r + \sigma^2/2F_{rr} - rF = 0 \\ F_t^T(T, r) = X(r). \end{cases} \tag{4.2.6}$$

and the $RNVF$ for the short rate, under the $\mathbb{Q} - measure$, is given as stated in the theorem below:

Theorem 4.3. *The derivative given by the payoff function $X(r(T))$ has the following pricing formula for the general short rate model:*

$$F(t, r) = E^{\mathbb{Q}}\left[e^{\int_t^T r(s)ds} X(r(T)) \middle| F_t \right] \tag{4.2.7}$$

where r(s) is defined by the stochastic differential equation (SDE) below

$$\begin{cases} dr(s) = (\mu(s, r(s)) - \sigma(s, r(s)\lambda(s))ds + \sigma(s, r(s))dW(s) \\ r(t) = r. \end{cases} \tag{4.2.8}$$

Before we describe some of the most common short-rate models that are in used and state some of their advantages and disadvantages, there are fundamental questions that need to be asked and answered in other to choose the short rate model that will suit your modeling purpose. According to *Damiano Brigo and Fabio Mercurio(2006)*, the questions that are necessary for understanding the theoretical and practical implications of any interest rate model are the following:

- Does the dynamics imply positive rates, i.e., $r(t) > 0$ almost surely for each time t?
- What distribution does the dynamics imply for the short rate r?

- Are the bond prices $P(t,T) = E_t\left\{e^{-\int_t^T r(u)du}\right\}$ (and therefore spot rates, forward rates and swap rates) explicitly computable from the dynamics?
- Is the model mean reverting, in the sense that the expected value of the short rate tends to a constant value as time grows towards infinity, while its variance does not explode?
- How do the volatility structures implied by the model look like?
- Does the model allow for explicit short-rate dynamics under the forward measures?
- How suited is the model for Monte Carlo simulation?
- How suited is the model for building recombining lattices?
- Does the chosen dynamics allow for historical estimation techniques to be used for parameter estimation purposes?

4.2.2 The Merton model

The Merton model for the short-rate is a very simple model that assumes that the short-rate dynamics is a **Brownian Motion** with drift, i.e.

$$dr(t) = \mu dt + \sigma dW(t) \tag{4.2.9}$$

The short-rates in the *Merton Model* are normally distributed with mean and variance,

$$E[r(t)] = \mu t \tag{4.2.10}$$

$$Var[r(t)] = \sigma^2 t \tag{4.2.11}$$

Observed that the *Merton Model* is not mean reverting and its variance grows linearly with time. The Morton Model is not a realistic model because it is possible to obtain negative short-rates. However its simplicity gives the model its charm and makes it very easy to compute with.

The price of a zero-coupon bond in the Merton model, can be obtain through the distribution of

$$\int_t^T r(s)ds = \int_t^T (r_t + \mu(s-t) + \sigma W_{s-t})ds$$

The above integral of the short-rate $r(t)$ is normally distribution with expectation

$$r(t)(T-t) + \frac{\mu(T-t)^2}{2}$$

and variance

$$\frac{\sigma(T-t)^3}{3}$$

thereby giving the price of a zero-coupon bond, by using the risk neutral valuation formula under the risk neutral measure, as *log-normally* distributed, as

$$P(t,T) = E_t\left\{e^{-\int_t^T r(u)du}\right\} = \tag{4.2.12}$$

$$= \exp\left\{-\frac{\mu(T-t)^2}{2} + \frac{\sigma^2(T-t)^3}{6} - r(t)(T-t)\right\} \tag{4.2.13}$$

4.2.3 The Vasicek model

Vasicek (1977) assumed that the dynamics of the instantaneous spot rate under the objective measure $\mathbb{P}$ is an *Ornstein-Uhlenbeck process* with constant coefficients. Thus, for a suitable value for the market price of risk λ, the above assumption implies that the *Vasicek model* is equivalent to the statement that the short-rate r follows an *Ornstein-Uhlenbeck process* with constant coefficients under the risk-neutral measure $\mathbb{Q}$, i.e.

$$dr(t) = \kappa[\theta - r(t)]dt + \sigma dW(t), \; r(0) = r_0 \tag{4.2.14}$$

where r_0, κ, θ and σ are positive constants. Hence, the short-rate $r(t)$ is obtain by integrating both sides of equation 4.2.14 which gives, for $s \leq t$

$$r(t) = r(s)e^{-\kappa(t-s)} + \theta(1 - e^{-\kappa(t-s)}) + \sigma\int_s^t e^{-\kappa(t-u)}dW(u) \tag{4.2.15}$$

The conditional expectation of the short-rate, in the *Vasicek model*, given the filtration $\mathcal{F}_s$ is normally distributed with mean

$$E\{r(t)|\mathcal{F}_s\} = r(s)e^{-\kappa(t-s)} + \theta(1 - e^{-\kappa(t-s)})$$

and as t goes to infinity, we have

$$E\{r(t)|\mathcal{F}_s\} \to \theta$$

The conditional variance is given by

$$Var\{r(t)|\mathcal{F}_s\} = \frac{\sigma^2}{2\kappa}\left[1 - e^{-2\kappa(t-s)}\right]$$

and as t goes to infinity, the variance converges to a constant, i.e.

$$Var\{r(t)|\mathcal{F}_s\} \to \frac{\sigma^2}{2\kappa}$$

The *Vasicek model* is a simple but realistic model. A major disadvantage of this model is that there is a possibility for the short-rates to take on negative values and they cannot exactly fit observed zero-coupon bond prices. However, the distribution of the short-rate in the Vasicek model is Gaussian and its analytical tractability is hard to match when compared to other short rate models.

The price of a zero-coupon bond can be obtain by computing the expectation of

$$P(t,T) = E_t \left\{ e^{-\int_t^T r(u)du} \right\}$$

which leads to the affine term structure representation, i.e.

$$P(t,T) = e^{(A(t,T)-B(t,T)r(t))}$$

where

$$A(t,T) = \exp\left\{ (\theta - \frac{\sigma^2}{2\kappa^2})[B(t,T) - T + t] - \frac{\sigma^2}{4\kappa^2} B(t,T)^2 \right\} \tag{4.2.16}$$

$$B(t,T) = \frac{1}{\kappa}[1 - e^{-\kappa(T-t)}] \tag{4.2.17}$$

4.2.4 The Cox, Ingersoll and Ross (CIR) model

The *CIR model (1985)*, which is an equilibrium approach, is an extension of the *Vasicek Model*. This extension introduces a *"square root"* term in the diffusion coefficient of the instantaneous short-rate dynamics of the Vasicek model. Thus, the *CIR model* of the short-rate under the risk-neutral measure is also based on the *Ornstein-Uhlenbeck process* and is given by

$$dr(t) = dr(t) = \kappa[\theta - r(t)]dt + \sigma\sqrt{r(t)}dW(t),\ r(0) = r_0 \tag{4.2.18}$$

where r_0, κ, θ and σ are positive constants. The *CIR model* is analytically tractable and unlike the Vasicek model, it ensures that the short-rates are always positive by imposing the condition that $2\kappa\theta > \sigma^2$.

The mean and variance of the short rate $r(t)$ condition on the filtration $\mathcal{F}_s$ are given by

$$E\{r(t)|\mathcal{F}_s\} = r(s)e^{-\kappa(t-s)} + \theta(1 - e^{-\kappa(t-s)}) \tag{4.2.19}$$

$$Var\{r(t)|\mathcal{F}_s\} = r(s)\frac{\sigma^2}{\kappa}(e^{-\kappa(t-s)} - e^{-2\kappa(t-s)}) + \theta\frac{\sigma^2}{2\kappa}(1 - e^{-\kappa(t-s)}) \tag{4.2.20}$$

Note that the *CIR model*, like the Vasicek model, is mean reverting. As time t approaches infinity ($t \to \infty$) its conditional mean and variance converges respective to

$$E\{r(t)|\mathcal{F}_s\} \to \theta \tag{4.2.21}$$

$$Var\{r(t)|\mathcal{F}_s\} \to \theta\frac{\sigma^2}{2\kappa} \tag{4.2.22}$$

It is not straight forward to compute the price of a zero-coupon bond in the CIR model, as was in the Vasicek model, because the distribution of the short-rates are now **non-central** χ^2. However, the model admits an affine term structure and generates the following ordinary differential equation

$$\begin{cases} \partial_t B(t,T) = -\kappa B(t,T) + \frac{\sigma^2}{2} B^2(t,T) - 1 \\ B(T,T) = 0 \end{cases} \tag{4.2.23}$$

$$\begin{cases} \partial_t A(t,T) = -\kappa\theta B(t,T) \\ A(T,T) = 0 \end{cases} \tag{4.2.24}$$

The solution to the **Riccati equation** (i.e equ. 4.2.23) is given by

$$B(t,T) = \frac{2(e^{-h(T-t)} - 1)}{(h+\kappa)(e^{-h(T-t)} - 1) + 2h} \tag{4.2.25}$$

$$h = \sqrt{\kappa^2 + 2\sigma^2} \tag{4.2.26}$$

and the expression for $A(t,T)$ is given by

$$A(t,T) = \frac{2\kappa\theta}{\sigma^2} ln\left[\frac{2he^{(\kappa+h)(T-t)/2}}{(h+\kappa)(e^{h(T-t)} - 1) + 2h}\right] \tag{4.2.27}$$

Hence the price of a zero-coupon bond using the *CIR model* is

$$P(t,T) = e^{A(t,T) - B(t,T)r(t)}$$

4.2.5 The Dothan model

The *Dothan model (1978)* is also an extension of the Vasicek model. It is the only known model that models the short-rates as a *log-normal* distribution and have an analytical formula for zero coupon bonds. The log-normal distribution for the short-rates implies that the short-rate are always positive, thus eliminating the main disadvantage of the Vasicek model. However, the variance grows *exponentially* and the risk-less account diverges. The Dothan model was introduced in two stages. First, under the objective probability measure $\mathbb{P}$, the short-rate dynamics was assumed to be a *geometric Brownian motion* with no drift part, i.e.

$$dr(t) = \sigma r(t) dW^{\mathbb{P}}(t), \; r(0) = r_0 \; ,$$

where σ and r_0 are positive constants. Later on, a risk neutral measure dynamics was assumed by introducing a constant market price of risk. Thus giving the dynamics

$$dr(t) = ar(t)dt + \sigma r(t) dW^{\mathbb{Q}}(t), \; r(0) = r_0 \; , \tag{4.2.28}$$

where a is a real constant.

For times $s \leq t$, the dynamics in equation(4.2.28) integrates to

$$r(t) = r(s)\exp\left\{\left(a - \frac{1}{2}\sigma^2\right)(t-s) + \sigma(W(t) - W(s))\right\} \tag{4.2.29}$$

Thereby giving us that the short-rate condition on the filtration $\mathcal{F}_s$ is *log-normally distributed* with mean value

$$E[r(t)|\mathcal{F}_s] = r(s)e^{a(t-s)}, \tag{4.2.30}$$

and variance

$$Var[r(t)|\mathcal{F}_s] = r^2(s)e^{2a(t-s)}\left(e^{\sigma^2(t-s)} - 1\right)$$

As mentioned earlier, the process is not mean reverting. Mean reversion can only be obtained if the real constant a is negative and the mean-reversion level that is necessary is equals to zero (as can be observed in equation(4.2.30).

From the *Dothan model*, the zero-coupon bond price is given by

$$P(t,T) = \frac{\bar{r}^p}{\pi^2}\int_0^\infty sin(2\sqrt{\bar{r}}\sinh y)\int_0^\infty f(z)\sin(yz)dzdy + \frac{2}{\Gamma(2p)}\bar{r}^p K_{2p}(2\sqrt{\bar{r}}) \tag{4.2.31}$$

where

$$\begin{aligned} f(z) &= \exp\left[\frac{-\sigma^2(4p^2 + z^2)(T-t)}{8}\right] z\left|\Gamma\left(-p + i\frac{z}{2}\right)\right|^2 \cosh\frac{\pi z}{2} \\ \hat{r} &= \frac{2r(t)}{\sigma^2} \\ p &= \frac{1}{2} - a, \end{aligned}$$

and K_q denotes the modified *Bessel function* of the second kind of order q.

Chapter 5

THE NELSON-SIEGEL MODELS

The need to accurately estimate, model and forecast the *yield curve* has lead to a variety of both parametric and nonparametric models. In the money market, estimates of the *yield curve* are used daily by *speculators* to take position on the market, *arbitragers* to make risk-less profits and *hedgers* to minimize *portfolio risk*.

In estimating the yield-curve, various quotes on the interest rate market are used depending on the type of model been applied. These quotes include the *spot rate*, the *zero-coupon rates* or the *forward rates*.
In this thesis, I will consider a specific parametric family, the *Nelson Siegel class*, and use some of its members to fit, estimate and forecast Swedish government yields. Before embarking on this task, an introduction of the various types of the *Nelson Siegel* models that will be of interest in this thesis are given.

5.1 Nelson-Siegel Model (NS)

The *Nelson Siegel model*, first introduced by *Nelson and Siegel (1987)*, is one of the most popular *parametric models* used in fitting the *yield curve*. It is both simple and flexible and provides statistically accurate and economically meaningful results.

Nelson and Siegel observed that functions that generates the varying shapes and forms the *yield curve* can assumed, are related to *solutions to difference or differential equations*. If the difference/differential equations represents *instantaneous rates*, then its solution will be the *forward rates*. They therefore suggested to fit the *forward rate curve*, at a given point in time, with approximating functions that consist of the product between a *polynomial* and an *exponentially* decaying term as

$$f(\tau) = \beta_1 + \beta_2 e^{-\tau\lambda} - \beta_3(\tau\lambda)e^{-\tau\lambda} \tag{5.1.1}$$

where β_1, β_2 and β_3 are model parameters that are determine by initial conditions and λ, the decay parameter, is a constant.

By averaging over these *forward rates*, the model for the *continuously compounded instantaneous rates* $y(\tau)$ is obtained. Thus, the *Nelson Siegel spot rate*

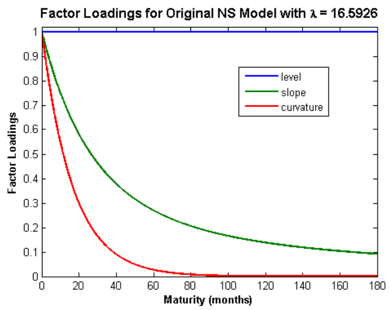

Figure 5.1. The factor loadings on the original Nelson Siegel yield curve representation.

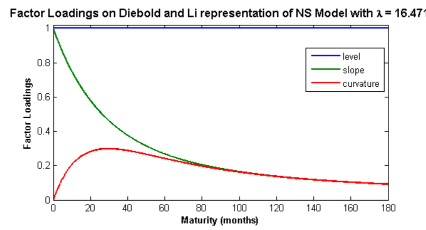

Figure 5.2. The factor loadings on the Dynamic Nelson Siegel yield curve representation by *Diebold and Li (2006)*

.

