COMENIUS UNIVERSITY IN BRATISLAVA
Faculty of Mathematics, Physics and Informatics
Department of Applied Mathematics and Statistics

Prediction of the Yield Curve Using a No-arbitrage Nelson-Siegel Model

Master Thesis

9.1.9 Applied Mathematics
Economic and Financial Mathematics

Supervisor:
Mgr. Juraj Katriak

Author:
Bc. Martin Kopča

Bratislava 2010

UNIVERZITA KOMENSKÉHO V BRATISLAVE
Fakulta matematiky, fyziky a informatiky
Katedra aplikovanej matematiky a štatistiky

Predikcia výnosovej krivky pomocou bezarbitrážneho Nelson-Siegel modelu

Diplomová práca

9.1.9 Aplikovaná matematika
Ekonomická a finančná matematika

Vedúci diplomovej práce:
Mgr. Juraj Kartiak

Diplomant:
Bc. Martin Kopča

Bratislava 2010

I declare this thesis was written on my own, with the only help provided by my supervisor and the referred-to literature and sources.

Bratislava 26. 04. 2010 ....

Acknowledgement

I would like to express thanks to my supervisor Mgr. Juraj Katriak for his guidance and time that he devoted to me while elaborating this thesis.

I am grateful to Lukáš for his valuable technical support. My warm thanks goes also to my family for their support and love.

Abstract

KOPČA, Martin: Prediction of the Yield Curve Using a No-arbitrage Nelson-Siegel Model [Master Thesis], Comenius University in Bratislava, Faculty of Mathematics, Physics and Informatics, Department of Applied Mathematics and Statistics; Supervisor: Mgr. Juraj Katriak, Bratislava, 2010, 52 pages.

We examine the in-sample fit and out-of sample forecasting performance of a recently developed no-arbitrage version of the popular Nelson-Siegel term structure model. We use the model and estimation procedure proposed in Christensen et al. (2007). In contrast to Christensen et al. (2007), who use data on U.S. Treasury security yields, our analysis is carried out on the data on German government bond spot rates. We cannot report that no-arbitrage restrictions imposed on the model result in gains in the in-sample fitting performace. We do observe a minor improvement in out-of sample forecasting, but even these improved predictions fail to outperform the random walk predictions.

Keywords: term structure, Nelson-Siegel model, no-arbitrage, prediction

Abstrakt

KOPČA, Martin: Predikcia výnosovej krivky pomocou bezarbitrážneho Nelson-Siegel modelu [Diplomová práca], Univerzita Komenského v Bratislave, Fakulta matematiky, fyziky a informatiky, Katedra aplikovanej matematiky a štatistiky; Vedúci diplomovej práce: Mgr. Juraj Katriak, Bratislava, 2010, 52 strán.

V práci vyhodnocujeme schopnosť nedávno navrhnutej bezarbitrážnej verzie známeho Nelson-Siegel modelu odhadnúť výnosovú krivku pre danú množinu spotových sadzieb ako aj jeho schopnosť správne predpovedať výnosové krivky do budúcnosti. Používame pritom model a postup pre jeho odhadnutie navrhnutý v Christensen et al. (2007). Na rozdiel od Christensen et al. (2007), kde odhadujú výnosové krivky pre americké vládne cenné papiere, naša analýza je vypracovaná na údajoch o výnosoch nemeckých vládnych dlhopisov. Na základe našich výsledkov nemôžeme konštatovať, že by aplikácia bezarbitrážnych obmedzení viedla k zlepšeniu schopnosti Nelson-Siegel modelu odhanúť výnosovú krivku pre daný súbor dát. Na druhej strane pozorujeme mierne zlepšenia v oblasti predpovedania budúcich výnosových kriviek. Avšak ani tieto zdokonalené predpovede nedosahujú kvality predpovení vykonaných na základe procesu náhodnej prechádzky.

Kľúčové slová: časová štruktúra úrokových mier, Nelson-Siegel model, arbitráž, predpoveď

Contents

Chapter 1

Introduction

Yield curve modelling is an important topic of the financial mathematics which is particularly relevant for bond traders, risk managers and monetary authorities. Forecasting the future yield curves correctly is the ultimate condition for substantial portfolio returns and responsible monetary policy.

Recently, after encouraging results of Diebold and Li (2006), a noticable effort has been paid to yield curve forecasting. The findings of Diebold and Li (2006) about superior forecasting performance of the Nelson-Siegel model, Nelson and Siegel (1987), started a rally to improve this performance even more. As the original Nelson-Siegel model does not explicitly account for no-arbitrage principle by constructing the yield curve, considerable attention has been paid to resolve this shortcomming.

We follow the procedure to construct and estimate a no-arbitrage version of the Nelson-Siegel model proposed in Christensen et al. (2007). They find a no-argitrage approximation of the original Nelson-Siegel model in the family of affine arbitrage-free models described in Duffie and Kan (1996). Estimation of the model is carried out by employing Kalman filter technique in maximizing the likelihood function.

In this master thesis, we estimate the no-arbitrage Nelson-Siegel model on the data on German government bond yields. To our best knowledge, there is no paper that examines the forecasting performance of this model on other than U.S. Treasury security yields. However minor innovation may our choise of data appear, our findings are very different from those presented in Diebold and Li (2006) and Coroneo et al. (2008). We find that although the forecasting performance of the no-arbitrage Nelson-Siegel model has slightly improved in comparison with the dynamic model described in Diebold and Li (2006) there is no reason for satisfaction. For our data set, forecasts based on the random walk predictions were still more successful than those of the enhanced Nelson-Siegel model. We conclude that a further research is needed and propose a path to resolve these disappointing results.

We proceed as follows. In the chapter 2 we provide a brief intoduction to the terminology of bond markets and the problematic of yield curve modelling. Chapter 3 introduces the well known Nelson-Siegel model as well as its modification proposed in Diebold and Li (2006) that is used for forecasting purposes. Chapter 4 provides a discussion about the no-arbitrage principle, arbitrage on the yield curve and the no-arbitrage approach in yield curve modelling. In chapter 5 we introduce the no-arbitrage version of the Nelson-Siegel model as proposed in Christensen et al. (2007). Chapter 6 presents our epmirical findings and concluding remarks are in chapter 7.

Chapter 2

Terminology

In this chapter we provide a brief introduction into the terminology of the bond market, where the topics addressed in this master thesis belong. The information provided in this chapter is mainly drawn from Melicherčík et al. (2005) and Cairns (2004).

2.1 Bonds

A bond is a security where the buyer of the bond pays the issuer a price P in return for a sequence of future payments. A *zero-coupon bond* (also called *discount bond*) is a bond for which the holder of the bond receives the nominal value (also called face value, par value or principal) of this bond at one time in the future. The date at which this payment is due is called *maturity date*, *maturity* or *redemption date*. A zero-coupon bond with maturity date T years after its issuance is sometimes referred to as a T*-bond*.

A bond that pays regular interest payments (*coupons*) in addition to the principal is called a *coupon bond*. Interest payments can have fixed or variable value, depending on the specification in the terms of the bond. Variable coupons can be linked to some index, e.g. LIBOR, EURIBOR or inflation. Dates at which coupons are paid are called *coupon dates*. Coupons are typically paid semi-annually or annually.

Coupon bonds are sometimes used to create new zero-coupon bonds. In that case, financial institutions separate coupons of the bond from its principal. Then, coupon payments and principal can be traded separately. Zero-coupon bonds that originated from coupon bonds by stripping the coupon payments, are called *STRIPS*.

Bonds are used as a way of raising money. They can be issued by national governments, local governments or companies. *Government bonds*[1] are bonds issued by national governments. Government bonds are usually regarded as credit risk free[2], since central banks can print money and repay (inflate away) the debt[3]. Corporate defaults are much more probable, therefore corporate bonds are usually traded at a discount to goverment bonds. However, credit risk among countries and companies differs substantially depending on a variety of factors, among others on the economic conditions of the particular debtor. In this thesis, we will only work with credit risk free bonds.

2.2 Interest rates

We denote $P(t,T)$ the present value of a zero-coupon bond at time t paying out 1 unit[4] at time T (in $T-t$ years). The interest rate $R(t,T)$ that must be used to discount the face value of a zero-coupon bond to get the present value of this bond, i.e.

$$P(t,T) = e^{-R(t,T)(T-t)},$$

is called a *spot rate*[5] (or *zero-coupon rate*).

Instantaneous risk-free rate (or *short rate*) r_t is defined as the spot rate of a zero-coupon bond with time to maturity infinitesimally small, i.e.

$$r_t = \lim_{T \to t} R(t,T).$$

For a coupon bond with principal 1 and coupons paid annually at a rate c_j, $j = 1, 2, \dots\ T-t$ an interest rate $y(t,T)$ used to discount its future cash flows

$$P(t,T) = \sum_{j=1}^{T-t} c_j e^{-y(t,T)j} + e^{-y(t,T)(T-t)}$$

is referred to as the *gross redemption rate.*

[1]In the UK bonds are called gilts, in the USA short term bonds are called bills, medium term bonds are called notes and long term bonds are bonds.

[2]It means that there is no risk govenment could default on debt. Although these bonds are often regarded as credit risk free there have been a few examples when governments were unable to meet their obligations and defaulted on their debt.

[3]This, however, poses foreign exchange risk, which makes government bonds of some countries truly risky investments.

[4]A price of a zero-coupon bond with face value 1 is also called a *discount factor*.

[5]Spot rates can be calculated also from coupon bonds. The method to extract spot rates from coupon bonds is called bootstraping. For a comprehensive explanation of this method see e.g. Berec (2010).

Another type of interest rate that we need to understand before we proceed is the *forward rate.* Forward rate is related to a *forward contract.* This is a contract to buy or sell an asset at a particular time in the future at the price agreed on today. (We want to point out that signing such a contract does not give us an option to buy/sell the asset in the future, instead it gives us an obligation to do so.)

Consider a forward contract to buy a zero-coupon bond at time S with maturity at time T, $S < T$. Setting the price $P_t(S,T)$ of this bond today, at time t, is equivalent to setting the future spot rate $R(S,T)$ of this bond at time S, $t \leq S < T$, because $P(S,T) = e^{-R(S,T)(T-S)}$. The spot rate $R(S,T)$ which is contracted in advance, at time t, is called the forward rate and we denote it $f_t(S,T)$. As will be shown in chapter 4 the fair value of the forward rate, that does not disadvantage any party involved in the contract, is determined by spot rates implied by observed bond prices at time t. Therefore, forward rates can be thought to reflect investors' expectations about future spot rates.

Instantaneous forward rate $f(S)$ is the short rate contracted for S year future,

$$f(S) = \lim_{T \to t+S} f_t(t+S,T) = \lim_{T \to t+S} R(t+S,T) = r_{t+S}.$$

2.3 Term structure of interest rates

Spot rates of different zero-coupon bonds may vary for various reasons, e.g. financial position of their issuer, time to maturity of the bond. If we consider zero-coupon bonds issued by the same debtor then the relationship between the spot rates of these bonds and their time to maturity defines the *term structure of interest rates*[6]. The relationship between discount factors and their time to maturity is known as *discount curve.*

A cash flow of any fixed income security[7] can be viewed as a stream of zero-coupon bonds. This allows us to calculate the present value of such security as the sum of the present values of zero-coupon bonds of the appropriate face value and maturity. Equivalently, if we know the term structure of interest rates, we can price this security as the sum of its cash flows, each discounted with the appropriate spot rate[8].

[6]Some authors call term structure of interest rates also the relationship between coupon bond's gross redemption rate and its time to maturity. However, these do not have the same properties as spot rates.

[7]Fixed income security is such that yields a regular return.

[8]Note, that using the gross redemption rate of coupon bonds would not be appropriate to discount these cash flows.

Term structure of interest rates, especially if depicted as a continuous curve, is often called the *yield curve*. Yield curves are usually derived from prices of bonds that are observed at the market of government bonds of some country. That means that we can observe only those spot rates for which there exist appropriate bonds. Yields can be derived also from money market instruments (short term deposits and loans) but both approaches have in common that we cannot compute the whole continuous curve but only particular points at this curve.

However, investors at financial markets work with an immense number of assets that may mature or generate cash flow at different dates than the ones for which we observe yields. To be able to correctly price these assets, they need to know the spot rates at all possible maturities and therefore different methods for constructing continuous yield curve have been proposed. Cairns (2004) provides the following categorization of term sturcture models.

2.4 Interest rate models

One group of interest rate models forms the no-arbitrage approach. Models from this group describe evolution of the short rate as a stochastic process with time dependent parameters. These parameters are then calibrated so that the bond prices implied by the no-arbitrage principal fit exactly the prices of observed bonds at time of estimation. Among the most prominent models from this class are Ho and Lee (1986) and Heath-Jarrow-Morton framework, Heath et al. (1992). Drawback of this approach is that its time dynamics properties may imply evolution of interest rates in the way which is not empirically justified. However, since these models fit observed yields accurately, they are widely used, especially for pricing short term derivatives.

Equilibrium and short rate models, on the other hand, are not designed to fit the present curve precisely. These are parametric models that focus on describing either the impact of the economy on the yield curve (equilibrium models) or their goal is to capture the evolution of interest rates in time (short rate models). Parameters of these models are usually estimated from historic data since they are assumed to be constant. This, however, causes them not to fit the curve so well, although asset prices under these models also evolve in the arbitrage free way. The most famous models from this class were proposed by Vašíček (1977) and Cox, Ingersoll, Ross (henceforth CIR), Cox et al. (1985). Both of these models describe evolution of the short rate as a mean reverting, stochastic process. The process used by Vašíček is also known as Ornstein-Uhlenbeck process and the one used in CIR is known as Bessel square root process. Unfortunately, these models are primarily designed to predict instantaneous interest rate and show little success in modelling whole term structure.

Besides these two main approaches, there is also a group of descriptive models that try to identify the term structure of interest rates only by estimating parameters of some predetermined functional form for the yield curve. Their sole aim is to provide a good fit to the observed yields and they give us no information about the time dynamics of the yield curve. For monetary authorities these models help to analyse monetary policy. For bond traders, these models are used to identify overpriced or underpriced bonds (so called cheap/dear analysis). Model proposed by McCulloch (1975) uses cubic splines to fit the discount curve. Vašíček and Fong (1982) construct discout curve by using exponential splines. Very popular is the class of models that make up a forward curve as a combination of exponentials and polynomials. To this class of models belongs also the model proposed by Nelson and Siegel which is the model of our interest in this thesis.

Chapter 3

Model of Nelson and Siegel

3.1 Original model

Charles Nelson and Andrew Siegel wanted to find a yield curve model that would be simple but flexible enough to generate the range of shapes typically observed at yield curves. Motivated by the expectation theory, which claims that different maturities are perfect substitutes[1] and the shape of the yield curve reflects only expectations of future interest rates, in Nelson and Siegel (1987) they proposed parsimonious, 4 parameter model. The model was based on the assumption that instantaneous forward rate follows a second order differential equation

$$f''(\tau) + af'(\tau) + bf(\tau) = 0. \tag{3.1}$$

Solution[2] to the equation 3.1 is

$$f(\tau) = \beta_1 + \beta_2 e^{-\lambda\tau} + \beta_3 \lambda\tau e^{-\lambda\tau}. \tag{3.2}$$

Spot rate of a zero-coupon bond maturing in τ years is the expected yearly return of this bond, if held until maturity. If instantaneous forward rate, $f(s)$, is the expected short rate s years ahead then the required yield of a bond maturing in $T - t$ years must be equal to the average forward rate during this time horizont[3]

[1] Perfect substitutes in the sense that investors should be indifferent between investing in a T-year bond or in a S-year bond and in S years reinvest the money to (T-S)-year bond.

[2] This is a solution with real, equal roots. Nelson and Siegel found that solution with unequal roots produces model that appears to be overparametrized. That means that nearly the same quality of fit could be achieved by different combination of parameters and therefore, estimation of the model was troublesome.

[3] See e.g. Nawalkha et al. (2005) for derivation of this formula.

$$R(t,T) = \frac{1}{T-t}\int_0^{T-t} f(s)\mathrm{d}s.$$

Thus, integrating right hand side of the equation 3.2 from 0 to $T-t$ gives us the following expression for spot rates

$$\begin{aligned} R(t,T) &= \beta_1 + (\beta_2+\beta_3)\left[\frac{1-e^{-\lambda(T-t)}}{\lambda(T-t)}\right] - \beta_3 e^{-\lambda(T-t)} \\ &= b_1 + b_2\left[\frac{1-e^{-\lambda(T-t)}}{\lambda(T-t)}\right] + b_3 e^{-\lambda(T-t)}. \end{aligned}$$

This model soon became popular among market practitioners. Not only it captures the most common shapes of the yield curves (monotonic, humped, S shaped) but, unlike some sophisticated no-arbitrage models, this one is simple and tractable. For a fixed value of parameter λ the remaining 3 parameters can be estimated by Ordinary Least Squares (OLS) mathod. Diebol and Li (2006) proposed a reasonable value[4] of λ to be used for fitting the yield curve hence it is possible to estimate the model simply by OLS.

3.2 Diebold and Li modification

Altough the model of Nelson and Siegel is successful in fitting the term structure of interest rates at given date, it says nothing about evolution of the yield curve in time. Diebold and Li examined the forecasting performance of term structure models and they used Nelson-Siegel (NS) framework as their starting point. In Diebold and Li (2006) they rearranged the terms of the model (for computational reasons and better interpretation) and introduced dynamic structure to the betas of the original NS model, resulting in the following specification:

$$R(t,T) = \beta_{1t} + \beta_{2t}\left[\frac{1-e^{-\lambda(T-t)}}{\lambda(T-t)}\right] + \beta_{3t}\left[\frac{1-e^{-\lambda(T-t)}}{\lambda(T-t)} - e^{-\lambda(T-t)}\right]. \quad (3.3)$$

The betas are assumed to follow AR(1) process:

$$\beta_{i,t+h/t} = c_i + d_i\beta_{it} + v_i\epsilon_t^i, \quad i = 1,2,3,$$

[4]In Diebold and Li (2006) it is showed that if the time is measured in months, than $\lambda = 0.0609$ produces a good fit to data of U.S. Treasury security yields and there is little gain in model's fitting performacne if optimal value of λ is calculated for every estimation date.

where ϵ_t^i are normally distributed errors. Parameter λ is, for computational reasons, set fixed.

Loadings of parameters $\beta_{1t}, \beta_{2t}, \beta_{3t}$ led Diebold and Li to interpret them as latent (unobserved) factors: level, negative slope and curvature of the yield curve. Level of the yield curve is defined as long term yield $R(t, \infty)$, slope as $R(t, t+10years) - R(t, t+3months)$ (or sometimes $R(t, \infty) - R(t, t)$) and curvature as $2R(t, t+2years) - [R(t, t+3months) + R(t, t+10years)]$. Factor loadings are depicted below.

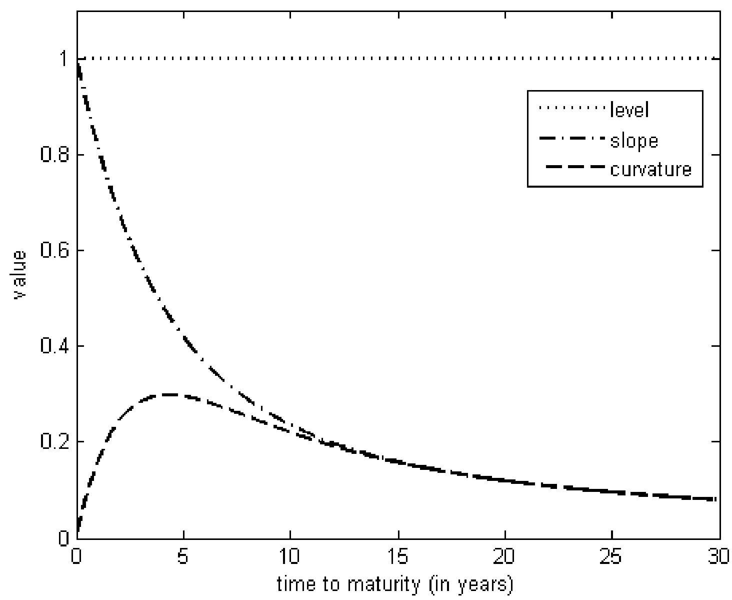

Figure 3.1: **Factor loadings of the Diebold-Li extension of the Nelson-Siegel model.**

We illustrate the justifiability of such an interpretation of beta factors on the figure 3.2.

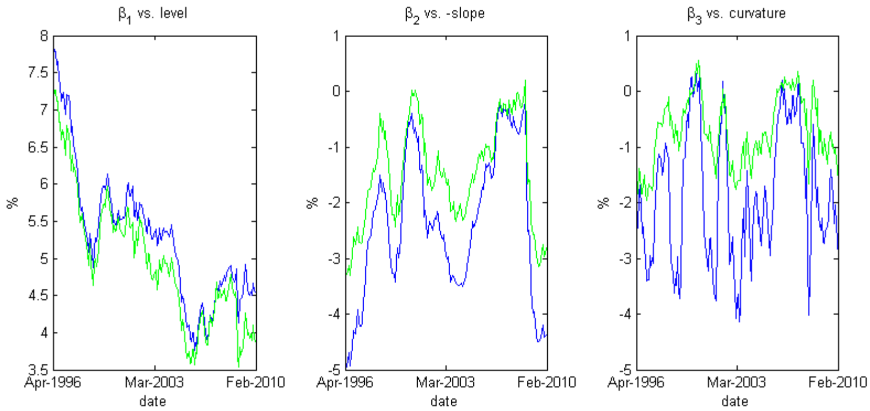

Figure 3.2: **Values of the estimated factors of the Nelson-Siegel model** for the data set of the spot rates used in this thesis. Time horizont is from April 1996 to February 2010. On the left picture, the blue line is the time series of the level factor β_{1t}, the green line is the time series of 30-year spot rate. Correlation coefficient between the changes in factor β_{1t} and the changes in the yield curve level is 0.9456. On the middle picture, the blue line is the factor β_{2t}, the green line depicts the difference between 3-month spot rate and 10-year spot rate. Correlation between the changes in β_{2t} and the changes in negative slope of the yield curve is 0.9075. The time series of β_{3t} (blue line) and $2y(2years) - [y(3months) + y(10years)]$ (green line) with correlation of the changes in these factors of 0.8003 is on the pictrue on the right.