$y_t(\tau)$ is given by:

$$y_t(\tau) = \beta_1 + (\beta_2 + \beta_3)\frac{1 - e^{-\tau\lambda}}{\tau\lambda} - \beta_3 e^{-\tau\lambda}. \tag{5.1.2}$$

In the above representation, the parameter λ governs the exponential rate of decay of the other model parameters. Thus, large values of λ induces a faster decay and can therefore better fit short-rates whereas small values induces slow decay thereby fitting the yield-curve at long maturities better. In particular, the decay parameter determines the maturity at which the loading on β_3 reaches its maximum.

The β's in the *Nelson Siegel* spot rate representation above, can be given a viable economic interpretation by examining its limiting properties. Observe that as the time to maturity τ approaches infinity, the *spot rate* $y(\tau)$ approaches β_1 and as τ approaches zero, the spot rate approaches $(\beta_2 + \beta_3)$. Thus we have the following limiting properties for the *Nelson Siegel spot rate*:

$$\lim_{\tau\to\infty} y(\tau) = \beta_1 \tag{5.1.3}$$

$$\lim_{\tau\to 0} y(\tau) = \beta_2 + \beta_3. \tag{5.1.4}$$

Hence β_1 corresponds to the *long-term component* of the yield curve and it governs the *level* of the yield curve. From equation 5.1.4, β_2 can be interpreted as the *short-term component* as it governs the *slope* of the yield curve and β_3, the *medium-term component* of the yield curve since it governs its *curvature*.

In the Nelson-Siegel model, the yield for a particular maturity can be seen as a sum of several different components, namely the long-run yield that is independent of the time to maturity (*level*), the effect on the short-end of the yield curve (*slope*) and a component that adds a hump to the yield curve (*curvature*). These components are denoted in the models in equation 5.1.2 as β_1, β_2, and β_3 and their factor loadings are shown in Figure 5.1.

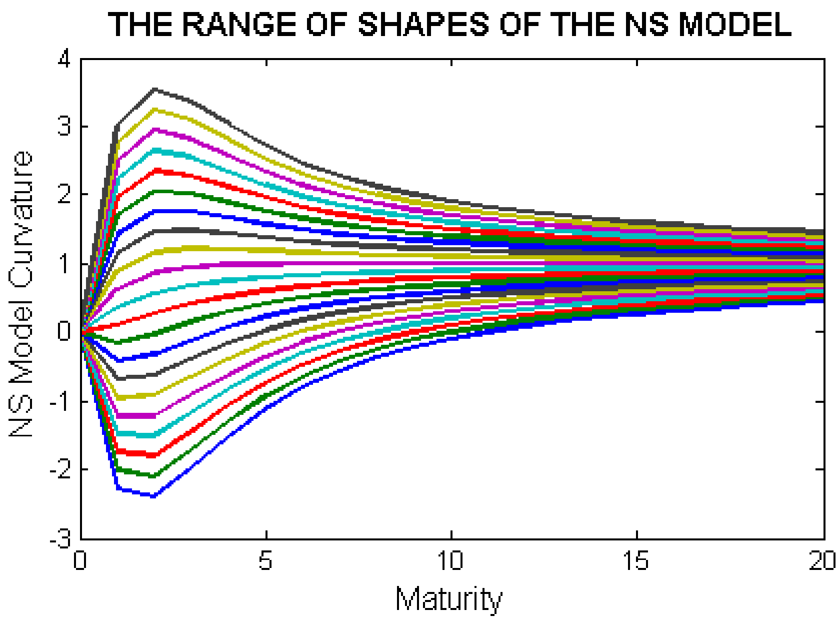

Figure 5.3. The varying shapes and forms the *Nelson Siegel* model can represent. The graph was constructed by fixing $(\beta_2 + \beta_3) = 0$ and $\beta_1 = 1$ in the original *NS*. The decay parameter $\lambda = 1$

.

In general, the Nelson Siegel model family, under the $\mathbb{P}$-*measures*, can be represented by the following state space set-up:

$$X_t = (I - A)\mu + AX_{t-1} + \eta_t, \qquad \eta_t \in N(0, \Sigma_s) \tag{5.1.5}$$

$$Y_t = Z_\lambda X_t + \epsilon_t, \qquad \epsilon_t \in N(0, \Sigma_m) \tag{5.1.6}$$

where the *state equation* X_t is a vector of factors, I an identity matrix, μ a vector of factor means and A is the transition matrix.
The *measurement equation* Y_t specify the vector of Swedish government yields with eight different maturities for any given time t and Z_λ is the constant matrix of factor loadings for a given decay parameter(s), λ.
This representation will by apply to the *DNS* and the *DNSS* models for both the *dependence* and *independence* factor models.

The simplicity of the *Nelson Siegel spot rate* representation and the economic interpretation attached to the model parameters, makes it very attractive to actors in the financial markets. The *Nelson Siegel* model, with only three model parameters, has the ability to capture the various shapes and forms the term structure of interest rates can exhibit as shown in Figure 5.3.
However, the *Nelson Siegel* model is a *static model* that does not take into account the dynamic behavior of its factors. This restriction leads to various extensions on the *Original Nelson Siegel (1987)* model, of which this thesis will consider the re-factorization by *Diebold and Li (2006)* and the dynamic version of its extension by *Svensson (1995)*, which are valuable tools in forecasting the term structure of interest rates. The arbitrage-ree version of the DNS models will also be studied and contrasted with the DNS and the DNSS models.

5.1.1 Dynamic Nelson-Siegel Model (DNS)

Diebold and Li (2006) developed a dynamic model for the yield curve by re-factorizing the *Nelson Siegel spot rate* representation as

$$y_t(\tau) = L_t + S_t\Big(\frac{1 - e^{-\tau\lambda}}{\tau\lambda}\Big) + C_t\Big(\frac{1 - e^{-\tau\lambda}}{\tau\lambda} - e^{-\tau\lambda}\Big) \tag{5.1.7}$$

where the latent factors, L_t, S_t and C_t, have a standard interpretation as level, slope and curvature, given their associated *Nelson Siegel factor loadings.* This re-factorization provides a model that performs as good as the *Original Nelson Siegel* model but have the extra advantage of allowing us to give a different interpretation for the model factors, as described above.

Diebold and Li (2006) show that their representation also corresponds exactly to a modern factor model, with yields that are *affine* in three latent factors. Thus, one can model the dynamics of the three factors with *time series models*, noticeably as *univariate independent stationary autoregressive processes* or as a *vector autoregressive process* using the *State Space Representation* in equation 5.1.5 and 5.1.6. Factor *independence* or *dependence* will be highlighted in our model framework by the form the transition matrix A and the state error matrix, Q attains.

For the independent factor models, the transition matrix and the state error structures are taken to be diagonal, i.e.

$$A = \begin{pmatrix} a_{1,1} & 0 & 0 \\ 0 & a_{2,2} & 0 \\ 0 & 0 & a_{3,3} \end{pmatrix} \quad Q = \begin{pmatrix} q_{1,1}^2 & 0 & 0 \\ 0 & q_{2,2}^2 & 0 \\ 0 & 0 & q_{3,3}^2 \end{pmatrix} \tag{5.1.8}$$

thus modeling the *DNS* factors as independent stationary *AR(1) processes.*

In the dependent case, correlation is induced in both the mean and the covariance matrix. Thus, the transition matrix A is a full matrix and the state error matrix Q is represented as a lower-, or upper triangular matrix as below:

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{pmatrix} \quad Q = \begin{pmatrix} q_{1,1}^2 & 0 & 0 \\ q_{2,1}^2 & q_{2,2}^2 & 0 \\ q_{3,1}^2 & q_{3,2}^2 & q_{3,3}^2 \end{pmatrix} \tag{5.1.9}$$

which gives a *VAR(1)* model for the *DNS* factors.

Empirically, the *DNS* provides good in-sample fit, but it does not impose the fundamental assumption of *absence of arbitrage.* This problem was solved in *Christenson, Diebold, and Rudebusch (CDR)*, where they derived the *affine arbitrage-free* class of dynamic Nelson-Siegel term structure models, referred to as the *Arbitrage Free Nelson-Siegel (AFNS)* model.

The results obtained by *CDR* that are essential for this thesis are reproduced below.

5.1.2 Arbitrage Free Nelson-Siegel model (AFNS)

CDR proposed modeling the yield curve by in-cooperating the empirical success of the *Dynamic Nelson Siegel model* and the theoretical absence of arbitrage assumptions underlying *Affine Processes.* They mimic a *multi-factor affine term structure model* and start their derivation from the standard continuous time affine arbitrage-free term structure.

Considering a three factor model with a constant volatility matrix Σ, *CDR* prove the following proposition

proposition 5.1. *Assume that the instantaneous risk-free rate is defined by*

$$r_t = X_t^1 + X_t^2$$

In addition, assume that the state variables $X_t = (X_t^1, X_t^2, X_t^3)$ *are described by the following system of stochastic differential equations (SDEs) under the risk-neutral* $\mathbb{Q}$*-measure*

$$\begin{pmatrix} dX_t^1 \\ dX_t^2 \\ dX_t^3 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \lambda & -\lambda \\ 0 & 0 & \lambda \end{pmatrix} \left[\begin{pmatrix} \theta_1^Q \\ \theta_2^Q \\ \theta_3^Q \end{pmatrix} + \begin{pmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{pmatrix} \right] dt + \Sigma \begin{pmatrix} dW_t^{1,Q} \\ dW_t^{2,Q} \\ dW_t^{3,Q} \end{pmatrix}, \ \lambda > 0 \tag{5.1.10}$$

Then, zero-coupon bond prices are given by

$$P(t,T) = E_t^{\mathbb{Q}}\big[\exp(-\int_t^T r_u du)\big] \tag{5.1.11}$$

$$= \exp(B^1(t,T)X_t^1 + B^2(t,T)X_t^2 + B^3(t,T)X_t^3 + C(t,T)) \tag{5.1.12}$$

where $B^1(t,T)$*,* $B^2(t,T)$*,* $B^3(t,T)$*,* $C(t,T)$ *are the unique solutions to the following system of ordinary differential equations (ODEs)*

$$\begin{pmatrix} \frac{dB^1(t,T)}{dt} \\ \frac{dB^2(t,T)}{dt} \\ \frac{dB^3(t,T)}{dt} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & -\lambda & \lambda \end{pmatrix} \begin{pmatrix} B^1(t,T) \\ B^2(t,T) \\ B^3(t,T) \end{pmatrix} \tag{5.1.13}$$

and

$$\frac{dC(t,T)}{dt} = -B(t,T)' K^{\mathbb{Q}} \theta^{\mathbb{Q}} - \frac{1}{2} \sum_{j=1}^{3} (\Sigma' B(t,T) B(t,T)' \Sigma)_{j,j'} \tag{5.1.14}$$

with boundary conditions $B^1(T,T) = B^2(T,T) = B^3(T,T) = C(T,T) = 0$*. The unique solution for the system of ODEs is:*

$$\begin{aligned} B^1(t,T) =& -(T-t), \\ B^2(t,T) =& -\frac{1-e^{\lambda(t-t)}}{\lambda} \\ B^3(t,T) =& (T-t)e^{-\lambda(T-t)} - \frac{1-e^{\lambda(t-t)}}{\lambda} \end{aligned}$$

and

$$C(t,T) = (K^Q\theta^Q)_2 \int_t^T B^2(s,T)ds + (K^Q\theta^Q)_3 \int_t^T B^3(s,T)ds \tag{5.1.15}$$

$$+\frac{1}{2}\sum_{j=1}^{3}\int_t^T (\Sigma' B(s,T) B(s,T)'\Sigma)_{j,j} ds \tag{5.1.16}$$

Finally, zero-coupon bond yields are given by

$$y(t,T) = X_t^1 + \frac{1-e^{-\lambda(T-t)}}{\lambda(T-t)} X_t^2 + \left[\frac{1-e^{-\lambda(T-t)}}{\lambda(T-t)} - e^{\lambda(T-t)}\right] X_t^3 - \frac{C(t,T)}{(T-t)} \tag{5.1.17}$$

where $B^1(t,T), B^2(t,T), B^3(t,T)$ *presents the factor loadings,* $K^{\mathbb{Q}}$ *is the mean-reversion matrix,* Σ *is the state-error matrix and* $\theta^{\mathbb{Q}}$ *represents the means of the latent factors.*

Observe that the factor loadings, in the yield function, in equation 5.1.17 exactly matches the factor loadings in the *DNS* as represented in equation 5.1.7. Hence, the significant difference between the *DNS* and the *AFNS* is the *yield-adjustment term* $-\frac{C(t,T)}{(T-t)}$.

From Proposition 5.1, we can also deduced that the *short rate* is a function of *level* and the *slope* of the yield curve only. Also observe that the *yield-adjustment term* is a function of the volatility matrix Σ, the decay parameter λ and the time to maturity $\tau = (T-t)$.