The most important observation for our future analysis concerning the factor loadings is that the short rate is the sum of the first two factors, $r_t = \beta_{1t} + \beta_{2t}$. This can be easily verified, because β_{1t} has weight 1 in 3.3 for all maturities, loading for β_{2t} has limit 1 for $T \to t$ and weight of β_{3t} is 0 for $T \to t$.

Model 3.3 retains very good in-sample-fit performance, as in the original Nelson-Siegel model, and is also reported to beat the benchmarks such as random walk and AR(1) in out of sample predictions[5]. Diebold and Li dynamic version of the NS model quickly became popular both among asset management firms and central banks. However, the cross sectional shape of the curve is purely arbitrary. It has no connection to short rate dynamics[6] that would be needed to fit the present curve by using the no-arbitrage

[5]More details on forcasting performance and its comparison with different models can be found in Diebold and Li (2006).

[6]Filipovič (1999) shows that there exists no short rate model within HJM framework

principle. This is considered as the biggest shortcoming of this model, since no-arbitrage appears to be reasonable assumption because bond markets are very liquid. Moreover, derivative pricing relies essentially on the assumption of absence of arbitrage.

with constant volatility that would generate a spot rate curve of the shapes NS model does.

Chapter 4

Arbitrage

The no-arbitrage principle is the key concept of asset pricing. Market without arbitrage opportunities is such where individuals cannot gain higher than risk free rate of return without undergoing any risk. Equivalently, the probability of earning higher than risk free rate of return must be 0 for zero net investments.

4.1 Bond versus bank account

As an example of arbitrage, consider a riskless zero-coupon bond that pays 100 units in one year and a bank that will lend you or take a deposit from you both for 10% p.a. interest rate. If the market price of the bond was 92, one could easily make profit if she borrowed and sold (went short) the bond and put $100/1.1 = 90.91$ units on the bank account. The $92 - 90.91 = 1.09$ units is her immediate profit because after one year she will only have to pay the bondholder 100 units that will be accumulated on her bank account since $90.91 \cdot 1.1 = 100$. Similarily, if the price of the bond was only 90, one could borrow 90.91 from the bank and buy the bond. She will profit 0.91 because in one year the face value of the loan will be repaid by the bond's principal.

In a liquid bond market it would not be possible to find this bond priced in the way that would allow arbitrge opportunities. Even if such situation occured, it would disappear very quickly. The reason is that it would be immediately exploited by first smart traders. In the case of the dear bond, investors' effort to short sell these bonds would increase their supply on the market, thus pushing the price down. In the latter case, increased demand for the bond would make the price rise. Eventually, the price of the bond in this setting could not be different from 90.91. (Only presence of transaction costs might cause a little deviation from this number, nevertheless after accounting for these costs there would still not be room for making a riskless profit.)

4.2 Fair value of forward rate

We have already mentioned that forward rates are determined by observed spot rates. At this point we clarify this assertion. Consider a riskless 1-year zero-coupon bond with the spot rate 5% and a riskless 4-year zero-coupon bond with the spot rate 6% at time t. The price of the 1-year bond with nominal value 1 is $P_t(1) = e^{-0.05} = 0.95$ and the price of the 4-year bond is $P_t(4) = e^{-0.06 \cdot 4} = 0.79$.

The pricpiple of no-arbitrage implies a fair value of the forward rate $f_t(t+1, t+4)$. It is because we can make a 4-year investment in 2 ways. One is to buy the 4-year bond which yields 6% yearly return or accumulated 27% return ($1/0.79 = 1.27$) in 4 years. Another possibility is to invest in the 1-year bond and contract a 3-year spot rate one year ahead, i.e. $f_t(t+1, t+4)$. Accumulated return from this stretegy at time $t+4$ is equal to $1/0.95 \cdot e^{f_1(t+1,t+4)\cdot 3}$ and from the no-arbitrage principlne, it must be equal to 27%, which was the cumulative return from investing in the 4-year bond. This implies that $f_t(t+1, t+4) = \ln(1.27 \cdot 0.95)/3 = 0.0626$ or 6.26%.

How could we make a riskless profit if $f_t(t+1, t+4)$ had different value? If the forward rate was higher than 6.26%, say 6.5%, it would mean that investing in the one-year bond with spot rate 5% at time t and then investing at the rate $f_t(t+1, t+4)$ for 3-years at time $t+1$ would yield higher profit than investing in the 4-year bond at time t. Thus, to make a profit, we would invest in the strategy with higher return and we would borrow and sell the 4-year bond, which has lower return. The sold 4-year bond creates cash flow -1 at time $t = 4$ and $+0.79$ at time $t = 0$. To offset the 1 unit at time $t = 4$ we need to ensure cash inflow of the same amount at time $t = 4$. This can be done by investing $1 \cdot e^{-0.065 \cdot 3} = 0.82$ at the forward rate at time $t = 1$. To ensure this 0.82 at time $t = 1$ we can invest $0.82 \cdot e^{-0.05} = 0.78$ in the one-year bond at time $t = 0$. Having done this, we have offset our future liabilities (1 unit at time $t = 4$) by investing 0.78 units at time $t = 0$. Subtracting this amount from the 0.79 we have received for selling the 4-year bond results in the immediate riskless profit 0.01 at time $t = 0$. In the case of forward rate being lower than its fair value a riskless profit would be gained by buying the 4-year bond and selling the one year-bond and borrowing money in the forward contract.

It is important to understand that every two spot rates determine a value of the approptriate forward rate in a way that prevents earning a riskless profit by investing in these bonds and the forward contract. Hence, it is impossible to make an arbitrage profit by investing in a bond on the yield curve and contracting a forward rate, regardless of the actual shape of the term structure of interest rates.

4.3 Yield curve arbitrage

The expression "arbitrage on the yield curve" refers to the situation when there exists a bond whose spot rate does not lie on the market yield curve. By market yield curve is meant, that we can borrow and lend resources for the rates at this curve. Consider a situation on the picture bellow, where the star is the spot rate of a particular riskless zero-coupon bonds and the line represents the market yield curve.

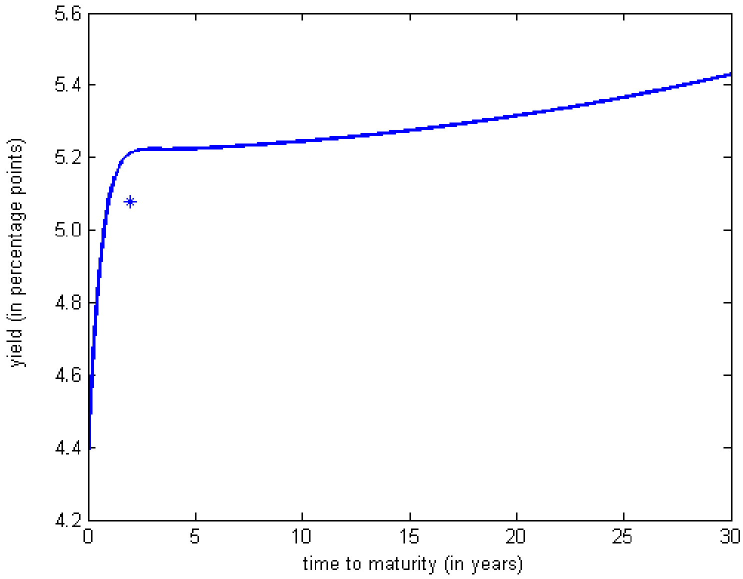

Figure 4.1: **Arbitrage on the yield curve.**

We can see that the bond maturing in 2 years has the spot rate below the market curve thus we can infer that this bond is overpriced. If the yields on the curve are $R(t, t+1) = 5.15\%$, $R(t, t+2) = 5.20\%$, $R(t, t+3) = 5.22\%$, then the corresponding bond prices are $P(1) = e^{-0.0515} = 0.9498$, $P(2) = e^{-0.0520 \cdot 2} = 0.9012$, $P(3) = e^{-0.0522 \cdot 3} = 0.8550$. The two-year bond marked with star has spot rate $R^*(t, t+2) = 5.08\%$ and the price is $P^*(2) = e^{-0,0508 \cdot 2} = 0.9034$. To make an arbitrage profit we would short the dear bond $P^*(2)$. By doing so we would get 0.9034 units at time t and we would need to pay 1 unit at time $t+2$. To offset this future payment, we can buy the cheaper two-year bond for 0.9012 and our immediate profit would be $0.9034 - 0.9012 = 0.0022$.

4.4 Arbitrage-free models

Now we explain what is the feature of the no-arbitrage models that assigns them their name. Consider the no-arbitrage model of Ho and Lee. The model describes evolution of the short rate by the stochastic differential equation

$$dr_t = \theta_t dt + \sigma dW_t.$$

If the price of a credit risk free zero-coupon bond evolves in the way that does not allow arbitrage opportunities, the holder of the bond should be indifferent between investing her money in the bond or putting them on a bank account because her wealth would rise at the same pace in both cases, i.e. $dP(t,T) = P(t,T)r_t dt$. More precisely, since the evolution of the short rate is not deterministic, we must account for the uncertainity factor, yielding to the following diffusion process for the no-arbitrage evolution of the bond price

$$dP(t,T) = P(t,T)(r_t dt + \sigma(t,T)dW_t^P), \tag{4.1}$$

where $\sigma(t,T)$ is volatility of this bonds price and W_t^P is a standard (with zero mean) Brownian motion (under real world measure P)[1].

Model 4.1 is still very inaccurate. Since this model expects that bond prices are influenced only by the short rate it implies that the spot rate cuvre should be, on average, quite flat. This is, however, in the contradiction with one stylized fact about yield curves, namely that their typical shape is upward sloping. There would be no room for such a rise in long term yields if they reflected only expectations of future yields, because future yield expecations on long horizonts are rather stable (see e.g. Tunc et al. (2009)).

Thus, there must be also some other factors than the short rate that determine yields of medium (between 1 and 10 years) and long (10 years and more) maturities. At this point, theories about what the other determinants are, differ. Liquidity preference theory says, that the longer the maturity is the higher is the risk of undesirable price moves, therefore investors require higher premiums over the short rate for long term investments[2]. Market

[1] A brief intoduction to the stochastic calculus and further refereces can be found in Appendix A.

[2] Liquidity preference theory accounts also for credit risk premium for bonds with longer maturities. However, here we consider credit risk free bond, hence with no credit risk.

segmentation theory says that investors preferences between purchasing short term or long term bonds differ with regard to the purpose of their investments and that resulting yields implied by prices of bonds are only a matter of demand and supply.

Whichever of the theories explaining the shape of term structure of interest rates is correct, it is a matter of fact that longer yields that we observe on the market are rarely equal to the short term rates. The difference between the yield of a bond maturing in τ years and the short rate is called the *risk premium*. This is the extra profit over the riskless rate of return that investors require to undertake a risk connected with the longer maturity of a bond. After deviding this risk premium by volatility of the bond price, we get the excess return required per unit of volatility, called the *market price of risk*.

Again, if we expressed the evolution of the price of a riskless bond in formal terms then we would need to use the actual yield of the bond in the drift term instead of the riskless rate in the equation 4.1. When we denote $\gamma(t) = \frac{R(t,T)-r_t}{\sigma(t,T)}$ the market price of risk[3] then we get

$$dP(t,T) = P(t,T)(r_t dt + \gamma(t)\sigma(t,T)dt + \sigma(t,T)dW_t^P). \tag{4.2}$$

It is importat to understand what infomation gives us equation 4.2. We know that price of a zero-coupon bond with maturity in $T - t$ years is, at time T, equal exactly to its principal (say 1). Hence, if we know also how the price evolved between the purchase of the bond at time t and its maturity at time T we can infer, what the price at time t should be. In other words, if we take the expected value of the bond price at time t it is

$$P(t,T) = E_P \left[exp \left(- \int_t^T r_s + \gamma(s)\sigma(s,T)ds \right) |\mathcal{F}_t \right],$$

where $\mathcal{F}_t$ is a filtration (information set available at time t). This can be compared with the market price we observe and if they are equal then the functions r_t and $\gamma(t)$ were specified correctly[4].

[3]Although it is not apparent from the formula the market price of risk does not depend on the time to maturity. Comprehensive explanation of this feature of γ can be found in Ševčovič, Stehlíková and Mikula (2009)[p.133].

[4]The implication that if the modelled and observed prices do not match then there is an arbitrage opportunity on the market, because our model is arbitrage-free, is not true. The model only ensures that the future term structure of interest rates implied by the present market data and model's dynamics will be consistent with the short rate evolution. The parameters of the Ho-Lee short rate model were estimated so that the modelled bond prices and actual prises are equal. Bond price dynamics consistent with the short rate

This is one possible way how to compute the term structure of interest rates. The difficult part here is how to find the function γ. There has been an extensive amount of literature written on this subject.

Alternative approach to find out what the expected prices of bonds are uses some advanced mathematical theory about martingales and change of measure. We provide a brief introduction into stochastic calculus in Appendix A and refer interested readers to Melicherčík et al. (2005) for an introduction and to Baxter and Rennie (1996) for a thorough explanation of the topic.

At this place we only state one important result, taken from Cairns (2004).

Theorem 4.4.1. *Let a diffusion processes of the short rate $r(t)$ and a zero-coupon bond price $P(t,T)$ (with principal 1) at time t maturing at time T follow stochastic differential equations*

$$\begin{aligned} dr(t) &= a(t)dt + b(t)dW(t)^P, && (4.3)\\ dP(t,T) &= P(t,T)\left[R(t,T)dt + \sigma(t,T)dW^P(t)\right], && (4.4) \end{aligned}$$

where $a(t)$, $b(t)$, $R(t,T)$, $\sigma(t,T)$ are foreseeable (potentionally stochastic) functions. There exists a measure Q equivalent to P such that the fair price of the zero-coupon bond defined by the above equations is

$$P(t,S) = E_Q\left[exp\left(-\int_t^S r(u)du\right)|\mathcal{F}_t\right], \tag{4.5}$$

where $t < S < T$,

$$dr(t) = (a(t) - \gamma(t)b(t))dt + b(t)dW^Q(t), \tag{4.6}$$

$\gamma(t) = \frac{R(t,T)-r(t)}{\sigma(t,T)}$, *$W^Q(t)$ is a standard Brownian motion under measure Q.*

Let us clarify the meaning and importance of this theorem. First, equivalency of measures Q and P is a mathematical term that refers to measures that have the same set of sets with measure 0. Second, and more important, expected value of a bond calculated under measure Q is the value that is already adjusted for the risk premium. (For this reason, P measure is usually called *real world* measure and the measure Q is often referred to as *risk neutral* or *risk adjusted*.) This means that equation 4.5 calculates the price of every bond with maturity in $S - t$ years at fixed time t. In other words,

model is such that after adjusting for the risk premium investors earn the same profit form holding the bond as if they put their money on a bank account. This kind of price dynamics is theoretically appealing but not allways empicrically justified.

equation 4.5 recovers the shape of the yield curve[5] at time t only by using diffusion process for the short rate and the no-arbitrage principle for bond price evolution. Moreover, the price of a bond under risk neutral measure follows

$$dP(t,S) = P(t,S)\left[r(t)dt + \sigma(t,S)dW^{Q}(t)\right], \tag{4.7}$$

where $r(t)$ evolves according to 4.6. This allows us to find the future no-arbitrage price $P(t+\Delta t, S-\Delta t)$ of a bond that costs $P(t,S)$ at time t and thus equations 4.6 and 4.7 ensure the no-arbitrage evolution of the yield curve implied by 4.5 at time t.

Specification of an interest rate model under risk neutral measure Q can be an advantage, since we do not need to specify a diffusion model for a risk premium. We still need to come up with a model for the short rate and this time the short rate model must be specified under the measure Q (i.e. adjusted for risk premium). Although this is also not trivial, it shows up to be often the more convenient way to specify an interest rate model than modelling the risk premium.

To conclude, there is no difference whether we describe underlying short rate model under real world measure and use the short rate plus the risk premium as the drift for evolution of a bond price, or we define the short rate model under risk adjusted measure and use the same short rate as the drift term in the stochastic differential equation for a bond price. Both of these approaches provide the same yield curve model that does not allow arbitrage opportunities. First, cross sectional arbitrage opportunies, in the sense that after adjusting for the risk premium there is no difference between yearly return of a 30 year zero-coupon bond and yearly return of a 1 year bond. Second, arbitrage opportunities in time, meaning that future yield curves implied by our model also evolved according to the no-arbitrage principle. This means that the risk adjusted price of a 10-year bond at time $t+1$ minus the risk adjusted price of this bond at time t (i.e. now 11-year bond) must earn the investor the same return as having her resources for one year on a bank account (with deposit rate equal to the current short rate.)

[5]Recall that the shape of the yield curve implied by the market prices can be allways referred to as arbitrage free.

Chapter 5

No-arbitrage Nelson-Siegel model

We are in a situation where we are quite satisfied with the performance of the Nelson-Siegel model of the term stucture. However, we take with discomfort the fact that it constructs the yield curve without regard to arbitrage free evolution of bond prices. This means that the model was not derived from a short rate model, thus the implied yield curve does not account for any known short rate dynamics. The question is, whether there exists a no-arbitrage term structure model that would imply factor loadings identical to the ones of Nelson and Siegel. Filipovič answered this question and his answer is 'not within HJM framework'. In Christensen et al. (2007) it is showed under what diffusion process for parameters of the NS model would the no-arbitrage derivation of the yield curve lead to Nelson-Siegel specification plus one additional term.

For the model that they call Arbitrage Free Dynamic Nelson Siegel they proposed the following[1] system of stochastic differential equations for betas under risk neutral measure Q:

$$\begin{pmatrix} d\beta_{1t} \\ d\beta_{2t} \\ d\beta_{3t} \end{pmatrix} = - \begin{pmatrix} 0 & 0 & 0 \\ 0 & \lambda & -\lambda \\ 0 & 0 & \lambda \end{pmatrix} \begin{pmatrix} \beta_{1t} \\ \beta_{2t} \\ \beta_{3t} \end{pmatrix} dt + \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{pmatrix} \begin{pmatrix} dW_t^{1,Q} \\ dW_t^{2,Q} \\ dW_t^{3,Q} \end{pmatrix} \qquad (5.1)$$

where $W_t^{1,Q}$, $W_t^{2,Q}$, $W_t^{3,Q}$ are indepentant standard Brownian motions. Then, using that short rate is an affine function of beta factors, namely $r_t = \beta_{1t} +$

[1]In fact, they allow the 3 Brownian motions to be correlated. However, we want to examine the forecasting performance of the model, therefore we use indepentent Brownian motions which are, as shown also in Christensen et al. (2007), more suitable for this purpose.

β_{2t}, and that to evolve in the argitrage-free way, bond price must follow $dP(t,T) = P(t,T)(r_t dt + dW_t^Q)$ it can be derived (see Christensen et al. (2007) or Appendix B in this thesis) that the resulting yield curve has the following form

$$\begin{aligned} R(t,T) &= \beta_{1t} + \beta_{2t}\left[\frac{1-e^{-\lambda(T-t)}}{\lambda(T-t)}\right] + \beta_{3t}\left[\frac{1-e^{-\lambda(T-t)}}{\lambda(T-t)} - e^{-\lambda(T-t)}\right] \\ &- \frac{C(T-t)}{T-t}, \end{aligned}$$

where, if we denote $\tau = T-t$,

$$\begin{aligned} \frac{C(\tau)}{\tau} &= \sigma_1^2\frac{\tau^2}{6} + \sigma_2^2\left[\frac{1}{2\lambda^2} - \frac{1-e^{-\lambda\tau}}{\lambda^3\tau} + \frac{1-e^{-2\lambda\tau}}{4\lambda^3\tau}\right] \\ &+ \sigma_3^2\left[\frac{1}{2\lambda^2} + \frac{e^{-\lambda\tau}}{\lambda^2} - \frac{\tau e^{-2\lambda\tau}}{4\lambda} - \frac{3e^{-2\lambda\tau}}{4\lambda^2} - \frac{2(1-e^{-\lambda\tau})}{\lambda^3\tau} + \frac{5(1-e^{-2\lambda\tau})}{8\lambda^3\tau}\right]. \end{aligned}$$

Dynamics of the beta coefficients specified under the risk neutral measure is sufficient to derive the cross sectional shape of a yield curve model that assumes no-arbitrage evolution of bond prices. Under 5.1, the resulting cross sectional specification of the spot curve is equal to the one of the NS model minus the adjustment term $\frac{C(\tau)}{\tau}$. The value of the adjustment term is constant in time and depends on time to maturity, coefficient λ that governs the mean reversion rate of β_{2t} to β_{3t}, and the volatility parameters σ_1, σ_2, σ_3.

To simulate the observed dynamics of the level, slope and curvature of the yield curve we can specify diffusion process for betas under the real world probability measure P. For reasons of siplicity and consistancy with Christensen et al. (2007) and Tunc et al. (2009) we use AR(1) process

$$\begin{pmatrix} d\beta_{1t} \\ d\beta_{2t} \\ d\beta_{3t} \end{pmatrix} = -\begin{pmatrix} K_1 & 0 & 0 \\ 0 & K_2 & 0 \\ 0 & 0 & K_3 \end{pmatrix}\begin{pmatrix} \theta_1 - \beta_{1t} \\ \theta_2 - \beta_{2t} \\ \theta_3 - \beta_{3t} \end{pmatrix} dt + \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{pmatrix}\begin{pmatrix} dW_t^{1,P} \\ dW_t^{2,P} \\ dW_t^{3,P} \end{pmatrix}, \quad (5.2)$$

where θ_1, θ_2, θ_3 are the mean reversion values and K_1, K_2, K_3 are the mean reversion rates of the betas under the real world measure.