By fixing the mean parameters of the state variable under the $\mathbb{Q}$-measure at zero, i.e. $\theta^{\mathbb{Q}} = 0$, *CDR* represented the *yield-adjustment term* in their *AFNS* as:

$$-\frac{C(T-t)}{T-t} = -\frac{1}{2}\frac{1}{T-t}\sum_{j=1}^{3}\int_t^T (\Sigma' B(s,T)B(s,T)'\Sigma)_{j,j}ds \qquad (5.1.18)$$

Given a general volatility matrix

$$\Sigma = \begin{pmatrix} \sigma_{1,1} & \sigma_{1,2} & \sigma_{1,3} \\ \sigma_{2,1} & \sigma_{2,2} & \sigma_{2,3} \\ \sigma_{3,1} & \sigma_{3,2} & \sigma_{3,3} \end{pmatrix}$$

the *yield-adjustment term* can be derived in analytical form as

$$\begin{aligned}
\frac{C(t,T)}{T-t} =& \frac{1}{2}\frac{1}{T-t}\int_t^T \sum_{j=1}^{3}(\Sigma' B(s,T)B(s,T)'\Sigma)_{j,j}ds \\
=& \bar{A}\frac{(T-t)^2}{6} \\
&+\bar{B}\left[\frac{1}{2\lambda^2} - \frac{1}{\lambda^3}\frac{1-e^{-\lambda(T-t)}}{T-t} + \frac{1}{4\lambda^3}\frac{1-e^{-2\lambda(T-t)}}{T-t}\right] \\
&+\bar{C}\Big[\frac{1}{2\lambda^2} + \frac{1}{\lambda^2}e^{-\lambda(T-t)} - \frac{1}{4\lambda}(T-t)e^{-2\lambda(T-t)} \\
&-\frac{3}{4\lambda^2}e^{-2\lambda(T-t)} - \frac{2}{\lambda^3}\frac{1-e^{-\lambda(T-t)}}{T-t} + \frac{5}{8\lambda^3}\frac{1-e^{-2\lambda(T-t)}}{T-t}\Big] \\
&+\bar{D}\left[\frac{1}{2\lambda}(T-t) + \frac{1}{\lambda^2}e^{-\lambda(T-t)} - \frac{1}{\lambda^3}\frac{1-e^{-\lambda(T-t)}}{T-t}\right] \\
&+\bar{E}\left[\frac{3}{\lambda^3}e^{-\lambda(T-t)} + \frac{1}{2\lambda}(T-t) + \frac{1}{\lambda}(T-t)e^{-\lambda(T-t)} - \frac{3}{\lambda^3}\frac{1-e^{-\lambda(T-t)}}{T-t}\right] \\
&+\bar{F}\left[\frac{1}{\lambda^2} + \frac{1}{\lambda^2}e^{-\lambda(T-t)} - \frac{1}{2\lambda^2}e^{-2\lambda(T-t)} - \frac{3}{\lambda^3}\frac{1-e^{-\lambda(T-t)}}{T-t} + \frac{3}{4\lambda^3}\frac{1-e^{-2\lambda(T-t)}}{T-t}\right]
\end{aligned}$$

where

- $\bar{A} = \sigma_{1,1}^2 + \sigma_{1,2}^2 + \sigma_{1,3}^2$

- $\bar{B} = \sigma_{2,1}^2 + \sigma_{2,2}^2 + \sigma_{2,3}^2$
- $\bar{C} = \sigma_{3,1}^2 + \sigma_{3,2}^2 + \sigma_{3,3}^2$
- $\bar{D} = \sigma_{1,1}\sigma_{2,1} + \sigma_{1,2}\sigma_{2,2} + \sigma_{1,3}\sigma_{2,3}$
- $\bar{E} = \sigma_{1,1}\sigma_{3,1} + \sigma_{1,2}\sigma_{3,2} + \sigma_{1,3}\sigma_{3,3}$
- $\bar{F} = \sigma_{2,1}\sigma_{3,1} + \sigma_{2,2}\sigma_{3,2} + \sigma_{2,3}\sigma_{3,3}$

The above derivation implies two important results. Firstly, the fact that yields in the *AFNS* class of models are given by an analytical formula greatly facilitates empirical implementation of these models. Secondly, the maximally flexible underlying volatility parameters in the *AFNS* specification that can be identified is triangular:

$$\Sigma = \begin{pmatrix} \sigma_{1,1} & 0 & 0 \\ \sigma_{2,1} & \sigma_{2,2} & 0 \\ \sigma_{3,1} & \sigma_{3,2} & \sigma_{3,3} \end{pmatrix} \tag{5.1.19}$$

Note: The choice of lower or upper triangular form is of no significance here.

Observe that the factor dynamics in the *AFNS* cannot be fitted with *time series models* because they are model in continuous time. A change of measure from the real world dynamics (the $\mathbb{P}$-measure) to the risk neutral dynamics (the $\mathbb{Q}$-measure) is therefore needed. The relationship that provides this change of measure is:

$$dW_t^{\mathbb{Q}} = dW_t^{\mathbb{P}} + \Gamma_t dt \tag{5.1.20}$$

where Γ_t represents the risk premium specification. As in *Duffee (2002)* affine dynamics under the $\mathbb{P}$-measure is retained by setting up an affine risk premium specifications as.
Thus, Γ_t, takes the form

$$\begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{pmatrix} = \begin{pmatrix} \gamma_{1,1} & \gamma_{1,2} & \gamma_{1,3} \\ \gamma_{2,1} & \gamma_{2,2} & \gamma_{2,3} \\ \gamma_{3,1} & \gamma_{3,2} & \gamma_{3,3} \end{pmatrix} + \begin{pmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{pmatrix} \tag{5.1.21}$$

This allows us to be able to used any mean vector θ and mean-reversion matrix $\mathbf{K}$, under the $\mathbb{P}$-*measure* and still preserve the affine structure under the $\mathbb{Q}$-*measure*.

With the above specification, the dynamics of the model factors under the $\mathbb{Q}$-*measure*, as in Equation 5.1.10, can be express under the $\mathbb{P}$-*measure* as:

$$\begin{pmatrix} dX_t^1 \\ dX_t^2 \\ dX_t^3 \end{pmatrix} = \mathbf{K}\left[\begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix} + \begin{pmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{pmatrix} \right] dt + \Sigma \begin{pmatrix} dW_t^1 \\ dW_t^2 \\ dW_t^3 \end{pmatrix} \tag{5.1.22}$$

From the above dynamic representation, independent and dependent factors are induced by the form imposed on the *mean reversion matrix* $\mathbf{K}$ and the *state volatility matrix* Σ.

Thus, for *Independent Arbitrage Free Nelson Siegel (IAFNS)* model, the mean reversion and state error matrix are expressed as:

$$\mathbf{K} = \begin{pmatrix} \kappa_{1,1} & 0 & 0 \\ 0 & \kappa_{2,2} & 0 \\ 0 & 0 & \kappa_{3,3} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_{1,1} & 0 & 0 \\ 0 & \sigma_{2,2} & 0 \\ 0 & 0 & \sigma_{3,3} \end{pmatrix} \tag{5.1.23}$$

respectively, and in the *Dependent Arbitrage Free Nelson Siegel (DAFNS)* these matrices take the following form:

$$\mathbf{K} = \begin{pmatrix} \kappa_{1,1} & \kappa_{1,2} & \kappa_{1,3} \\ \kappa_{2,1} & \kappa_{2,2} & \kappa_{2,3} \\ \kappa_{3,1} & \kappa_{3,2} & \kappa_{3,3} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_{1,1} & 0 & 0 \\ \sigma_{2,1} & \sigma_{2,2} & 0 \\ \sigma_{3,1} & \sigma_{3,2} & \sigma_{3,3} \end{pmatrix} \tag{5.1.24}$$

The main *in-sample* problem with the fitted *NS* yield is that, for reasonable choices of the decay paramter, the factor loadings on the *slope* and the *curvature* decays rapidly to zero as maturity increases. This makes it difficult to fit *long rates* as only the factor loading on the *level* factor is left to fit these rates. As decribed in *CDR*, this limitation is highlighted by *lack of fit* in empirical estimation.

A solution to the *NS* lack-of-fit to the cross-section of yields, by adding a second curvature factor with a factor loading component that has a different decay parameter, was proposed by *Svensson (1995).* This extension to the static *Nelson Siegel* model is generally referred to as the *Nelson Siegel Svensson (NSS)* model.

5.1.3 Dynamic Nelson-Siegel Svensson model (DNSS)

The *NSS* model is an extension of the original *NS* model. This model adds a second curvature term, which allows for more flexibility and a better in-sample fit at long maturities. The *NSS* model is one of the most popular models used by Central Banks all around the world to model, estimate and forecast the term structure of interest rates *(BIS (2005))*.

Accordingly, a dynamic version of *NSS* model corresponds to a modern four-factor term structure model. Let $y(\tau)$ be the zero rate for maturity τ, then

$$y_t(\tau) = L_t + S_t\left(\frac{1-e^{(-\lambda_1\tau)}}{\lambda_1\tau}\right) + C_t^1\left(\frac{1-e^{(-\lambda_1\tau)}}{\lambda_1\tau} - e^{(-\lambda_1\tau)}\right) + C_t^2\left(\frac{1-e^{(-\lambda_2\tau)}}{\lambda_2\tau} - e^{(-\lambda_2\tau)}\right)$$

where $y(\tau)$ is the spot-rate curve with τ time to maturity. L_t, S_t, C_t^1 C_t^2 are the model factors and λ_t^1 governs the rate of exponential decay of S_t and C_t^1 whiles λ_t^2 governs the exponential growth and decay rate of C_t^2.

Observe that the added *curvature* mainly affects *medium-term* maturities, which in turn makes it possible and easy to fit yields with more than one local *minima/maxima* along various maturities.

The fact that the *DNSS* model has two decay parameters, $\lambda_{1,t}$ and $\lambda_{2,t}$, makes the model highly *nonlinear*. This nonlinearity is highlighted by a high degree of *multicollinearity* between factors, especially when λ_t^1 assumes similar value as λ_t^2, thereby making it impossible to identify the curvature factors , C_t^1 and C_t^2, separately. Hence the difficulty in the estimation of model parameters.

In general, multicollinearity arises when two or more parameters cannot be identify separately because there exist a high level of dependence between them.

There is therefore the need to define a restriction criteria that will allow us to identify the two curvature factors separately. In this thesis , I followed *Michiel De Pooter (June 5, 2007)* and restrict the factor loading on the second curvature factor, in the *DNSS*, to reach its maximum for a maturity which is at least 1-year shorter than the corresponding maturity for the first curvature's factor loading. *Michiel De Pooter (June 5, 2007)* argues that the above restriction can be expressed as a *minimum distance restriction* as follows:

$$\lambda_{1,t} \geq \lambda_{2,t} + 6.69, \qquad \text{for monthly data.}$$

Observe that, in the *DNSS* models, the two decay parameters and the two curvature factors and hence their role in the minimum distance restriction, are interchangeable.

As in the *DNS* model, the factor dynamics in the *DNSS* model can be represented by the *State Space representation* in equation 5.1.6 and model independently as stationary *AR(1)* or as correlated factors with a *VAR(1)* process, in exactly the same way. The only difference been the size of the *transition matrix A* and the *state error matrix* Σ, which in the *DNSS* model are 4×4 matrices.

Thus for factor *Independent Dynamic Nelson Siegel Svensson (IDNSS)* model, the transition- and state-error matrices are respectively given by:

$$\mathbf{A} = \begin{pmatrix} a_{1,1} & 0 & 0 & 0 \\ 0 & a_{2,2} & 0 & 0 \\ 0 & 0 & a_{3,3} & 0 \\ 0 & 0 & 0 & a_{4,4} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_{1,1} & 0 & 0 & 0 \\ 0 & \sigma_{2,2} & 0 & 0 \\ 0 & 0 & \sigma_{3,3} & 0 \\ 0 & 0 & 0 & \sigma_{4,4} \end{pmatrix} \tag{5.1.25}$$

and in the factor *Dependent Dynamic Nelson Siegel Svensson (DDNSS)* as:

$$\mathbf{A} = \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} \\ a_{2,1} & a_{2,2} & a_{2,3} & a_{2,4} \\ a_{3,1} & a_{3,2} & a_{3,3} & a_{3,4} \\ a_{4,1} & a_{4,2} & a_{4,3} & a_{4,4} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_{1,1} & 0 & 0 & 0 \\ \sigma_{2,1} & \sigma_{2,2} & 0 & 0 \\ \sigma_{3,1} & \sigma_{3,2} & \sigma_{3,3} & 0 \\ \sigma_{4,1} & \sigma_{4,2} & \sigma_{4,3} & \sigma_{4,4} \end{pmatrix} \tag{5.1.26}$$

Chapter 6

APPLICATION OF THE NELSON SIEGEL MODELS

6.1 Data

The data used in this thesis was obtained from the Swedish Central Bank homepage *(www.riksbanken.se)*. The data consist of observed monthly averages of Swedish government bonds in the period *January 1997 to December 2011*. For each month, eight different maturities namely, 1-month, 3-months, 6-months, 1-year, 2-years, 5-years, 7-years and 10-years, were observed. Thus implying that we have both a *time series* and a *cross-sectional* dataset.

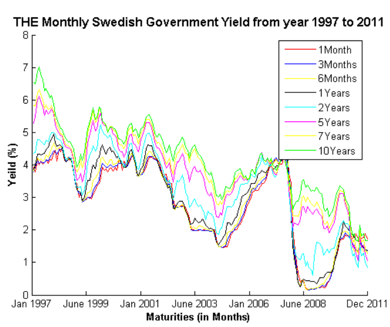

Figure 6.1. The time series of the observed monthly Swedish government yields from the period January 1997 to December 2011.

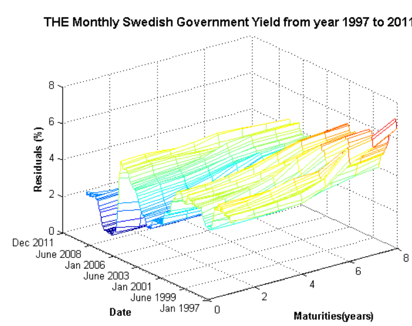

Figure 6.2. A 3D plot of Swedish government yields with maturities ranging from 1-month to 10-years for the period January 1997 to December 2011.

Given that our dataset is both a time series and cross-sectional data, implies for any given month t, we have a yield curve with a set of unknown parameters that fits the curve at that very specific month. Therefore all the model parameters are time varying, which motivates the use of dynamic modelling to represnt our dataset.

A first inspection of our dataset reveals that there are some missing observations for the 1-year bonds. These missing observations were approximated by using the matlab function *missdata* and *iddata*. I then replaced the column of 1-year bonds with missing observations with the results obtained from the matlab interpolation function *missdata*.

Given that Swedish government bills are quoted as *simple rates* and bonds as *effective rates*, I converted the simple rates to effective rates in order to ensure that we have a yield curve for bonds of the same type and hence the same credit risk.

I choose this representation of the yield curve instead of *bootstrapping* the government bonds to obtain/extract its zero-coupon bonds. Not *bootstrapping* the government rates, which is the preferred option, is due mainly to lack of historic Swedish government bond prices for the period under study.

Table 6.1. Descriptive Statistics for Swedish Government Yields

Maturity	Mean	Std. dev.	minimum	maximum	$\hat{\rho}(1)$	$\hat{\rho}(12)$	$\hat{\rho}(30)$
1	2.802	1.278	0.152	4.524	0.9840	0.509	-0.035
3	2.809	1.316	0.152	4.522	0.9850	0.505	-0.012
6	2.870	1.337	0.166	4.725	0.9840	0.510	0.022
12	3.022	1.358	0.257	5.085	0.9843	0.536	0.062
24	3.335	1.304	0.611	5.364	0.9726	0.532	0.208
60	3.884	1.198	1.069	6.300	0.9622	0.487	0.340
84	4.079	1.185	1.279	6.543	0.9620	0.510	0.389
120(Level)	4.291	1.178	1.677	7.268	0.9605	0.494	0.384
Slope	1.482	0.867	-0.586	3.286	0.9563	-0.083	-0.290
Curvature	-0.431	0.648	-2.230	0.926	0.8959	0.087	-0.225

The Descriptive statistics for observed monthly Swedish government yields and the empirical level,slope and curvature, for the period January 1997 to December 2011 in percent. The data based level-, slope- and curvature factors are defined as in section 6.3. The sample autocorrelation functions at displacements of 1,12 and 30 months are also included.