Under this specification Christensen et al. (2007) report that the forecasting performance of the no-arbitrage Nelson-Siegel model has improved in comparison with the Diebold-Li dynamic version of the NS model.

Chapter 6

Empirical analysis

6.1 Data

In our analysis we use the end of month Bloomberg spot rates of euro-denominated fixed-rate bonds issued by the German government. The sample dates from April 1996 to February 2010 and contains spot rates for the following maturities: 3, 6 months, 1, 2, 3, 4, 5, 7, 8, 9, 10, 15, 20, 25, and 30 years. Table 6.1 provides some descriptive statistics of our data set.

maturity	mean	st.dev.	min	max	$\rho(1)$	$\rho(6)$	$\rho(12)$
3m	2.95	1.10	0.20	5.03	0.985	0.748	0.272
6m	3.01	1.08	0.41	5.12	0.984	0.739	0.292
1y	3.12	1.06	0.53	5.20	0.980	0.720	0.307
2y	3.32	0.96	0.94	5.24	0.971	0.693	0.350
3y	3.56	0.91	1.36	5.28	0.970	0.693	0.387
4y	3.78	0.87	1.81	5.32	0.968	0.693	0.412
5y	3.91	0.85	2.17	5.59	0.969	0.716	0.456
7y	4.22	0.84	2.71	6.19	0.972	0.774	0.550
8y	4.33	0.85	2.84	6.37	0.974	0.796	0.591
9y	4.38	0.85	2.86	6.51	0.974	0.804	0.597
10y	4.42	0.83	3.02	6.54	0.973	0.796	0.579
15y	4.75	0.78	3.44	6.91	0.976	0.830	0.650
20y	4.95	0.82	3.54	7.09	0.979	0.867	0.715
25y	5.02	0.85	3.61	7.20	0.981	0.876	0.746
30y	4.97	0.89	3.54	7.26	0.982	0.884	0.766

Table 6.1: **Descriptive statisctics of the data set.**

6.2 Estimation method

Our aim is to estimate the values of the 10 parameters [K_1, K_2, K_3, θ_1, θ_2, θ_3, σ_1, σ_2, σ_3, λ], introduced in chapter 5, that produce the best fit of the model to the data available at time of estimation.

Maximum likelihood estimation is a technique commonly used for this purpose. Parameters estimated by maximizing the likelihood function are such that no other parameter values would produce the observed data with higher probability than the maximum likelihood estimates. However, since the dimensionality of our problem is rather high, searching directly for the global maximum of the loglikelihood function may produce unstable results, which would reduce the trustworthiness of the conclusions based upon such results.

Thus, following Christensen et al. (2007) and Tunc et al. (2009), we employ Kalman filter technique to improve the stability of our results. The Kalman filter technique enables us to filter the observed spot rates from the noise produced by imperfections in measurements (e.g. bid/ask spread, liquidity problems) and to recover unobserved state variables (in our case the betas) from the measurement equation (i.e. Nelson-Siegel yield curve specification). Values of these state variables are also adjusted for noise that arises from their stochastic character. More details on the Kalman filter estimation procedure can be found in Appendix C.

We used Matlab® by The MathWorks^{TM} to solve all the optimization problems.

6.3 Parameters

In this section we present some observations concerning the optimal values of the 10 parameters of interest [K_1, K_2, K_3, θ_1, θ_2, θ_3, σ_1, σ_2, σ_3, λ] as well as their implications on the values of the factors beta. In the following we will refer to the Nelson-Siegel model without the no-arbitrage structure as NS model, and the model proposed in Christensen et al. (2007) will be called NANS model.

Recall that the parameters K_i determine the mean reversion rate of the beta factors and the parameters θ_i are the long term (mean reversion) values of the beta factors. Parameter λ governs the decay of the Nelson-Siegel factor loadings and in the NANS model it also enters in the adjustment term $\frac{C(\tau)}{\tau}$. Volatility parameters σ_i do not directly influence the shape of the NS model, they only appear in the Kalman filter estimation of the true values of the beta factors and thereby affect the estimated time dynamics of the betas.

Parameters σ_i do, however, influence the cross sectional shape of the NANS model through the adjustment term $-\frac{C(\tau)}{\tau}$. For illustration of the impact of this adjustment term on the shape of the yield curve at different maturities see figure 6.1.

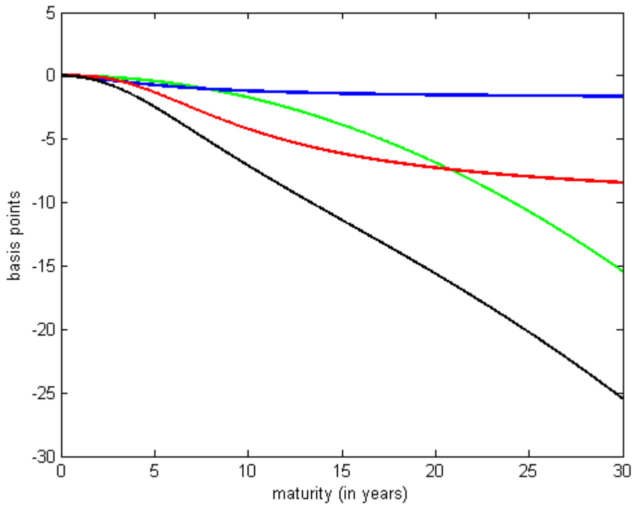

Figure 6.1: **The yield adjustment term** $-\frac{C(\tau)}{\tau}$ constructed using the optimal values of the parameters to fit the whole data set. The black line is the sum of the green, the blue and the red line and represents the effect of the adjustment term on the spot rates at different maturities. The green line represents the contribution of the term containing volatility σ_1 of the level factor β_{1t} on the value of the adjustment term. The blue line presents the contribution of the terms with σ_2 and the red line the contribution of σ_3 terms on the adjustment term.

Table 6.2 contains the optimal values of the parameters used to fit the whole data set. In the table 6.3, we provide descriptive statistics of the optimal parameter values that were estimated each month from March 2003 until February 2010 to best fit the data set available at that time. To get a better view on their evolution in time we plot also the time series of these parameters for both NS and NANS model on the figures 6.2, 6.3, 6.4.

We can say that despite some variance in the optimal values of the ten parametes there were no extreme deviations in the estimated values which suggests that the estimation algorithm was chosen wisely and produced stable results. Indeed, we have also performed an analysis of sensitivity of the solution to the starting point leading to the same conclusion of very solid

parameter	NS	NANS
K_1	0.06741	0.02375
K_2	0.14522	0.12811
K_3	1.33391	1.61848
θ_1	0.05853	0.06300
θ_2	-0.03492	-0.03912
θ_3	-0.01947	-0.01849
σ_1	0.00503	0.00321
σ_2	0.00846	0.00807
σ_3	0.01958	0.01954
λ	0.52218	0.42104

Table 6.2: **Optimal values of the parameters calculated to fit the whole data set.**

stability[1].

We start with a discussion about similarities and differences in the evolution of the parameters θ_i, depicted on the figure 6.3, extimated in the NS and the NANS model. Correlation of the values calculated for each model is apparent[2]. The difference in the levels of these parameters can be attributed to the presence of the $\frac{C(\tau)}{\tau}$ term in the NANS model. Since the adjustment term subtracts approximately 25 basis points from the value of the 30-year spot rate the parameter θ_1, which represents the mean reversion value of the level factor, must be in the NANS model shifted upwards to compensate for this adjustment. Similarily can be explained the shifts in the values of θ_2 and θ_3 because they are defined as $R(t,\infty) - R(t,t)$ and $2R(t,t+2years) - R(t,t+3months) + R(t,t+10years)$, respectively.

On the figure 6.3 we can see the evolution of the parameters K_i that govern the mean reversion rate of the factors β_{it}. The values calculated for the NS and the NANS model are again highly correlated[3]. This correlation can be clearly attributed to the fact that the only difference between the models is the presence of the adjustment term $\frac{C(\tau)}{\tau}$. The reason for the lower absolute values of the parameters K_1 and K_2 in the NANS model is less clear. One

[1] There were a minor stability issues when we allowed the parameters K_i to be negative. However, after restricting K_i to be positive (which is necessary for srationary processes), these issues disappeared.

[2] Correlation coefficient of the first differences are $\rho(\Delta\theta_1^{NS}, \Delta\theta_1^{NANS}) = 0.98$, $\rho(\Delta\theta_2^{NS}, \Delta\theta_2^{NANS}) = 0.99$, $\rho(\Delta\theta_3^{NS}, \Delta\theta_3^{NANS}) = 0.98$.

[3] Correlation coefficient of the first differences are $\rho(\Delta K_1^{NS}, \Delta K_1^{NANS}) = 0.95$, $\rho(\Delta K_2^{NS}, \Delta K_2^{NANS}) = 0.99$, $\rho(\Delta K_3^{NS}, \Delta K_3^{NANS}) = 0.41$.

	mean		st.dev.		min		max	
	NS	NANS	NS	NANS	NS	NANS	NS	NANS
K_1	0.074	0.021	0.028	0.006	0.041	0.012	0.138	0.033
K_2	0.135	0.116	0.045	0.041	0.058	0.051	0.192	0.165
K_3	1.247	1.509	0.131	0.152	0.967	1.196	1.415	1.697
θ_1	0.060	0.064	0.002	0.002	0.057	0.060	0.063	0.067
θ_2	−0.031	−0.034	0.004	0.004	−0.035	−0.039	−0.024	−0.028
θ_3	−0.021	−0.020	0.002	0.002	−0.024	−0.023	−0.018	−0.018
σ_1	0.005	0.003	0.000	0.000	0.005	−0.003	0.005	−0.003
σ_2	0.008	0.007	0.001	0.001	0.007	0.006	0.009	0.008
σ_3	0.020	0.020	0.001	0.001	0.018	0.019	0.021	0.022
λ	0.537	0.449	0.011	0.015	0.519	0.418	0.551	0.468

Table 6.3: **Descriptive statisctics of the estimated parameters.**

possible explanation is that imposition of the no-arbitrage structure on the Nelson-Siegel model improves the stability of the beta factors. This, however, does not seem plausible if we look at the figure 6.5 where the dynamics of the betas in both models suggests less persistent β_1^{NANS} than would imply the value of K_1^{NANS}[4]. Another explanation of the lower value of K_1 in the NANS model offers figure 6.4. The value of the parameter σ_1 in the NANS model is considerably lower than its value in the NS model. Besides lower volatility od the beta factors in the NANS model this could be also attributed to the presence of the σ_1 in the adjustment term $\frac{C(\tau)}{\tau}$. Higher values of σ_1 cause higher reduction in the long term yields which is usually not desirable. Hence, to avoid inappropriate deformation of the yield curve the value of the σ_1 is lower in the NANS model. By lowering σ_1, the variance of the innovation term in the state space equation for β_1 is reduced. This reduction in variance of innovations may not be desirable (recall figure 6.5), thus, the value of K_1 is reduced to boost this variance (see B.22 in Appendix B). Similar reasoning might be used to explain the differences in K_2^{NS} and K_2^{NANS} and in σ_2^{NS} and σ_2^{NANS}. The curvature factor is less persistent in the NANS model. (Note that the contribution of the σ_3 to the value of the $\frac{C(\tau)}{\tau}$ is very modest, hence there is no reason to suppress the value of σ_3.)