It is important to note that I did not consider bond *liquidity* and *tax effects* that might vary the government rates due to the effect they might have on bond prices, upon which these quoted rates are based. I assumed that the data have already been cleanup by the Swedish Central Bank.

After converting all government rates into effective rates, the effective rates are then continuously compounded and expressed in *decimals* and ready to use for modelling.

Using the whole tenor structure, the data is presented both as a *time-series* and a *three-dimensional* plot as shown in Figure 6.1 and Figure 6.2 respectively. From Figure 6.2, variation in the level, slope and curvature of the yield curve can be observed. The large decline in investments in *long-rates bonds* at the beginning of the global financial crisis is highly visible in Figure 6.1. Also, observed that the interest rates with longer maturities (i.e the long-rates), varies less than the interest rates with short maturities (i.e. the short-rates). In Figure 6.6, the median yield curve of the data with a 25% and 75% inter-percentile range are shown. From Figure 6.6, we observed that the interval is much wider for *short-rates* than for *long-rates*, indicating that the *long-rates* are more stable.

Table 7.1 shows the descriptive statistics of the continuously compounded Swedish government rates, including the empirically defined level, slope and curvature factors that are shown in Figure 6.3, Figure 6.4 and Figure 6.5 respectively

These factors are defined and interpreted as level, slope and curvature as in *Diebold and Li (2006).* Table 7.1 also includes the sample auto-correlation functions with displacements of 1-, 12- and 30-months.

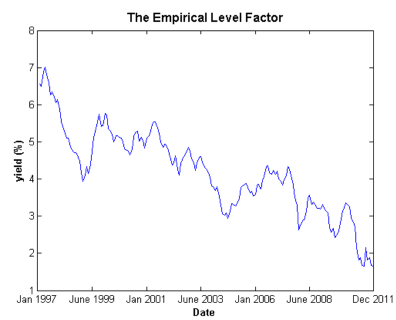

Figure 6.3. The empirical level defined as the yield with the longest maturity, which in our case is the 10-years yield.

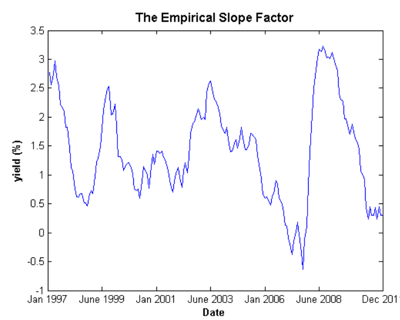

Figure 6.4. The empirical slope defined from data as the difference between the 10-years and the 3-months yields.

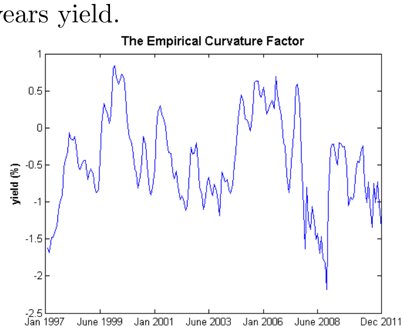

Figure 6.5. The empirical curvature defined from data as twice the 2-years yield minus the sum of the 10-years and 3-months yields.

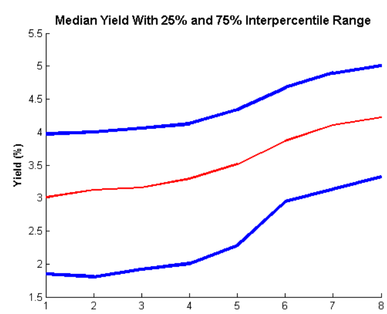

Figure 6.6. The median yield curve for Swedish government bonds with 25% and 75% inter-percentile range.

Diebold and Li (2006) defined the level factor as the yield with the longest maturity, (which in our case is the 10-years yield), They then load the level factor with a constant, one. Hence an increase in the level factor increases all yield equally, as the loading is identical for all maturities. The level factor governs the long term component of the yield curve.

The slope of the yield curve is governs by the short-term components of the term structure. The slope-factor was empirically defined as the difference between the 10-years yield and the 3-months yield. The slope factor is loaded by an exponentially decaying component that starts at one and decays rapidly towards zero. This introduces the decay parameter, λ, that governs the speed at which the slope factor decays.

The medium term component of the yield curve is govern by the curvature factor. The curvature factor is empirically defined as twice the 2-years yield minus the sum of the 3-months and 10-years yields. The curvature factor is loaded by a component that starts at zero, grows to a maximum and then decreases towards zero at a rate determined by the decay parameter, λ.

Statistically, the significance of the empirically defined factors should not be exaggerated. However, two reasons as to why these factors are defined in the following way can be given. Firstly, in the economic world, the factors are defined such that they reflex the time it takes for economic impulses to take effect. Secondly, the factors are defined such that there exist as little correlation as possible between them.
In principle, minimal correlation between factor is desired in order to avoid factor *identifiability problems*, and thereby causing *collinearity issues*. The empirical factors should be seen as references to the Nelson Siegel model class factor dynamics that we are interested in estimating.

Let $(\beta_l; \beta_s; \beta_c)$ represent the empirically defined factors, then their pairwise correlations are $\rho(\beta_l; \beta_s) = 0.15362$, $\rho(\beta_s; \beta_c) = -0.30586$ and $\rho(\beta_l; \beta_c) = 0.16434$, which are very weak indeed.

6.2 Estimation Procedure Overview

Six different dynamic representations of the Nelson Siegel model, that were discussed in Chapter 5, will be investigated. I will estimate *independent* and *dependent* factors for the *dynamic Nelson Siegel DNS* model and the *dynamic Nelson Siegel Svensson DNSS* model. For the *DNS* models, I will lift the arbitrage free restriction they implied and estimate their arbitrage free models, namely the *Arbitrage Free Nelson Siegel AFNS* models.

For the Nelson Siegel model family, the main challenge is in estimating the *decay parameter(s)*. Fixing or freeing the decay parameter determines whether we have a *linear* or a *non-linear* optimization problem.

The size of the decay parameter also determines the exponential rate at which the slope- and curvature factors decays towards zero. In particular, the decay parameter determines the maturity at which the curvature factor achieve its maximum.

As mentioned in Chapter 5, the Nelson Siegel model family, under the $\mathbb{P}$-*measures*, can be represented by the following state space representation:

$$X_t = (I - A)\mu + AX_{t-1} + \eta_t, \qquad \eta_t \in N(0, \Sigma_s) \tag{6.2.1}$$
$$Y_t = ZX_t + \epsilon_t, \qquad \epsilon_t \in N(0, \Sigma_m) \tag{6.2.2}$$

where the *state equation* X_t is a vector of factors, I an identity matrix, μ a vector of factor means and A is the transition matrix.
The *measurement equation* Y_t specify the vector of yields with eight different maturities for any given time t and Z is the constant matrix of factor loadings for a given decay parameter(s), λ.
This representation will by apply to the *DNS* and the *DNSS* models for both *dependence* and *independence* factors. The factors under the $\mathbb{P}$-*measure* will then be model as *univariate independent AR(1)* and also as *multivariate dependent VAR(1)* processes.

The continuous-time counterpart of the *DNS* models are the *AFNS* is models, which are under the $\mathbb{Q}$-*measure*. In order to conform to the same modeling framework in continuous-time and facilitate model comparison, we need to change the probability measure from the $\mathbb{Q}$-*measures* to the $\mathbb{P}$-*measures* and discretized the continuous dynamics under the $\mathbb{P}$-*measure* in order to obtain state equation for the continuous-time models, as was detailed in Proposition 5.1.

Using our state-space representation in equation 6.2.1, the conditional mean and conditional covariance, for our continuous-time models respectively are given by:

$$E^{\mathbb{P}}[X_T|F_t] = (I - \exp(-K^{\mathbb{P}}\Delta t))\theta^{\mathbb{P}} + \exp(-K^{\mathbb{P}}\Delta t)X_t \tag{6.2.3}$$
$$V^{\mathbb{P}}[X_T|F_t] = \int_0^{\Delta t} e^{(-K^{\mathbb{P}})u}\Sigma\Sigma^\top e^{(-K^{\mathbb{P}})^\top u} du \tag{6.2.4}$$

where $\Delta t = T - t$ is the time difference between observations, $K^{\mathbb{P}}$ is the mean-reversion matrix, $\theta^{\mathbb{P}}$ is the vector of factor means, X_t the vector of model factors and Σ_s is the state-error matrix.

The continuous dynamics under the $\mathbb{P}$-*measure* is discretized to obtain the state equation

$$X_i = (I - \exp(-K^{\mathbb{P}}\Delta t_i))\theta^{\mathbb{P}} + exp(-K^{\mathbb{P}}\Delta t_i)X_{i-1} + \nu_t, \tag{6.2.5}$$

where $\Delta t_i = t_i - t_{i-1}$ is the time difference between observations, K^P is the mean-reversion matrix and θ^P is the vector of factor means.

By considering the one-month conditional transition matrix, the relation between the transition matrices under the $\mathbb{P}$-*measures* and the $\mathbb{Q}$-*measures* can be expressed as

$$A = e^{-K\Delta t},$$

The measurement equation in continuous-time is given by

$$Y_t = C + ZX_t + \epsilon_t \tag{6.2.6}$$

where C is the yield adjustment term, Z is a matrix of factor loadings and X_t is a vector of factors. The conditional covariance matrix for the state errors is given by

$$\Sigma_s = \int_0^{\Delta t_i} e^{(-K^P)u} \Sigma\Sigma^\top e^{-(K^P)^\top u} du \tag{6.2.7}$$

Observe that the yield adjustment term C, in the *measurement equation* of the continuous-time model, is the significant difference between the Nelson Siegel models under the *risk neutral dynamics (i.e. the AFNS models)* and the Nelson Siegel models under the *real world dynamics (i.e. the DNS models)*.

To obtain our parameter estimates, a Kalman Filter is used with all the unknown parameters as input. To enforce *stationarity* in our modeling framework, the Kalman Filter is initialized at the unconditional mean and the unconditional covariance of the latent factors that are given in Equation 6.2.3 and Equation 6.2.4 respectively. *Kalman Filter Maximum Log- Likelihood Estimates* are achieved by optimizing the state parameters and the decay parameter λ to maximize the *loglikelihood*.

$$l = \sum_t \big(\frac{-1}{2} [\ log(|S_t|) + v_t' S_t^{-1} v_t] \ \big) \tag{6.2.8}$$

where the predicted error matrix S_t and the predicted error v_t are computed using the Kalman Filter described in Section 3.2.1.

Due to the sensitivity of the estimates of the Nelson Siegel parameters to initial values, extra work is put in identifying optimal initial values for our algorithm. This will go a long way in providing accurate parameter estimates. Below I described the procedure used to initialize and estimate our models.

6.2.1 Estimation Procedure DNS

As mentioned above, initial values provided to the Kalman Filter are indeed decisive in obtaining accurate parameter estimates especially the initial value(s) assign to decay parameter(s). I therefore took extra steps to obtain good initial estimate of the decay parameter in the *DNS* model. The estimates from *DNS* play a are very important role in all other models, as will be shown later.

For any given time in our data, I estimated the decay parameter, in the *DNS*, by minimizing the sum of squares error over a grid of decay parameters. We are interested in

$$min_{\lambda,\beta}(Y - Z_\lambda X)^\top (Y - Z_\lambda X), \tag{6.2.9}$$

where Y is a vector of yields, Z is the matrix of factor loadings and X is the vector of factors. Thus for a given λ, the factors X can be obtained by

$$X_\lambda = (Z_\lambda^\top Z_\lambda^{-1}) Z_\lambda^\top Y$$

By substitution, Equation 6.2.9 can be re-written to optimize over only the decay parameter *lambda* as:

$$\begin{aligned}&min_\lambda(Y - Z_\lambda(Z_\lambda^\top Z_\lambda)^{-1}Z_\lambda^\top Y)^\top(Y - Z_\lambda(Z_\lambda^\top Z_\lambda)^{-1}Z_\lambda^\top Y)\\&= min_\lambda(I - Z_\lambda(Z_\lambda^\top Z_\lambda)^{-1}Z_\lambda^\top)^\top(I - Z_\lambda(Z_\lambda^\top Z_\lambda)^{-1}Z_\lambda^\top)Y\end{aligned}$$

Observe that

$$(I - Z_\lambda(Z_\lambda^\top Z_\lambda)^{-1}Z_\lambda^\top)$$

is an orthogonal projector, hence we only need to optimize

$$min_\lambda(Y^\top Y - Y^\top Z_\lambda(Z_\lambda^\top Z_\lambda)^{-1}Z_\lambda^\top Y) \tag{6.2.10}$$

since if P is an orthogonal projector, then $P = P^\top P$.

The decay parameter that will be used to initialize the Kalman Filter, in the *DNS*, is chosen as the *median* of the time series of decay parameters obtained by optimizing Equation 6.2.10. Three different *matlab* optimization routines, i.e. *fminsearch*, *fmincon* and *fminunc* were used and the results obtained are , 17.4494, 17.5184 and 16.4945 respectively.

Figure 6.7 shows the trajectory of the decay parameter estimated by the three different optimization procedures. Note that the decay parameter in the conditional optimization routine was restricted such that the medium-term component reaches its maximum at a maturity that is between 1- and 3-years.

The median of the time series produced by $fminunc$, i.e. $\lambda = 16.4945$, is taken as the initial optimal decay parameter estimate, as it produces the most stable results and there were no restrictions imposed on the parameters.

Observed that I did not fix the decay parameter to the value recommended by *Diebold and Li (2006)*, i.e. 16.471 for monthly data or 0.0604 for yearly data. The reason been that *Diebold and Li (2006)* used *US government zero coupon bonds* with a different tenor structure which is different from the data and tenor structure used in this thesis.

The decay parameter obtained from the above procedure was then used in the initialization of our Kalman Filter. I then followed the proposal of *Diebold,Rudebusch and Aruoba(2006 b)* and estimate all the model parameters simultaneously, by optimizing them to maximize the *log-likelihood* of the state-space system, i.e.

$$l = \sum_t \big(-\frac{1}{2}[\ log(|S_t|) + v_t' S_t^{-1} v_t]\ \big) \tag{6.2.11}$$

through a Kalman Filter.

The Kalman Filter is used in this thesis because it gives the best linear unbiased estimator for the conditional mean and conditional covariance of a linear dynamic system with latent factors, where the state and measurement errors are assumed to be Gaussian. Also, the Kalman Filter used on a linear state-space model has the advantage of treating the latent factors as unknown, which is much better than setting the short-rates as proxy for the unobserved factors.

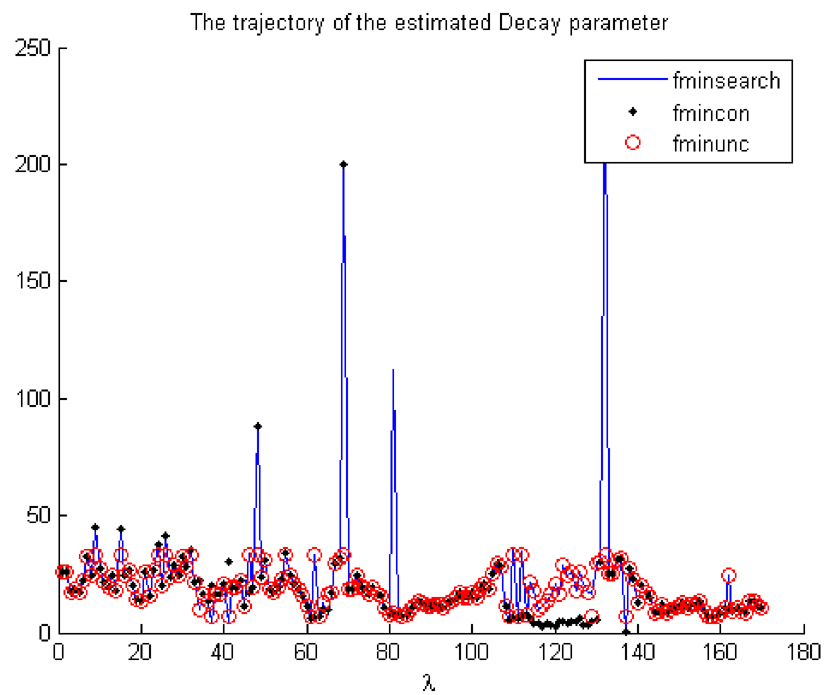

Figure 6.7. The time series of the decay parameters estimated using three different matlab optimization procedure.

After obtaining the *Maximum Likelihood Estimates(MLE)* of the model parameters, I then used the *MLE* estimate of the decay parameter to fixed the factor loading matrix *Z*. The *DNS* model factors were then estimated by *Ordinary Least Squares (OLS)*. The model factors thus obtained are then model as independent factors, i.e.*AR(1)* processes or as correlated factors, i.e. *VAR(1)* process.

The procedure described above offers three main advantages.

- Firstly, it allows us to perform joint optimization in both cross-section and time series dimension.

- Secondly, by fixing the decay parameter, we establish a stable state space framework that provides statistically meaningful factor estimates, which are essential for predicting the dynamics of the yield curve.

- Finally, it converts the estimation problem from the *non-linearity* in our state-space system to a *linear* problem, by allowing us to work with a constant factor loading matrix Z.

6.2.2 Estimation Procedure DNSS

The *DNSS* model have two curvature factors that are govern by two different decay parameters, that allows these factors to decay exponentially towards zero at different rates. This makes estimating model parameters much more challenging compared to the *DNS* models.

The *DNSS* model can exhibit high factor collinearity if the decay parameters are approximately equal, because these parameters are the only difference between the two curvature factors. To avoid factor unidentifiable issues, I imposed the *minimum distance restriction* proposed by *Michiel De Pooter (June 2007)*, as discussed in Section 5.1.3.

The addition of a minimum distance restriction on the decay parameters, to an already difficult nonlinear optimization problem, makes finding optimal starting values for the *DNSS* models even more challenging. There are various initialization procedures outlined in the literatures, of which the initialization proposed by *Michiel De Pooter (June 2007)* is the preferred method.

To initialize the decay parameters, *Michiel De Pooter (June 2007)* proposed setting the first decay parameter (i.e. λ_t^1) to the optimal decay parameter estimates from the *DNS*. To enforce the minimum distance restriction on the decay parameters, for monthly data, the second decay parameter is set as follows:

- If $\hat{\lambda}_t^1$ is larger than twice the minimum allowed value of 6.69 then λ_t^2 is initialized to $0.5\hat{\lambda}_t^1$.

- If $\hat{\lambda}_t^1$ is smaller than 13.38 then λ_t^1 and λ_t^2 are initialized at 13.38 and 6.69 respectively.

To initialize the *level-*, *slope-* and *first curvature* factors, *Michiel De Pooter (2007)* proposed using the optimal factor estimates computed in the *DNS* model and then set the *second curvature factor* to zero.

The above procedure was implemented in *matlab* to optimize

$$\min_{\lambda_1,\lambda_2} \sum (y^{data} - y^{model})^2$$

for every given time t. The *median* of the time series obtained are then taken as our initial optimal decay parameters. Given our dataset, the following result were obtained, $\lambda_1 = 16.9440$ and $\lambda_2 = 7.4436$.

As in the *DNS*, after obtaining our optimal decay parameters, we used them to initialize the Kalman Filter and optimized all the model parameters simultaneously by maximizing the *log-likelihood* in Equation 6.2.11. The information matrix Z is constructed, using the *MLE* decay parameters estimates produced through our Kalman Filter and the *DNSS* factors, i.e. L_t, S_t $C_{1,t}$ and $C_{2,t}$, are computed by *OLS* in exactly the same way as in the *DNS* models.

The discussion on independence and dependence factors, and the form of the *transition-* as well as *state covariance* matrix in the *DNS* case, also applies to the *DNSS* models. The only difference here is the size of these matrices, as was detailed in Section 5.1.3.

6.2.3 Estimation Procedure AFNS

The arbitrage-free models are formulated under the $\mathbb{Q}$-*measure*. Therefore in order to enhance comparison of the estimated parameters for all six models under study, we need a change of measure that relates the *real-world dynamics* under the $\mathbb{P}$-*measure* to the *risk-neutral dynamics* under the $\mathbb{Q}$-measure. This relationship was described in detailed in Section 5.1.2 and will not be repeated here. However, to enhance a smooth flow in the structure of this thesis, the *state-* and *measurement equation* for the *AFNS* model are given below.

In *continuous time*, the *state equation* under the $\mathbb{P}$-*measure* is given by

$$X_t = (I - e^{-K^{\mathbb{P}}\Delta t})\theta^{\mathbb{P}} + e^{-K^{\mathbb{P}}\Delta t} X_t + \nu_t \tag{6.2.12}$$

$$\Sigma_s = \int_0^{\Delta t} e^{(-K^{\mathbb{P}})u} \Sigma \Sigma^{\top} e^{(-K^{\mathbb{P}})^{\top} u} du \tag{6.2.13}$$

and the *measurement equation* under the *P-measure* is given by

$$y_t = C_t + Z_t X_t + \epsilon_t \tag{6.2.14}$$

MLE estimates of the parameters in the *AFNS* modesl are obtained by optimizing these parameters to maximize the *loglikelihood* in equation 6.2.11. Observed that in *continuous time*, a *time-series* model cannot be set on the model factors, which implies that the dynamics are obtained as output of our Kalman Filter.

Stationary of the system under the $\mathbb{P}$-*measures* is imposed by restricting the real component of each eigenvalue of the mean-reversion matrix K^P to be *positive*. The Kalman Filter for the arbitrage free models are started at their unconditional mean and unconditional covariance.

$$\hat{X}_0 = \theta^{\mathbb{P}} \quad \text{and} \quad \hat{\Sigma}_0 = \int_0^{\infty} e^{(-K^{\mathbb{P}})u} \Sigma\Sigma^{\top} e^{(-K^{\mathbb{P}})^{\top}} du. \tag{6.2.15}$$

6.2.4 Forecasting Procedure

To accurately forecast yields is of vital importance to financial actors. It helps in making long term decisions possible. It is of utmost important, for any good yield curve model to be able to perform well both in-sample and out-of-sample. Good fixed income model that also forecast well leads to accurate asset pricing, better portfolio returns and risk management. Certain models performed extremely well *in-sample* but provides unacceptable *out-of sample* results, which may be due to over-parameterization.

To perform out-of-sample forecast for models in the Nelson Siegel family, it is sufficient to only forecast the dynamics of the model factors. In this thesis, I intend to follow *CDR (2009)* approach in constructing out-of-sample forecast for Swedish government yields by applying a *recursive procedure* as described below.

For the *DNS* models, the h period ahead forecast of the yield with maturity τ at time t $\hat{Y}_{t+h}(\tau)$, is given by the conditional expectation

$$E_t^{\mathbb{P}}[Y_{t+h}(\tau)] = E_t^{\mathbb{P}}[L_{t+h}] + E_t^{\mathbb{P}}[S_{t+h}]\left(\frac{1-e^{-\tau\lambda}}{\tau\lambda}\right) + E_t^{\mathbb{P}}[C_{t+h}]\left(\frac{1-e^{\tau\lambda}}{\tau\lambda} - e^{-\tau\lambda}\right)$$

Thus, given parameter estimates for the transition matrix A and the factor means μ, and assuming independent and identically distributed innovations, recursive iteration implies that the conditional expectation of the factors $X_t = (L_t, S_t, C_t)$ in period $t+h$ are given by:

$$E_t^{\mathbb{P}}[X_{t+h}] = (\sum_{i=0}^{h-1} A)(I-A)\mu + A^h X_t. \tag{6.2.16}$$

Note that for the *DNSS* models, a second curvature factor and an additional decay parameter are introduced.

For the *AFNS* models, the forecast in time $t+h$ based on information available at time t is simply the conditional expectation

$$E_t^{\mathbb{P}}[Y_{t+h}(\tau)] = E_t^{\mathbb{P}}[X_{t+h}^1] + E_t^{\mathbb{P}}[X_{t+h}^2]\left(\frac{1-e^{-\tau/\lambda}}{\tau/\lambda}\right) + E_t^{\mathbb{P}}[X_{t+h}^3]\left(\frac{1-e^{\tau/\lambda}}{\tau/\lambda} - e^{-\tau/\lambda}\right) - \frac{C(\tau)}{\tau}.$$

In this case,the conditional expectations are given by

$$E_0^{\mathbb{P}}[X_t] = (I - \exp^{-K^{\mathbb{P}}t})\theta^{\mathbb{P}} + \exp^{-K^{\mathbb{P}}t} X_0, \tag{6.2.17}$$

where $X_t = (X_t^1, X_t^2, X_t^3)$.

Thus with estimates for the mean-reversion matrix $K^{\mathbb{P}}$, the mean vector $\theta^{\mathbb{P}}$, the decay parameter λ and the state error matrix Σ_s along with the optimally filtered factors, we can compute the future expected yields.

The forecast performances of our models are compared by using the root mean squared error ($RMSE$) for the forecast errors that are given by

$$\epsilon_t(\tau, h) = \hat{Y}_{t+h}(\tau) - Y_{t+h}(\tau) \tag{6.2.18}$$

6.3 Estimation Results and Analysis

In this section, I will present the results obtained from the above described procedures for our six *Nelson Siegel* models. The models will be assessed in terms of *goodness of fit* and their ability to *forecast* the Swedish government yields. Based on the *in-sample* and *out-of-sample* results, a model will be chosen as the *best* model to fit Swedish government yields. The matlab-function, *fminsearch* will be used, for parameter estimation in all our models, and the *Nelder-Mead simplex direct search*, which is the default algorithm for *fminsearch*, will be used and iterated 1000 times.

6.3.1 DNS Estimation Results

Here I present and contrast the results of the estimated parameters for the *DNS* models. In Table 6.2 and Table 6.3, estimates of the transition matrix A , the mean vector μ , the estimates for the transition errors q, the decay parameter λ and the maximum log likelihood, for the independent and dependent *DNS* models respectively, are displayed.

Table 6.2. Estimated Independent Factor Dynamics (DNS)

A	L_{t-1}	S_{t-1}	C_{t-1}	μ	q	λ	loglike
L_t	0.9112	0	0	3.21	0.0027	19.4370	-6696.4
St	0	0.8583	0	-2.56	0.0020		
C_t	0	0	0.9416	-1.06	0.0021		

The table shows the estimated transition matrix, A, the mean vector (μ) and the estimated transition error parameters q, in the independent *DNS* model on Swedish government bonds for the period January 1997 to December 2011. The maximum negative log-likelihood value is -6696.4, The estimated value for the decay parameter is 19.4370 on monthly maturities.

Table 6.3. Estimated dependent Factor Dynamics (DNS)

A	L_{t-1}	S_{t-1}	C_{t-1}	μ	q_L	q_S	q_C	λ	loglike
L_t	0.7015	0.0108	0.0137	3.3604	0.0020	0	0	18.4859	-6928.0
St	0.0101	0.9819	0.0096	-2.0005	0.0017	0.0022	0		
C_t	0.0112	0.0089	0.8740	-1.1879	0.0021	0.0023	0.0020		

The table shows the estimated transition matrix, A, the mean vector (μ) and the estimated transition error parameters q, in the dependent *DNS* model on Swedish government bonds for the period January 1997 to December 2011. The maximum negative log-likelihood value is -6928.0, The estimated value for the decay parameter is 18.4859 on monthly maturities.

From Table 6.2 and Table 6.3, we observed, from their respective transition matrices, that the *curvature* factor is the most persistent for the independent *DNS* model, whereas the *slope* factor is the most persistent in the dependent *DNS* model. We also observed that the independent *slope* factor has the fastest *mean-reversion* *rate* whereas in the correlated model, the *level* factor reverts to its mean faster than the other two factors. The persistence in the independent *curvature* factor is greatly reduced in the correlated model while the persistence of the independent *slope* factor increases with almost the same amount as the reduction in persistent in the curvature.

The reduction in persistent for the dependent level can be attributed to the introduction of factor interaction that was induced by the full transition matrix. Observed that all the off-diagonal elements in the dependent *DNS* are almost of the same size. The interesting off-diagonal element that plays an important role in the derivation of the the *AFNS* model is that of (S_t, C_{t-1}), which is estimated to 0.0096. The low estimated value for (S_t, C_{t-1}) might be indicating that our dataset does not support the *arbitrage-free* models as proposed by *CDR (2008)*, and its contribution to the *in-sample-fit* will be minimal if not negligible. This will be investigated later on in this thesis.

The decay parameter λ, for the independent model was estimated to 19.4370. The estimated value of the decay parameter implies that the curvature factor for the independent *DNS* model, given our dataset, reaches its maximum near the three-years maturity, i.e. at approximately 2-years and 9-months. In the dependent model, the decay parameter was estimated to 18.4859, which is slightly lower than in the independent model. It maximizes the dependent curvature factor near the 2-years 8-months maturity. This indicates that the correlated *DNS* model fit long maturity bonds better than the independent *DNS* model.

Given that I used a different dataset and tenor structure to that used by *Diebold and Li (2006)*, the decay parameters obtained here still maximizes the curvature factor for the *DNS* models at a maturity between 1- and 3-years, as *Diebold and Li* recommended.

Observed that the deviation of the factor-mean vector in the *DNS* models is quite significant. This may be indicating the difficulty our models are experiencing in locating these means.