The values of the parameter λ in the two models again exhibit high positive correlation ($\rho(\Delta\lambda^{NS}, \Delta\lambda^{NANS}) = 0.76$) and downward shift in the values in the NANS model. Recall that λ governs the pace of decay of the Nelson-

[4]The averge value of $K_1^{NS} = 0.074$ implies half life of β_1^{NS} about 4 years. The average value of $K_1^{NANS} = 0.021$ implies half life of β_1^{NANS} more than 14 years.

Siegel loadings. Since the adjustment term $-\frac{C(\tau)}{\tau}$ in the NANS model has stronger negative effect on the yields on the longer horizont the yield curve produced by the NANS model must "rise" more than the NS curve before the adjustment term is added which can be achieved by lower values of λ.

The parameters $[K_1, K_2, K_3, \theta_1, \theta_2, \theta_3, \sigma_1, \sigma_2, \sigma_3, \lambda]$ were estimated in order to indentify the latent factors beta along with the optimal values of AR(1) coefficients that would describe their time dynamics. On the figure 6.5 we can see the evolution of the beta factors estimated from fitting the models to the whole data set. The correlation between the values for NS and NANS model is visually apparent.

We can see that the level factor β_{1t} in the NANS model evolves along with the level factor in the NS model shifted approximately for 30 basis points upwards. Interpretation of this relationship is straightforward. Since the NANS model contains the adjusment term $-\frac{C(\tau)}{\tau}$ which subtracts cca 25 basis points from the 30-year yields, the higher value of β_{1t} compensates for this adjustment.

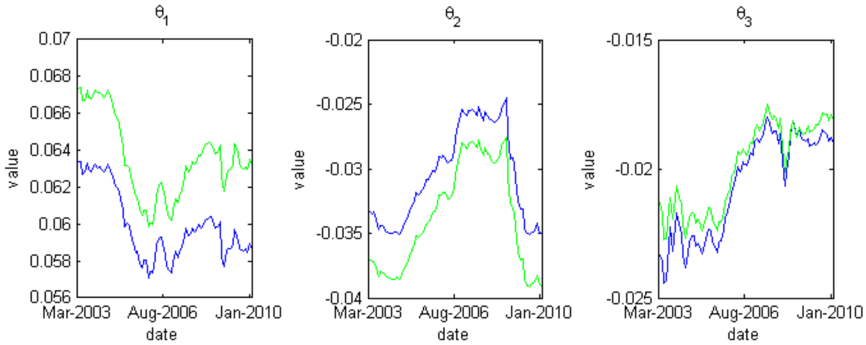

Figure 6.2: **Time series of the parameters θ_1, θ_2, θ_3.** The blue lines represents the NS and the green lines NANS estimates.

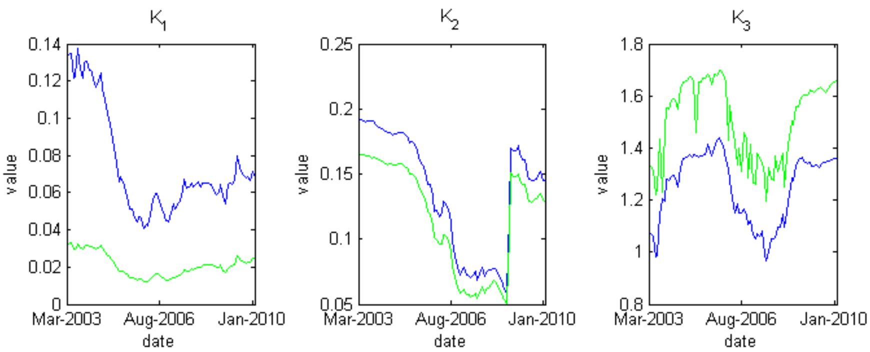

Figure 6.3: **Time series of the parameters K_1, K_2, K_3.** The blue lines represent the values in the NS model and the green lines are the estimates in the NANS model.

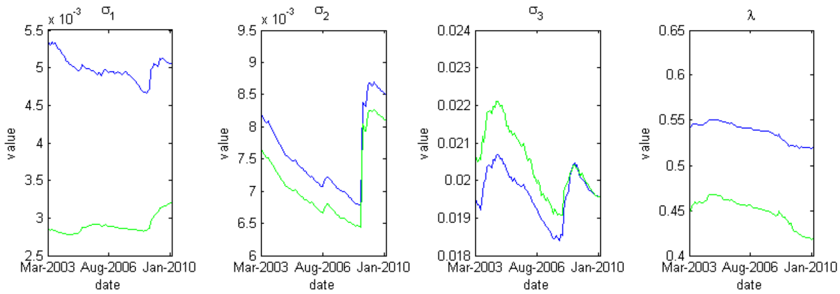

Figure 6.4: **Time series of the parameters σ_1, σ_2, σ_3, λ.** Blue lines stand for the NS estimates and green lines for the estimates in the NANS model.

The same relationship as between the level factors of the two models can be observed between the slope factors. Again, since the negative slope of the yield curve is defined as the short rate minus level, and the level factor of the NANS model is upward shifted relative to the level of the NS model, the β_{2t} of the NANS model is generally only a shifted value of the β_{2t} factor of the NS model.

The highest variance between the factor values estimated in the NS and NANS model can be seen in β_{3t}. Very high correlation between the two curvature factors is, however, still apparent.

An unpleasant observation about all the beta factors for both NS and NANS model is that they do not appear to be stationary. Indeed, if we look at augmented Dickey Fuller test statistics, preseted in the table 6.4, we cannot rejcet the null hypothesis of unit root process at five percent signifficance level for none of the betas. Recall that stationarity of the beta factors is necessary condition for our AR(1) modelling framework to be justified.

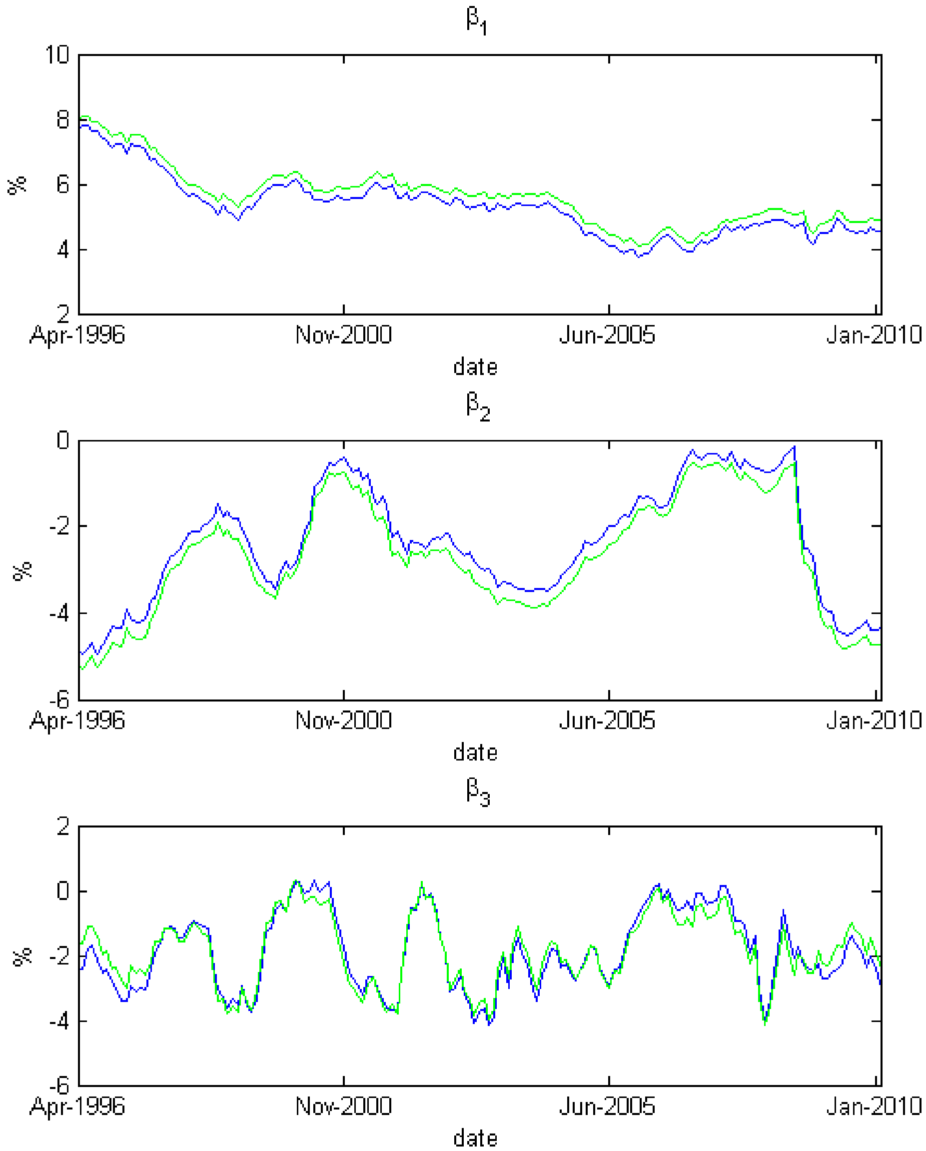

Figure 6.5: **Evolution of the estimated beta factors in time.**
The top blue lines are the estimate in the NS model, the green lines stand for the estimate in the NANS model.

	NS	NANS
β_1	-2.66	-2.25
β_2	-2.33	-2.39
β_3	-2.63	-2.59

Table 6.4: **Augmented Dickey Fuller test statictics** for the estimated time series of the beta factors . Critical value at the five pecrent signifficance level is -2.80.

6.4 In-sample fit

As mentioned in the previous section the estimated values of the parameteres beta were very similar for both models. This suggests that the models should produce similar fit to the data. Also a comparison of the final values of the loglikelihood functions, which was 13432 for the NS model and 13377 for the NANS, refers to a similar in sample fit performance with a slight advantage for the NS model. Another measure of the curve's fit to the data gives us the statistics Root Mean Squared Error (RMSE) which is defined as follows

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{y}_t)^2}, \tag{6.1}$$

where y_t denotes the actual spot rate and $\hat{y}_t$ denotes its counterpart on the modelled curve[5]. T is the number of dates for which we estimate the quality of the fit. We state the values of RMSE for spot rates of all the maturities in our data set in the table 6.4.