The correlation between the estimated-slope and the estimated-level, for the *DNS*, should be weak in order to avoid *collinearity*. This correlation for the independent factor model was shown to be -0.6484, whereas in the dependent model the correlation was shown to be -0.60991 . The slightly higher correlation in the independent factor *DNS* model may be cause by the slightly higher value of its decay parameter compared to that obtained in the dependent case.

Our reference factors, i.e the empirically defined level, slope and curvature $(\beta_l, \beta_s, \beta_c)$, were plotted against the estimated independent *DNS* factors
i.e. $\hat{\beta}_l^{indep}, \hat{\beta}_s^{indep}, \hat{\beta}_c^{indep}$ as shown in Figure 6.8, 6.9 and 6.10.
Observe that there are three time series in Figures 6.8, 6.9 and 6.10. The trajectories for the estimated independent level, slope and curvature, labeled β_t^1, β_t^2 and β_t^3 respectively, were obtained by fixing λ at its MLE estimate of 19.4370 and the model factors computed by *OLS*. The trajectories labeled as X_t^1, X_t^2 and X_t^3 were obtained by using the MLE factor dynamics estimates directly.

The significance for including both trajectories is to show the effect on the factor estimates that my approach have produced. It could be observed that the factor estimates obtained by fixing the decay parameter to its MLE estimate is much more smooth compared to the factor estimates generated by directly using the state parameter estimates. In the remainder of the thesis, all model factors will be estimated as in β_t^1, β_t^2 and β_t^3.

In Figures 6.14, 6.15 and 6.16 the dependent- against the independent factor estimates for the *DNS* model are shown. There is not much difference between the two models as shown by the fit they exhibit. Observed that even though we modeled the factors as *latent* by using a state-space representation, the factors indeed can be interpreted as *level*, *slope* and *curvature* as they display the same form as the empirically defined factors.

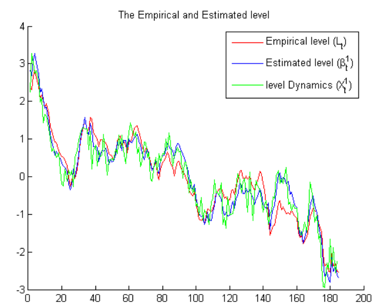

Figure 6.8. The empirical- and estimated level for the independent *DNS* model.

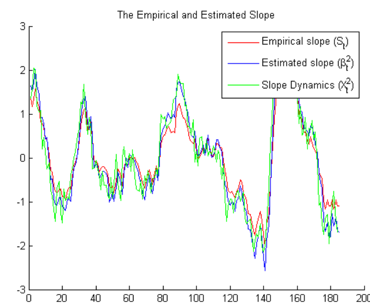

Figure 6.9. The empirical- and estimated slope for the independent *DNS* model.

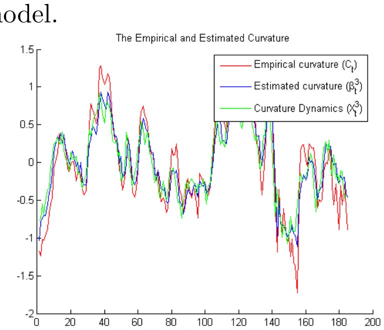

Figure 6.10. The empirical- and estimated curvature for the independent *DNS* model.

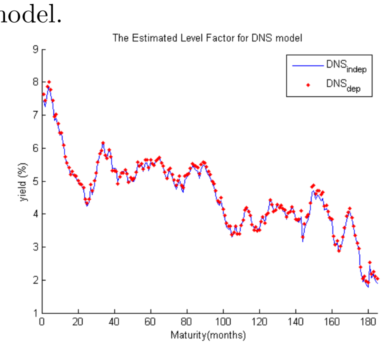

Figure 6.11. The estimated independent and dependent level factor for the *DNS* model.

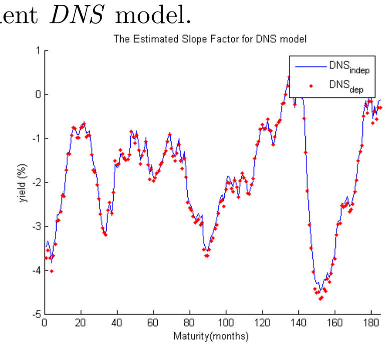

Figure 6.12. The estimated independent and dependent slopefactor for the *DNS* model.

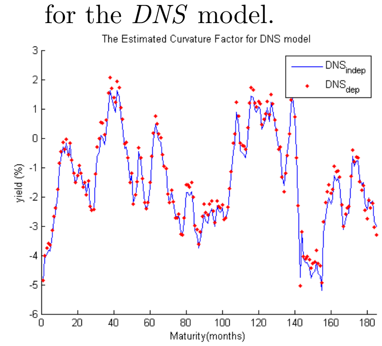

Figure 6.13. The estimated independent and dependent curvature factor for the *DNS* model.

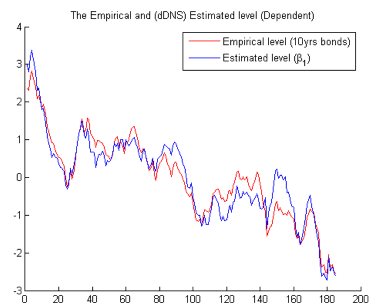

Figure 6.14. The empirical- and estimated level for the independent *DNS* model.

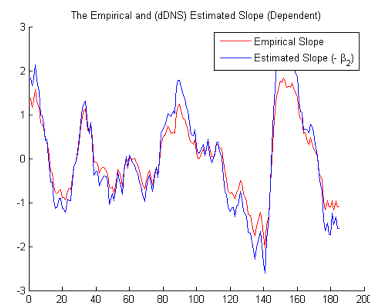

Figure 6.15. The empirical- and estimated slope for the independent *DNS* model.

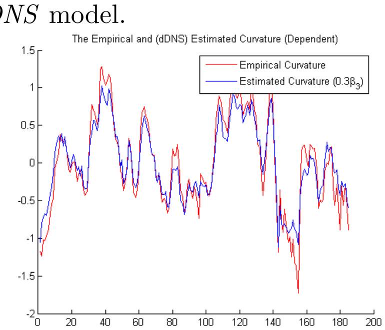

Figure 6.16. The empirical- and estimated curvature for the independent *DNS* model.

To compare the transition errors, in the two models, across various maturities, I followed *CDR (2008)* and constructed the 1-month conditional covariance matrix Q by using the state error estimates in Table 6.2 and 6.3.
Thus, for the *DNS* models, the estimated conditional covariance matrix, given by $Q = qq'$ were estimated as

$$Q^{DNS}_{indep} = \begin{pmatrix} 0.4056 \times 10^{-5} & 0 & 0 \\ 0 & 0.7085 \times 10^{-5} & 0 \\ 0 & 0 & 0.4488 \times 10^{-5} \end{pmatrix} \tag{6.3.1}$$

$$Q^{DNS}_{dep} = \begin{pmatrix} 0.405 \times 10^{-5} & 0.339 \times 10^{-5} & 0.417 \times 10^{-5} \\ 0.339 \times 10^{-5} & 0.764 \times 10^{-5} & 0.851 \times 10^{-5} \\ 0.417 \times 10^{-5} & 0.851 \times 10^{-5} & 0.1337 \times 10^{-5} \end{pmatrix} \tag{6.3.2}$$

From the conditional covariance matrices above, the first observation that can be made is that the volatility of the *level* and *slope* factor, in both models, are very similar and the *slope* factors varies less. On the other hand, the variance of the *curvature* factor increases in the dependent *DNS* model.

Note that little can be said about the differences between the two models, if any, based on their transition- and one-month conditional covariance matrices, discussed above. However, if the *negative log-likelihood* for the two models are considered, we observed that the negative log- likelihood for the independent *DNS* model is slightly less than that of the dependent *DNS* model, which might be indicating that the independent model better fits dataset. Less is better in this case because we optimizing the *negative-log-likelihood* by using the matlab *fminsearch* function. Based on the structures of the two *DNS* models, we observed that the two models are nested, which facilitates the use of *log-likelihood ratio test* to help choose a model that best fits our dataset, *in-sample*. The log-likelihood ratio *(llr)*, is given by

$$llr = 2[\log L(\theta_{dep}) - \log L(\theta_{indep})] \approx \chi^2(p) \tag{6.3.3}$$

where the number of restricted parameters p equals 9. Using matlabs likelihood-ratio test function *lratiotest* gives a critical value of 16.9190 and indicates that at the 5% level, the test rejects the restricted model, i.e the independent *DNS* model. Thus, the restrictions in the independent factor *DNS* model are not supported by our dataset and indicates that there is some interaction between factors. The increased parameterization in the dependent *DNS* model and their estimates thus obtained, will represent well our dataset and improved the *in-sample fit*.

In Table 6.4, the summary statistics for the error generated by the *DNS* and the *AFNS* model are presented.

The deductions made from our log-likelihood ratio test indicates that the dependent *DNS* model represents the Swedish government bonds better than the independent *DNS* model. From Table 6.4, we observed that the difference in the fitted errors for our *DNS* models are quite small and both models preformed equally well across maturities.

Table 6.4. Summary Statistics of In-Sample Fit For DNS and AFNS

Maturity in months	indep-DNS		dep-DNS		indep-AFNS		dep-AFNS	
	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE
1	0.0336	0.0584	0.0312	0.0568	0.0320	0.0574	0.0209	0.0518
3	-0.0082	0.0437	-0.0087	0.0436	-0.0085	0.0436	-0.0100	0.0429
6	-0.0236	0.0533	-0.0222	0.0531	-0.0227	0.0531	-0.0151	0.0526
12	-0.0299	0.1102	-0.0271	0.1091	-0.0280	0.1094	-0.0155	0.1050
24	0.0273	0.0796	0.0284	0.0800	0.0281	0.0799	0.0292	0.0828
60	0.0161	0.0778	0.0126	0.0762	0.0138	0.0767	-0.0034	0.0712
84	-0.0116	0.0572	-0.0132	0.0573	-0.0127	0.0573	-0.0188	0.0582
120	0.0038	0.0523	-0.0009	0.0523	0.0019	0.0523	0.0127	0.0543

.

6.3.2 AFNS Estimation Results

Now we present the model parameter estimates for the arbitrage-free counterpart of the *DNS* models, i.e. the *dependent* and *independent AFNS* models. First we display the *mean-reversion matrix* **K** and the *factor-means* θ, which are shown in Table 6.5 for the independent AFNS model and in Table 6.6 for the dependent AFNS model.

Given that the *AFNS* models are estimated in continuous time, only the factor means from Table 6.5 and Table 6.6 can be compared to the *DNS* models discussed above. Overall, the factor means in the *AFNS* models are slightly higher compared to the *DNS* models factor mean estimates. The difference between the factor mean estimates in the *DNS* models and the *AFNS* models might be indicating that our models are having difficulties in estimating these means. This difficulty might be attributed to the high persistence of our model factors.

Table 6.5. Estimated Independent Factor Dynamics (AFNS)

K matrix				Mean (θ)
	K_l	K_s	K_c	
K_l	0.9307	0	0	3.7541
K_s	0	0.8047	0	-2.0265
K_c	0	0	0.6810	-1.22287

The table shows the mean-reversion matrix K and the factor means for the independent AFNS model

In Tables 6.7 and 6.8, we present the estimates for the state error matrix, the decay parameter and the *negative log-likelihood estimate* for the *independent* and

Table 6.6. Estimated dependent Factor Dynamics (AFNS)

K matrix				Mean (θ)
	K_l	K_s	K_c	
K_l	0.6298	0.0095	0.0097	3.2172
K_s	0.0148	0.7854	0.0106	-1.9829
K_c	0.0104	0.0115	0.7047	-0.8125

The table shows the mean-reversion matrix K and the factor means for the dependent AFNS model

dependent AFNS models respectively.

In the *independent AFNS*, the decay parameter was estimated to 18.7911. This implies that the curvature factor in the *independent AFNS* model reaches its maximum at around the same maturity as our *DNS* models , i.e. at 2-years 8-months. This is not surprising since the estimated decay parameters are very similar indeed. The slope and curvature factors in the independent *AFNS* decays at the same rate towards zero as in the the *DNS* models.

This observation might be highlighting that, compared to the *DNS* models, perhaps there will be no significant gain in in-sample-fit for our dataset by using the independent *AFNS* model.

The decay parameter in the *dependent AFNS* model was estimated to 15.3915. This value of λ, which is slightly less compared to the other models considered so far, indicates that there is a slow exponential decay rate for the slope factor and as well as a slow growth and exponential decay for the curvature factor in the *dependent AFNS* model. This slow rate of decay allows the model to fit long maturity bonds well since the curvature factor reaches its maximum at a maturity of 2-years and 3-months.

Using the above decay parameter estimates to fixed the factor loading matrices in the *AFNS* models, the model factors were then estimated by *OLS* and their trajectories are plotted in Figures 6.17, 6.18 and 6.19 against our reference factors and the factor estimates from our *DNS* models. From Figures 6.17, 6.18 and 6.19, we observed that all four models coincided, thereby making it impossible to identify the model that best fits our dataset in-sample.

If model superiority is base on the value of the *negative log-likelihood*, then we will prefer the *dependent AFNS* model to the *independent AFNS* model, as it has a much lower *negative log-likelihood* value, as shown in Table 6.7 and 6.8.

However, a *log-likelihood ratio test* can be performed on the *AFNS* models since the dependent *AGNS* nests the independent *AFNS* model. Using Equation 6.3.3 with 9 degrees of freedom, our likelihood ratio test indicates that we can reject the independent factor *AFNS* model at the 5% level, indicating that the correlated *AFNS* model best represents our dataset.

The dependent *AFNS* helps give a greater flexibility but this is not reflected by

Table 6.7. Estimated state errors for the Independent (AFNS)

Σ matrix				λ	loglike
	σ_l	σ_s	σ_c		
σ_l	0.0018	0	0	18.7911	- 4657.7 .
σ_s	0	0.0019	0		
σ_c	0	0	0.0025		

The table shows the estimated transition error matrix, the decay parameter estimated to 18.7911 and the negative log-likelihood of the parameter estimates is -4657.7 for the *independent AFNS* model

Table 6.8. Estimated state errors for the dependent (AFNS)

Σ matrix				λ	loglike
	σ_l	σ_s	σ_c		
σ_l	0.0023	0	0	15.3915	-6530.8 .
σ_s	0.0022	0.0024	0		
σ_c	0.0020	0.0020	0.0017		

The table shows the estimated transition error matrix, the decay parameter estimated to 15.3915 and the log-likelihood of the parameter estimates is -6530.8 the *dependent AFNS* model

the results of the fitted error-means or the $RMSE$ as shown in Table 6.4. Again there is no superiority, in in-sample fit, between the independent and dependent *AFNS* models. Observe that the fitted-mean error and the $RMSE$ are slightly smaller than those of the *DNS* models.

The correlation between the *independent AFNS* slope and curvature factors was estimated as -0.62274 and the correlation, between the same factors, for the *dependent AFNS* model as -0.45141. The low correlation in both models implies that we can identify the slope- and curvature factors in the *AFNS* models separately and thereby avoiding *collinearity* problems associated with the *Nelson Siegel* model family.

To compare the transition matrix for the *DNS* models and the *AFNS*, we need to convert the continuous state *mean-reversion matrix* K, in the *AFNS*, to the transition matrix A under the P-measure. Since we are modeling monthly data, we therefore need to convert K and the state error matrix Σ_s into their respective one-month conditional matrices.

The *mean-reversion matrix* K is converted into a one-month transition matrix A by

$$A = e^{-K\Delta t},$$

where $\Delta t = \frac{1}{12}$ *and* K *is the transition matrix under the* $\mathbb{Q}$*-measure.*

The transition matrices for the independent and dependent *AFNS* models un-

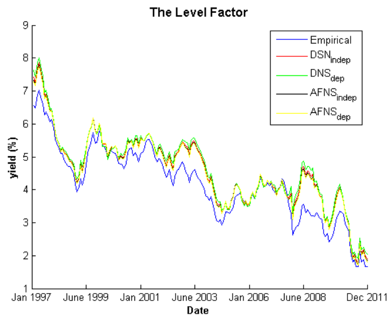

Figure 6.17. The empirical level defined as the yield with the longest maturity against the independent and dependent level estimates for the *DNS* and the *AFNS* models

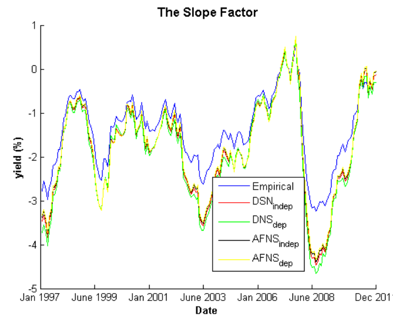

Figure 6.18. The empirical slope defined as the yield with the longest maturity against the independent and dependent slope estimates for the *DNS* and the *AFNS* models

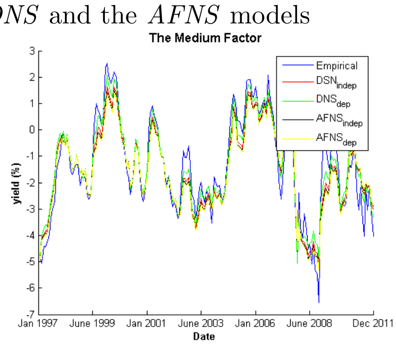

Figure 6.19. The empirical curvature defined as the yield with the longest maturity against the independent and dependent curvature estimates for the *DNS* and the *AFNS* models

der the $\mathbb{P}$-*measure*, as well as their one-month conditional state error matrices are presented in Table 6.9 and Table 6.10 respectively.

The one-month conditional covariance matrix, for the *AFNS* models are obtained by:

$$\int_0^{\frac{1}{12}} e^{-K^P u} \Sigma\Sigma\top e^{-(K^P)\top u} du \tag{6.3.4}$$

The state error matrices, that are used to compute the covariance matrices, are presented as lower triangular matrices because empirically only these parameters can be identified, according to *CDR (2008)*. The choice between lower- and upper triangular matrix for the conditional covariance is irrelevant in this framework.

Table 6.9. Estimated A and Q matrices for independent (AFNS)

	A matrix			Q matrix		
	L_{t-1}	S_{t-1}	C_{t-1}	q_l	q_s	q_c
L_t	0.9254	0	0	0.2765×10^{-6}	0	0
St	0	0.9351	0	0	0.3115×10^{-6}	0
C_t	0	0	0.9448	0	0	0.5399×10^{-6}

The table shows the estimated mean-reversion matrix A, and the one-month conditional covariance matrix for the *independent AFNS* model.

Table 6.10. Estimated dependent (AFNS) model parameters

	A matrix			Q matrix		
	L_{t-1}	S_{t-1}	C_{t-1}	q_l	q_s	q_c
L_t	0.9489	-0.0007	-0.0008	0.4556×10^{-6}	0.4480×10^{-6}	0.3990×10^{-6}
St	-0.0011	0.9366	-0.0008	0.4480×10^{-6}	0.9483×10^{-6}	0.8246×10^{-6}
C_t	-0.0008	-0.0009	0.9430	0.3990×10^{-6}	0.8246×10^{-6}	0.9681×10^{-6}

The table shows the estimated mean-reversion matrix A, and the one-month conditional covariance matrix for the *dependent AFNS*

Considering the transition matrices in Table 6.9 and Table 6.10, we observed that there is not much difference between them. However, compared across models, the transition matrices in the *DNS* and the *AFNS* models are quite different. All factors show more persistence in the AFNS models. This was quite a surprising results since we are making the same assumption about the $\mathbb{P}$-*dynamics* and the only difference between the two models is the *yield-adjustment term* in the *AFNS* models. An observation that can be made here is that, our tenor structure does not include maturities more than 10-years which are necessary to observe the effects of the *AFNS* models. We could also observed that there is minimal factor interaction as indicated by the very low off-diagonal values of the dependent *AFNS* model as Table 6.10 indicates.

From Table 6.9 and Table 6.10, we observed that the volatility of the factors have been greatly reduced compared to the *DNS* models. In both the independent and dependent *AFNS* models, their *curvature* factors has the smallest variance and therefore are less volatile whereas their *level* factors displays the highest volatility.

6.3.3 DNSS Estimation Results

So far, we have presented and interpreted the estimated results obtained for the *DNS* and *AFNS* models, both of which have three latent factors. The flexibility of the *DNS* models was increased by adding a second curvature factor, as detailed in Section 5.1.3.
In Table 6.11 and Table 6.12, we present the estimated transition matrix A, the estimated vector of means μ, the estimated parameters for the conditional covariance matrix Q as well as the estimated decay parameters and the *negative log-likelihood* for the *independent DNSS* model.

Table 6.11. Estimated independent Factor Dynamics (DNSS)

A matrix					Mean (θ)
	L_{t-1}	S_{t-1}	C^1_{t-1}	C^2_{t-1}	
L_t	0.8421	0	0	0	3.60151
S_t	0	0.7588	0	0	-2.2478
C^1_t	0	0	0.9894	0	-0.8611
C^2_t	0	0	0	0.9331	-0.9558

The table shows the estimated transition matrix A, and the factor means for the independent DNSS model using monthly Swedish government yields from the period January 1997 to December 2011.

Table 6.12. Estimated state errors for the independent (DNSS)

Σ matrix					λ_1	λ_2	loglike
	σ_l	σ_s	$\sigma_{c^1_t}$	$\sigma_{c^2_t}$			
σ_l	0.0021	0	0	0	19.0106	8.5619	-6465.2
σ_s	0	0.0022	0	0			
$\sigma_{c^1_t}$	0	0	0.0020	0			
$\sigma_{c^2_t}$	0	0	0	0.0021			

The table shows the estimated transition error matrix, the decay parameters estimated and the negative log-likelihood of the parameter estimates for the *independent DNSS* model using monthly Swedish government yields from the period January 1997 to December 2011

In Table 6.13 and Table 6.14 the corresponding model parameter estimates for the *dependent DNSS* model are presented.

From the transition matrix in Table 6.11, we observed that the *first curvature* factor is the most persistent and the *slope* factor is the least persistent, whereas in the correlated DNSS model the *slope* factor is most persistent as shown in Table 6.13

Compared to the transition matrices for the independent *DNS* and the *AFNS*

Table 6.13. Estimated dependent Factor Dynamics (DNSS)

A matrix					Mean (θ)
	L_{t-1}	S_{t-1}	C^1_{t-1}	C^2_{t-1}	
L_t	0.8343	0.0102	0.0098	0.0100	4.0667
S_t	0.0105	0.9912	0.0106	0.0081	-1.7233
C^1_t	0.0120	0.0108	0.8868	0.0106	-1.0802
C^2_t	0.0108	0.0109	0.0097	0.8024	-0.1059

The table shows the estimated transition matrix A, and the factor means for the dependent DNSS model

Table 6.14. Estimated state errors for the dependent (DNSS)

Σ matrix					λ_1	λ_2	loglike
	σ_l	σ_s	$\sigma_{c^1_t}$	$\sigma_{c^2_t}$			
σ_l	0.0037	0	0	0	16.8379	7.9193	7900.6
σ_s	0.0040	0.0040	0	0			
$\sigma_{c^1_t}$	0.0038	0.0041	0.0040	0			
$\sigma_{c^2_t}$	0.0041	0.0042	0.0046	0.0042			

The table shows the estimated transition error matrix, the decay parameters estimated to and the log-likelihood of the parameter estimates is -7900.6 for the *dependent DNSS* model

models discussed earlier, two observations can be made on the estimated parameters of the *independent DNSS* model. Firstly, the persistent in the *level* factor, in the uncorrelated *DNS* model, is greatly reduced by the introduction of a second curvature term. Secondly, we notice that the levels mean is much higher in the *independent DNSS* model. These two observations shows that without the additional curvature term, only the level factor was used to fit long maturity bonds, whereas with its inclusion, the level factor is release to fit other components of the yield curve. This argument can also be used to justify the reduction in persistence for both the level and first curvature factor in the dependent DNSS model, as Table 6.13 indicates.

The factor means in the *dependent DNSS* model are much higher than those in the independent model and across all the models discussed so far. This is another indication that the factor means are indeed very difficult to locate in our models.

For the *independent DNSS* model, the estimated decay parameter that affects the loading on the slope and first curvature factor is 19.0106, which implies that the first curvature factor reaches its maximum at around the 2-years 9-months maturity. The decay parameter that affects the loading on the second curvature factor was estimated to 8.5619, implying that the second curvature factor is maximized at around the 15-months maturity.

In the dependent factors DNSS model, the estimated decay parameters are slightly lower as Table 6.12 Table 6.14 shows. These values for the decay parameters makes the factor loadings for the slope and the two curvature factors to decay at a slower rate than in the uncorrelated DNSS model and thereby fitting long-rates well.

The decay parameter that affects the slope and first curvature factors, in the correlated DNSS model was estimated to 16.8379 which maximizes the first curvature at around the 2-years and 5-months maturity. The second curvature factor is maximized at around the 1-year and 2-months maturity since we estimated the decay parameter that affects its factor loading to 7.9193.

Note that the interval in which the curvature factors reaches their maximum are still within the desired bound of 1- to 3-years and they satisfy the minimum distance restriction imposed on the decay parameters. Also, the slope and curvature factors in the *DNSS* models are maximized at around the same maturities.

In Figure 6.20 we plotted the factor loadings on the *independent DNSS* model with the decay parameters fixed at their MLE estimates to illustrate the effect of these parameters on the factor loadings.

Given their respective decay parameter estimates, the factor loading matrix Z for our *DNSS* models were fixed and their respective model factors estimated by *OLS*. In Figures 6.21, 6.22 and 6.23, estimated factors for the independent DNSS model are plotted against the independent factors- *DNS* and *AFNS* models for comparison.

In Figures 6.21, 6.22 and 6.23, we observed that there is no significant changes in the *level* and *slope* factors. The only noticeable change is seen in the curvature factor. This was indeed expected as there are now two curvatures to fit the humps in the yield curve. The second curvature factor helps in fitting yields with maturities more than 10-years.

Similarly, Figures 6.24, 6.25 and 6.26 shows the time-series for the corresponding dependent factor estimates.

As mentioned earlier, the second curvature factor was included to achieved increased flexibility and thereby help fit yields with maturities more that 10-years. However, our tenor structure does not support such maturities and we therefore cannot study the relation between the second curvature and yields more than 10-years.

The *negative log-likelihood* for the *dependent DNSS* was estimated to -7900.6 whcih is less than the value obtained for the independent DNSS model, i.e -6465.2. These values does not say much to us but since the independent DNSS model is nested in the dependent DNSS, a likelihood ratio test can be performed using Equation 6.3.3 with 18 degrees of freedom. The likelihood ratio test indicates that the uncorrelated model does not fit our dataset well.

The residual mean and RMSE for the DNSS models are shown in Table 6.15. Again the results in Table 6.15 shows no overall superiority between the DNSS models. However, across models, residual mean and RMSE for the DNSS models are much smaller than all the other models. Thus indicating that the DNSS models, with our estimated parameters provides a better in-sample fit to the Swedish

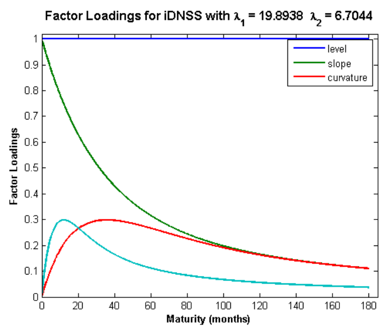

Figure 6.20. The independent DNSS factor loading.

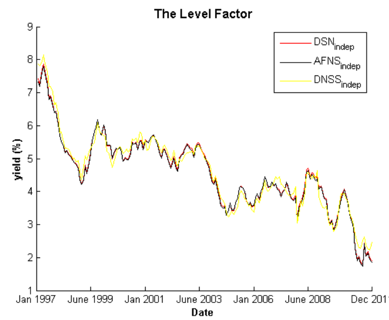

Figure 6.21. The level factor in the DNSS model plotted against the level factor for the the *DNS* and the *AFNS* models.

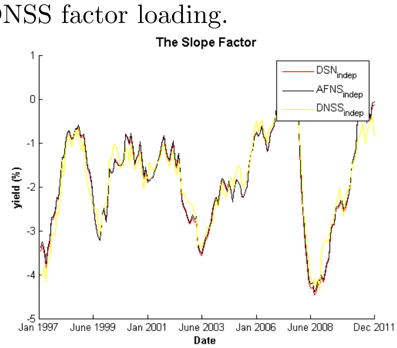

Figure 6.22. The slope factor in the DNSS model plotted against the slope factor for the the *DNS* and the *AFNS* models.

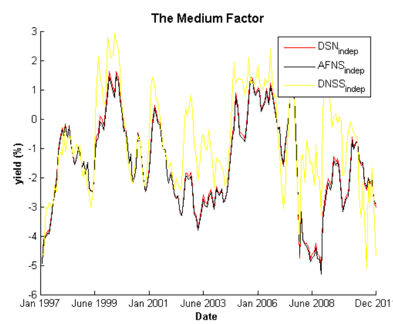

Figure 6.23. The first curvature factor in the DNSS model plotted against the curvature factor for the the *DNS* and the *AFNS* models.

government rates. The results of the mean-error and the RMSE should not be over-interpreted because they might be indicating that the model is over-parameterized and therefore over-fitting our data.