Also this measure of fit suggests that the NS model is better able to fit the curves in our data set. Only for the maturities of 5, 20 and 30 years has NANS model achieved better results.

Although the fitting preformance is generally very similar, there are situations where one of the models is better able to capture the shape of the curve. On the figure 6.6 we present examples of the fit at 3 different dates.

[5]In our case we have either $\hat{y}_t(\tau)^{NS} = \beta_{1t} + \beta_{2t}\left[\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right] + \beta_{3t}\left[\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right]$ or $\hat{y}_t(\tau)^{NANS} = \beta_{1t} + \beta_{2t}\left[\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right] + \beta_{3t}\left[\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right] - \frac{C(\tau)}{\tau}$ where τ denotes time to maturity and $\frac{C(\tau)}{\tau}$ is defined as ich chapter 5.

maturity	NS	NANS
3m	8.50	10.05
6m	4.14	4.17
1y	8.40	8.85
2y	7.49	9.14
3y	5.56	6.76
4y	4.51	5.37
5y	3.28	2.80
7y	5.03	5.44
8y	5.89	6.31
9y	7.44	7.64
10y	10.13	10.55
15y	10.26	10.94
20y	10.10	9.04
25y	8.79	8.88
30y	10.3	9.48

Table 6.5: **RMSE of the in-sample fit** (in basis points).

6.5 Forecasting performance

In this section we examine the forecasting performance of both the NS and the NANS model. As a measure of prediction quality we use the statistics of Root Mean Squared Forecasting Error (RMSFE) which is constructed as the ratio of RMSE produced by the forecast of some model to RMSE of the forecast using the random walk (forecast of no change).

We execute the forecasting exercise as follows. We estimate the models using data from April 1996 to March 2003 and perform predictions for one month, six month and twelve month horizont. Then we estimate the models using data from April 1996 to April 2003 and perform predictions. In every step we add one additional month to our data set and make predictions. Last one month predictions are made after models were estimated on the data from April 1996 to January 2010. Last 6 month predictions use data from April 1996 to August 2009 and last twelve month predictions use data form April 1996 to February 2009. Then we calculate the RMSFE for all the three prediction horizonts. Table 6.5 presents our results[6].

[6]We have also tried to employ the rolling horizont forecasting approach, where the parameters of the model would be estimated from a fixed number of dates before the forecasting date. The additional volatility in estimated parameters has not helped to improve the results.

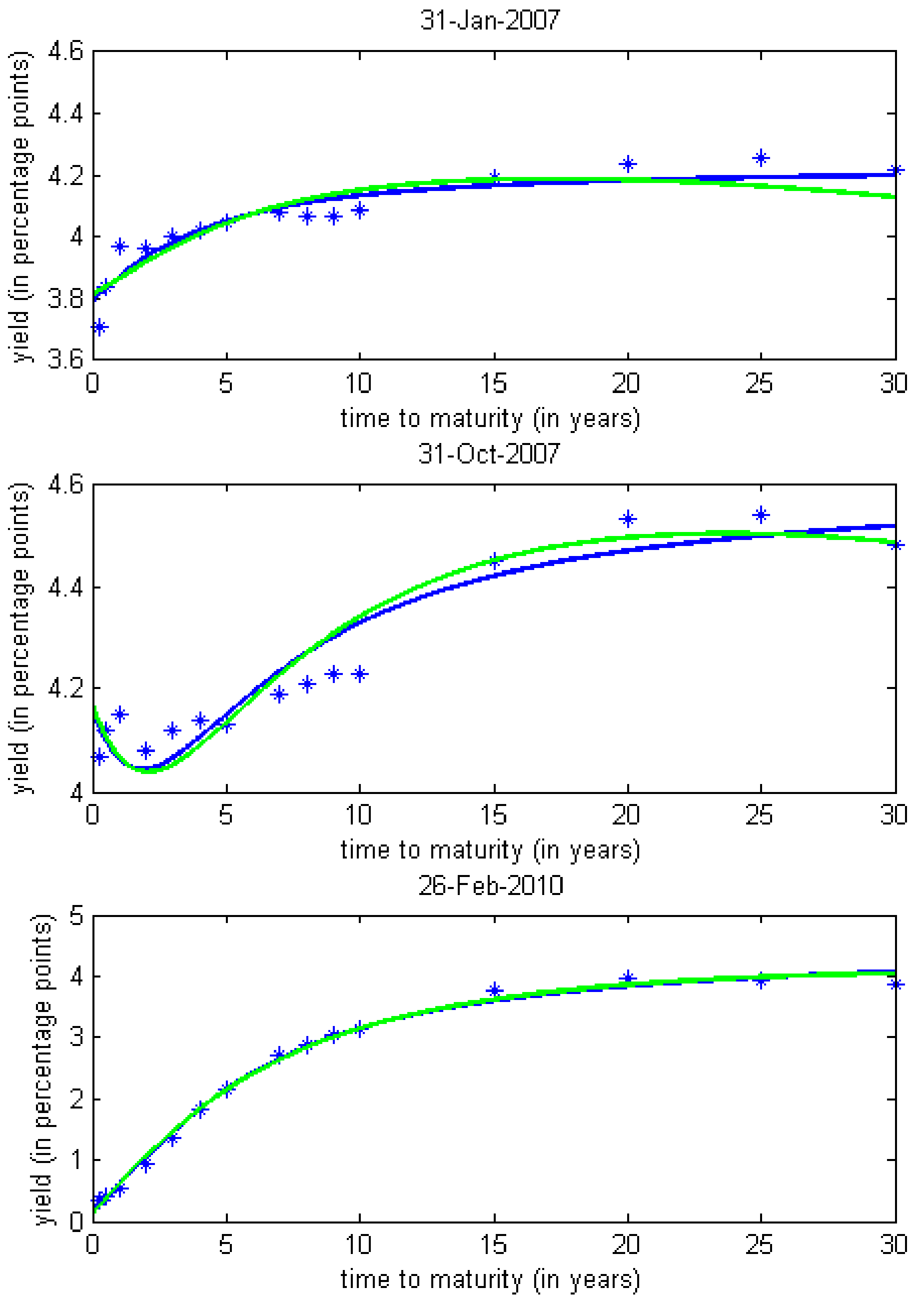

Figure 6.6: **Examples of the yield curve fit.** The blue line represents the NS curve, the green line is the curve fitted by the NANS model. Asterisks are data on spot rates.

matu-	1 month forecast			6 month forecast			12 month forecast		
rity	RW	NS	NANS	RW	NS	NANS	RW	NS	NANS
3m	20.98	1.08	1.08	89.31	1.03	1.03	147.17	1.02	**1.01**
6m	20.60	1.07	**1.06**	87.89	1.03	**1.02**	142.10	1.02	1.02
1y	22.10	**1.13**	1.14	87.92	1.01	1.01	135.78	1.02	1.02
2y	24.39	**1.03**	1.04	78.58	1.02	1.02	111.61	**1.06**	1.07
3y	22.67	1.04	1.04	72.04	1.01	1.01	98.87	1.06	1.06
4y	22.33	**1.01**	1.02	65.75	0.99	0.99	87.09	**1.05**	1.06
5y	21.77	1.02	**1.01**	61.31	1.02	**1.01**	79.80	1.09	1.09
7y	19.87	1.00	1.00	51.04	1.03	**1.02**	66.59	1.10	**1.09**
8y	19.14	1.00	**0.99**	48.92	1.04	**1.03**	63.72	1.11	**1.09**
9y	18.68	1.04	1.04	47.36	1.06	**1.04**	61.69	1.13	**1.11**
10y	18.45	1.06	1.06	45.28	1.07	**1.06**	57.96	1.16	**1.14**
15y	16.65	1.07	**1.05**	39.60	1.01	1.01	52.18	1.09	**1.07**
20y	15.68	1.19	**1.16**	37.63	1.03	1.03	52.46	1.07	**1.06**
25y	15.57	**1.07**	1.09	38.89	1.03	1.03	53.07	1.09	**1.07**
30y	16.27	1.28	**1.23**	41.86	1.09	**1.06**	56.34	1.14	**1.09**

Table 6.6: **Forecsting performance.** Columns marked RW contain RMSE statistics of the random walk predictions on the one month, six month or twelve month prediction horizont. Columns NS and NANS contain RMSFE statistincs for the out-of-sample predictions on the particular time horizont. In bold is marked the better result of the NS and the NANS.

Table 6.5 provides 2 things worth noting. One is that there is only minor difference between the forecasting performance of the two models suggesting a little dominance of the no-arbitrage version of the Nelson-Siegel model. The second, more striking observation is, however, that the values in the table 6.5 are mainly higher than 1. This reflects the fact that the RMSE produced by forecasting no change in interest rates is a more successful way of making predictions than using either of the models. This is a truly surprising result because Christensen et al. (2007) interpret their findings about better forecasting performance of the NANS model with comparison to the NS as success and they do not provide the comparison with random walk forecasts[7].

[7]We have also examined, whether our conclusions regarding the forecastig performance of the NS and the NANS model were affected by the situation during the financial crisis. Random walk forecasts were more successful than the NS or the NANS forecasts also after September 2008 but with less striking dominance.

Chapter 7

Conclusions

In this master thesis we implemented a recently developed no-arbitrage version of the well-known Nelson-Siegel term structure model on the data of German government bond spot rates. Most of the literature examines qualities of the model only on data on U.S. Treasury security yields and the results are then often (from our viewpoint recklessly) adoped by institutions in other countries.

The no-arbitrage model that we estimate belongs to the class of affine arbitrage-free term structure models. Following Christensen et al. (2007) we assume that the factors level, slope and curvature are first order autoregressive processes. Our empirical results suggests that this assumption might be inappropriate.

Our in-sample fit results are similar to those reported in Christensen et al. (2007). We conclude that there was no improvement in the fitting performance gained by imposing the no-arbitrage structure with uncorrelated factors to the original Nelson-Siegel model.

Our interpretation of the out-of sample forecasting performance is, however, different. Although we find that the no-arbitrage version of the Nelson-Siegel model produces slightly better forecasts than the dynamic version of the original Nelson-Siegel model we cannot claim this success because these forecasts are worse than random walk predictions. Christensen et al. (2007) compares only the forecasting performance of the two Nelson-Siegel models but omits to provide comparison with the random walk forecasts. Coroneo et al. (2008), paper from the European Central Bank, does offer such a comparison (although with a different approach to build the no-arbitrage Nelson-Siegel model) but provides only results estimated on the data sample of U.S. Treasury security yields from 1970 to 2000, which are neither current nor really relevant for the primary region of activity of this institution.

The failure of the no-arbitrage Nelson-Siegel to beat the random walk in forecasting can be, in our view, attributed to incorrect specification of the dynamics of the factors level, slope and curvature. Our results suggest that these factors are unit root processes which cannot be captured by the affine specification of the real world dynamics of betas as constructed in Christensen et al. (2007). Our suggestion for further research would be to find the diffusion process under the real world measure that would allow us to model the betas as first order integrated processes. Then would be needed to examine whether there exists the market price of risk which would allow the change from the real world to the risk neutral measure under which the dynamics of the betas is affine.