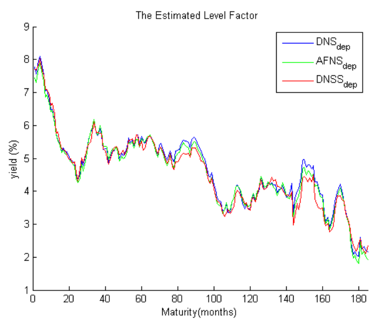

Figure 6.24. The level factor for the dependent DNSS model plotted against the dependent level factors for the the DNS and the AFNS models

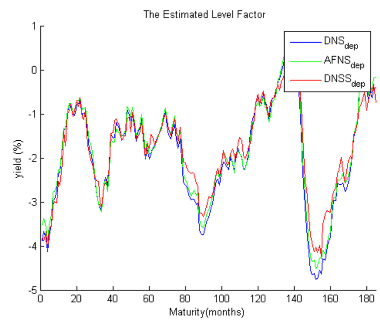

Figure 6.25. The slope factor for the dependent DNSS model plotted against the dependent slopel factors for the the DNS and the AFNS models

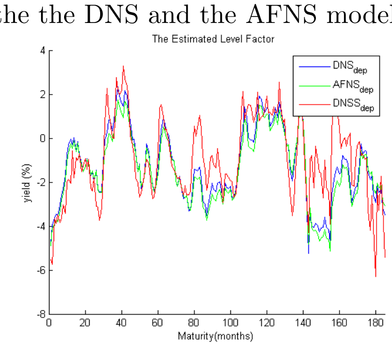

Figure 6.26. The first curvature factor estimate for the dependent DNSS model plotted against the dependent curvature factors for the the DNS and the AFNS models

The conditional covariance matrices, for the *DNSS* models, given by $Q = qq'$, are shown below:

$$Q_{indep}^{DNSS} = \begin{pmatrix} 0.2034 \times 10^{-4} & 0 & 0 & 0 \\ 0 & 0.2007 \times 10^{-4} & 0 & 0 \\ 0 & 0 & 0.1879 \times 10^{-4} & 0 \\ 0 & 0 & 0 & 0.1908 \times 10^{-4} \end{pmatrix}$$

$$Q_{dep}^{DNSS} = \begin{pmatrix} 0.1372 \times 10^{-4} & 0.1469 \times 10^{-4} & 0.1396 \times 10^{-4} & 0.1511 \\ 0.1469 \times 10^{-4} & 0.3194 \times 10^{-4} & 0.3165 \times 10^{-4} & 0.3292 \\ 0.1396 \times 10^{-4} & 0.3165 \times 10^{-4} & 0.4756 \times 10^{-4} & 0.5113 \\ 0.1511 \times 10^{-4} & 0.3292 \times 10^{-4} & 0.5113 \times 10^{-4} & 0.7292 \end{pmatrix}$$

From Q_{indep}^{DNSS}, we observed that for the uncorrelated DNSS model, the level factor is the most volatile and the curvature factors displays less volatility. However, in the correlated DNSS model, the variance on the level and slope factors stays almost the same as those of the uncorrelated model whiles the volatility on the two curvature factors increases.

Table 6.15. Summary Statistics of In-Sample Fit For DNSS

Maturity in months	indep-DNSS		dep-DNSS	
	Mean	RMSE	Mean	RMSE
1	0.0098	0.0401	0.0085	0.0391
3	-0.0083	0.0435	-0.0082	0.0447
6	-0.0050	0.0496	-0.0039	0.0488
12	-0.0074	0.0754	-0.0061	0.0752
24	0.0209	0.0739	0.0209	0.0662
60	-0.0083	0.0589	-0.0118	0.0616
84	-0.0173	0.0580	-0.0189	0.0582
120	0.0157	0.0383	0.0195	0.0421

The table shows the residual means and the RMSE for the DNSS models in basis points.

6.4 Forecasting

From the discussions and in-sample fit analysis in the previous section, we can conclude that the dependent models can better represent our dataset in-sample. The residual means and RMSE shows that the dependent DNSS model provides the best in-sample fit for Swedish government yields.

Now we investigate if the in-sample-fit superiority of the dependent models, especially the dependent DNSS, translates to better *out-of-sample* fit or if it was due to *over-fitting*. We also study the contribution of the *yield-adjustment term*, that were introduced in the AFNS models, to out-of-sample forecast.

The estimation procedure introduced in this thesis mimic the *two-step-approach* proposed by *Diebold and Li (2006)*. Thus the estimation of the model factors almost totally ignores the dynamics of the factors which are crucial in obtaining a good forecast. To forecast the yield curve in the Nelson Siegel model class, for a given decay parameter(s), it is enough to forecast the factor dynamics. Therefore, to perform out-of-sample forecast for our dataset, we implement the *recursive procedure* proposed by *CDR (2009)*, as described in Section 6.2.4.

For all six models discussed, 3-, 6- and 12- months forecast horizons will be computed and the model with the least RMSE forecast error chosen as the best model.

Table 6.16 present the RMSE for the 3-months, 1-year, 2-years, 5-years,7-years and 10-years yields. In Table 6.16, for each forecast horizon, the model with the least RMSE forecast error is mark with an asterisk.

The results in Table 6.17 shows that, for all maturities and all forecast horizons, the *independent dynamic Nelson Siegel model*, which is the simplest of all the six models considered in this thesis, is the most accurate model. This contradicts the results obtained from the *in-sample-fit*, which shows the *dependent Dynamic Nelson Siegel Svensson* model as the most accurate. Out-of-sample, the correlated factor models are the worst performing models for all forecast horizon.

This indicates that the more complex models do over-fit our data and the yield adjustment terms contribution to the out-of sample forecast is not very significant as the *AFNS* models are outperformed by the independent *DNS* model.

Thus, given our dataset and tenor structure, the in-sample-fit superiority of the *DNSS* models can be attributed to over-fitting and the lack of significant contribution of the complex *AFNS* models, both in-sample and out-of-sample fit, may be due to lack of bonds with maturities more than 10-years. This highlight that fact that the extensions of the *DNS* model helps in fitting long-term yields, which are not present in our dataset and hence the effect of these models cannot be observed.

Table 6.16. Out-of-sample Forecast RMSE for Six Models

	Forecast horizon		
Model	Three-month	Six-months	Twelve-months
	3-months yields		
DNS_{indep}	0.5405*	0.8308*	1.0886*
DNS_{dep}	0.8496	0.8496	1.3602
$AFNS_{indep}$	0.5433	1.1915	1.2553
$AFNS_{dep}$	0.9821	1.0949	1.2533
$DNSS_{indep}$	0.6535	0.9471	1.1489
$DNSS_{dep}$	0.6728	1.0440	1.3695
	1-year yields		
DNS_{indep}	0.5491*	0.8283*	1.0658*
DNS_{dep}	0.8662	0.8662	1.3084
$AFNS_{indep}$	0.5755	1.2329	1.2596
$AFNS_{dep}$	1.2406	1.1343	1.2576
$DNSS_{indep}$	0.6289	0.9341	1.1358
$DNSS_{dep}$	0.6949	1.0536	1.3098
	2-years yields		
DNS_{indep}	0.5289*	0.7813*	1.0017*
DNS_{dep}	0.8734	0.8734	1.2097
$AFNS_{indep}$	0.6592	1.1910	1.2017
$AFNS_{dep}$	1.3021	1.0736	1.1993
$DNSS_{indep}$	0.6301	0.9294	1.1290
$DNSS_{dep}$	0.6960	1.0090	1.2003
	5-years yields		
DNS_{indep}	0.4820*	0.6914*	0.8485*
DNS_{dep}	0.8210	0.8210	0.9943
$AFNS_{indep}$	0.5788	1.0552	1.0366
$AFNS_{dep}$	1.2096	0.9074	1.0338
$DNSS_{indep}$	0.5969	0.8624	1.0098
$DNSS_{dep}$	0.6266	0.8678	0.9681
	7-years yields		
DNS_{indep}	0.4585*	0.6684*	0.8318*
DNS_{dep}	0.7919	0.7919	0.9720
$AFNS_{indep}$	0.5452	1.0340	1.0176
$AFNS_{dep}$	1.1776	0.8763	1.0146
$DNSS_{indep}$	0.5697	0.8280	0.9817
$DNSS_{dep}$	0.5918	0.8277	0.9378
	10-years yields		
DNS_{indep}	0.4273*	0.6195*	0.7742*
DNS_{dep}	0.7650	0.7650	0.9150
$AFNS_{indep}$	0.4833	0.9887	0.9556
$AFNS_{dep}$	1.1335	0.8180	0.9523
$DNSS_{indep}$	0.5332	0.7673	0.9072
$DNSS_{dep}$	0.5550	0.7701	0.8719

6.5 Conclusion

In this thesis, three different classes of the Nelson Siegel model family namely the Dynamics Nelson Siegel model, the Dynamics Nelson Siegel Svensson model and the Dynamics Arbitrage-Free Nelson Siegel model, were examined. For each model, I studied both the effect correlated- and independent factors have on the *in-sample-fit* and *out-of-sample forecast* for Swedish government bonds under the period January 1997 to December 2011.

The results of the *in-sample-fit* shows the correlated Dynamic Nelson Siegel Svensson model, were factor interaction and dependence are introduced, as the most accurate model to represent our dataset and the given tenor structure, as it improves considerably the fit on the Dynamic Nelson Siegel models.

However, the *out-of-sample forecast* for the six models reveals that the superiority of the DNSS in-sample-fit was due to over-fitting as they were outperformed by the simple independent factor DNS model. The effect of over-fitting was also visible by the fact that the dependent models and the complex AFNS models performed worst out-of-sample.

This shows that the contribution of the *yield-adjustment term*, in the measurement equation of the AFNS models, is not very significant in forecasting the yield curve. To my surprise, the model that incorporate the arbitrage-free tradition of *Affine Processes* and Nelson Siegel model class tradition of providing good in-sample fit, was out-performed by the independent three factor DNS model given our dataset.

6.6 Extension

In this thesis, I do not consider the effects macroeconomic factors, such as inflation, unemployment and productivity have on Swedish government yields. The modeling framework used in this thesis can easily be extended to include these factors in the state representation and their interaction, if any, studied. In this vain, it could also be interesting to study the relation between Swedish fixed income securities and the European Union bond market.

There is also the possibility to investigate an arbitrage-free counterpart for the DNSS models. This was proposed by *CDR (2008)*, where they introduced a five factors Dynamic Nelson Siegel model, namely the *Generalized Arbitrage-Free Nelson Siegel (AFGNS)* models that is a very clever extension of the three factors Arbitrage-Free Nelson Siegel model.

BIBLIOGRAPHY

[1] *Interest Rate Models, Theory and Practice.Damiano Brigo and Fabio Mercurio, August 2001.*

[2] *Time Series Analysis. Henrik Madsen .*

[3] *Arbitrage Theory in Continuous Time. Björk 3th Edition.*

[4] *Finance Market,Instruments and Investments. Hans Byström, 2nd Edition.*

[5] *Modeling Non-Linear and Non-Stationary Time Series. Henrik Madsen and Jan Holst, December 2000.*

[6] *Stationary Stochastic Processes. Georg Lindgren, Holger Rootzën and Maria Sandsten, September 2009.*

[7] *Derivative Pricing part I and II. Sebastian Ȧberg, 2010.*

[8] *Räntebärande Instrument, Värdering och Riskhantering. Hossien Asgharian and Lars Nordén.*

[9] *De Finansiella Marknaderna i ett Internationellt Perspektiv. Lief Hässel, Marie Norman and Christian Andersson .*

[10] *Bank for International Settlements 2006. ISSN 1683-0121 (print), ISSN 1683-013X (online)*

[11] *Diebold, Francis X. & Li, Canlin, 2006. "Forecasting the term structure of government bond yields," Journal of Econometrics, Elsevier, vol. 130(2), pages 337-364, February*

[12] *" On the Estimation of Term Structure Models and An Application to the United States" ; IMF Working Papers from International Monetary Fund, No 10/258. Giancarlo Gasha, Ying He, Carlos Medeiros, Marco Rodrigues, Jean Salvati, and Jiangbo Yi, November 2010.*

[13] *Berec R. (2010), Extraction of Nelson-Siegel Factors from Bond Prices.*

[14] *Harrison, J. Michael and Pliska, Stanley R., "Martingales and Stochastic Integrals in the Theory of Continuous Trading", in Stochastic Processes and their Applications, North Holland Publishing Company, 1981*

[15] *Term Structure Forecasting. Does a Good Fit Imply Reasonable Simulation Results, by. Emile A.L.J. van Elen (74.30.15).*

[16] *Annaert, Jan; Claes, Anouk; Ceuster, de, Marc; Zhang, Hairui - Estimating the spot rate curve using the Nelson-Siegel model : a ridge regression approach - In: International review of economics and finance, 27(2013), p. 482-496. .*

[17] *An Arbitrage-Free Generalized Nelson-Siegel Term Structure Model. (CDR): Jens H.E. Christensen, Francis X. Diebold, Glen D. Rudebusch, November 2008.*

[18] *Manfred Gilli & Stefan Groe & Enrico Schumann, 2010. "Calibrating the NelsonSiegelSvensson model," Working Papers 031, COMISEF.*

[19] *Svensson, L. E. O. 1994, Estimating and Interpreting Forward Interest Rates: Sweden 1992-1994, NBER Working Paper Series 4871.*

[20] *Gregory R. Duffee, 2002. "Term Premia and Interest Rate Forecasts in Affine Models," Journal of Finance, American Finance Association, vol. 57(1), pages 405-443, 02.*

[21] *Michiel De Pooter, 2007. "Examining the Nelson-Siegel Class of Term Structure Models," Tinbergen Institute Discussion Papers 07-043/4, Tinbergen Institute.*

[22] *Cox, John C & Ingersoll, Jonathan E, Jr & Ross, Stephen A, 1985. "A Theory of the Term Structure of Interest Rates," Econometrica, Econometric Society, vol. 53(2), pages 385-407, March.*

[23] *Vasicek, Oldrich, 1977. "An equilibrium characterization of the term structure," Journal of Financial Economics, Elsevier, vol. 5(2), pages 177-188, November*

[24] *Nelson, C. R.& Siegel, A. F. (1987). Parsimonious modeling of yield curves, Journal of Business 60(4): 473489*

[25] *Dothan, L. Uri, 1978. "On the term structure of interest rates," Journal of Financial Economics, Elsevier, vol. 6(1), pages 59-69, March*

[26] *Diebold, Francis X. & Rudebusch, Glenn D. & Borag[caron]an Aruoba, S., 2006. "The macroeconomy and the yield curve: a dynamic latent factor approach," Journal of Econometrics, Elsevier, vol. 131(1-2), pages 309-338.*