Bibliography

Ang, A. and Piazzesi, M. (2003). A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables, *Journal of Monetary Economics* **50**: 745–787.

Baxter, M. and Rennie, A. (1996). *Financial Calculus: An Introduction to Derivative Pricing*, Cambridge University Press.

Berec, R. (2010). *Extraction of Nelson-Siegel Factors from Bond Prices*, Master thesis.

Bishop, G. and Welch, G. (2006). An introduction to the kalman filter.

Björk, T. and Christensen, B., J. (1999). Interest rate dynamics and consistent forward rate curves, *Mathematical Finance* **9**: 323–348.

Cairns, A. J. G. (2004). *Interest Rate Models. An Introduction*, Princeton University Press.

Christensen, J., Diebold, F. and Rudebusch, G. (2007). The affine arbitrage-free class of nelson-siegel term structure models, *NBER working paper series* **13611**.

Christensen, J., Diebold, F. and Rudebusch, G. (2008). An arbitrage-free generalized nelson-siegel term structure model, *NBER working paper series* **14463**.

Coroneo, L., Nyholm, K. and Vidova-Koleva, R. (2008). How arbitrage-free is the nelson-siegel model, *European Central Bank working paper series* **874**.

Cox, J., Ingersoll, J. and Ross, S. (1985). A theory of the term structure of interest rates, *Econometrica* **53**: 385–408.

Diebold, F. and Li, C. (2006). Forecasting the term structure of government bond yields, *Journal of Econometrics* **130**: 337–364.

Diebold, F., X., Rudebusch, G., D. and Aurora, S., B. (2006). The macroeconomy and the yield curve: a dynamic latent factor approach, *Journal of Econometrics* **131**: 309–338.

Duffie, D. and Kan, R. (1996). A yield-factor model of interest rates, *Mathematical Finance* **6**: 379–409.

Favero, C., Niu, L. and Sala, L. (2007). Term structure forecasting. no-arbitrage restrictions vs. large information set.

Filipovič, D. (1999). A note on the nelson-siegel family, *Mathematical Finance* **9**: 341–350.

Heath, D., Jarrow, R. and Morton, A. (1992). Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation, *Econometrica* **60**: 77–105.

Ho, S. and Lee, S. (1986). Term structure movements and pricing interest rate contingent claims, *The Journal of Finance* **41**(5): 1011–1029.

Janková, K. (n.d.). Poznámky k prednáškam pravdepodobnosť a štatistika 1-mat-180.

Kotian, J. (1999). *Vyhľadávanie arbitrážnych príležitostí na výnosovej krivke*, Master thesis.

La Grand, F. (2007). Nelson and siegel, no-arbitrage and risk premium, *Job Market Paper* **1**.

McCulloch, J., H. (1975). The tax adjusted yield curve, *Journal of Finance* **30**: 811–830.

Melicherčík, I., Olšarová, L. and Úradníček, V. (2005). *Kapitoly z finančnej matematiky*, Epos.

Nawalkha, S., K., Soto, G., M. and Beliaeva, N., A. (2005). *Interest Rate Risk Modeling: The Fixed Income Valuation Course*, John Wiley&Sons, Inc.

Nelson, C. and Siegel, A. (1987). Parsimonious modeling of yield curves, *Journal of Business* **60**: 473–489.

Sutóris, I. (2009). *Odhady parametrov modelov časovej štruktúry úrokových mier.*, Master thesis.

Tunc, J., Ungari, S., Huang, C. and Kurpiel, A. (2009). Filtering the interest rate curve. the menir framework., *Societe Generale Cross Asset Research* .

Vašíček, O. (1977). An equilibrium characterization of the term structure, *Journal of Financial Economics* **5**: 177–188.

Vašíček, O. and Fong, H., G. (1982). Term structure modeling using exponential splines, *Journal of Finance* **37**: 339–348.
ΩŠevčovič et al.

Ševčovič, D., Stehlíková, B. and Mikula, K. (2009). *Analytické a numerické metódy oceňovania finančných derivátov*, Nakladateľstvo STU.

Appendix A

Introduction to stochastic calculus

At this place we provide[1] a brief introduction to the terminology of stochastic calculus. We begin with a few definitions from basic course in probability and statistics.

Definition A.0.1. *(sigma algebra) Let* $\Omega \neq \emptyset$. *A subset* $\mathcal{F}$ *of a set* Ω *is called a* σ*-algebra if it has the following properties*
(i) $\emptyset \in \mathcal{F}$,
(ii) If $A \in \mathcal{F}$ *then* $A^C \in \mathcal{F}$,
(iii) If $A_i \in \mathcal{F}$ $i = 1, 2, ...$ *then* $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

If Ω is a given set, $\mathcal{F}$ is a σ-algebra of measurable sets on Ω and P is a probability measure on Ω, then the tripple $(\Omega,\ \mathcal{F},\ P)$ is called a probability space.

Definition A.0.2. *A random variable is an arbitrary function* $X : \Omega \to \mathbb{R}$ *such that for every* $x \in \mathbb{R} : \{\omega \in \Omega : X(\omega) < x\} \in \mathcal{F}$.

Definition A.0.3. *Stochastic process is a set of random variables* $X = \{X_t; 0 \leq t < \infty\}$ *on a propability space* $(\Omega,\ \mathcal{F},\ P)$ *with values is* $\mathbb{R}^d$. *For every* t *is* $\omega \to X_t(\omega)$; $\omega \in \Omega$ *a random variable. If we fix* $\omega \in \Omega$, *we get a function* $t \to X_t(\omega)$; $0 \leq t < \infty$ *which is called a trajectory of* X *assigned to* ω.

Next we define a special type of stochastic process which plays a vital part in stochastic calculus.

[1]Sources were Melicherčík et al. (2005), Janková (n.d.)

Definition A.0.4. *A standard Brownian motion is a stochastic process with the following properties:*
(i) trajectories $W_t(\omega)$ are continuous with probability 1 and $W_0 = 0$
(ii) a random variable W_t is normally distributed $N(0, t)$
(iii) $W_{t+s} - W_s$ has distribution $N(0, t)$. Moreover, increments of W_t are independent, i.e. W_{t_1}, $W_{t_2} - W_{t_1}$,..., $W_{t_k} - W_{t_k-1}$ are independent for all $0 \leq t_1 < ... < t_k$.

By Brownian motion is in some literature meant the process $B_t = \mu t + \sigma W_t$. Then, W_t as defined in A.0.4 is called a Wiener process. For our purposes, however, we do not need to distinguish these two, thus we follow the definition A.0.4 from Melicherčík et al. (2005).

Definition A.0.5. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, $X : \Omega \to \mathbb{R}$ a random variable for which $E(|X|) < \infty$. Let $\mathcal{H} \subset \mathcal{F}$ be a σ-algebra. Conditional mean $E(X|\mathcal{H})$ is a random variable with the following properties:*
(i) $E(X|\mathcal{H})$ is $\mathcal{H}$-measurable
(ii) $\int_H E(X|\mathcal{H})dP = \int_H XdP \ \forall \ H \in \mathcal{H}$.

Definition A.0.6. *Let $\{N_t\}_{t\geq 0}$ be a growing system of σ-algebras on Ω (i.e. $N_{t_2} \supset N_{t_1}$ for $t_2 > t_1$). Stochastic process $g(t, \omega) : [0, \infty) \times \Omega \to \mathbb{R}$ is N_t-adapted if for every $t \geq 0$ is a function $\omega \to g(t, \omega)$ N_t-measurable.*

N_t-adaptation means that the value of a function g is not computed using information that will be known to us only in the future.

Definition A.0.7. *Let $W_t(\omega)$ be a Brownian motion on $(\Omega, \mathcal{F}, P)$. One-dimensional Itô process is a stochastic process of the form*

$$X_t = X_0 + \int_0^t u(s, \omega)ds + \int_0^t v(s, \omega)dW_x(\omega)$$

or in differential form

$$dX_t(\omega) = u(t, \omega)dt + v(t, \omega)dW_t(\omega),$$

where $u(t, \omega)$, $v(t, \omega)$ are $\mathcal{F}_t^W$-adapted. Function u is called a drift and function v is called a volatility.

Functions $u(t, \omega)$, $v(t, \omega)$ in the above definition must have also some additional properties, but we do not state them here, because they have rather technical character.

Definition A.0.8. *Filtration on a probability space $(\Omega, \mathcal{F}, P)$ is called a system of σ-algebras $\{M_t\}_{t\geq 0}$, $M_t \subset \mathcal{F}$ such that $0 \leq s < t \Rightarrow M_s \subset M_t$.*

Filtration is a notion that stands for an information set available to us at time t. The condition that filtration "grows" as time proceeds means that we do not forget any past information and every additional time step gives us some new information.

Definition A.0.9. *Let P, Q be probability measures on $(\Omega, \mathcal{F})$. P and Q will be called equivalent if $P(A) > 0 \Leftrightarrow Q(A) > 0$.*

Lemma A.0.1. *Let $X_t(\omega)$ be an Itô's process*

$$dX_t(\omega) = u(t,\omega)dt + v(t,\omega)dW_t(\omega).$$

Let $g(t,x) \in C^2([0,\infty) \times \mathbb{R})$. Then $Y_t(\omega) = g(t, X_t(\omega))$ is also Itô's process and

$$dY_t = \frac{\partial g}{\partial t}(t, X_t)dt + \frac{\partial g}{\partial x}(t, X_t)dX_t + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(t, X_t)v^2 dt.$$

Appendix B

Derivation of a no-arbitrage Nelson-Siegel model

A thorough derivation of the no-arbitrage Nelson-Siegel model is described in Christensen et al. (2007). Since it is rather complex, we provide here only an outline of the derivation. Duffie and Kan (1996) prove that if the instantenous risk-free rate is an affine function of state variables β_t and the diffusion process for these state variables under the risk neutral measure Q is of a particular form[1] then zero-coupon bond prices, derived from the no-arbitrage principle, are *exponential-affine* functions of the state variables

$$P(t,T) = E_t^Q[exp(-\int_t^T r_u du)] = exp(B(t,T)'\beta_t + C(t,T)). \tag{B.1}$$

The first equality in B.1 represents the no-arbitrage pricing approach and the second equality, where $B(t,T)$ and $C(t,T)$ are the solutions to a particular system of differential equations, is proved in Duffie and Kan (1996).

Since prices of zero-coupon bonds can be also computed as

$$P(t,T) = exp(-R(t,T)(T-t))$$

we get that

$$R(t,T) = -\frac{1}{T-t}\log P_t(T) = -\frac{B(t,T)'\beta_t}{T-t} - \frac{C(t,T)}{T-t}. \tag{B.2}$$

[1] Affine arbitrage free class of term structure models.

Hence, if we want to obtain a specification of the spot rates similar to that of Nelson and Siegel, it must be the case that

$$-\frac{B_1(t,T)}{T-t} = 1, \tag{B.3}$$

$$-\frac{B_2(t,T)}{T-t} = \frac{1-e^{-\lambda(T-t)}}{\lambda(T-t)}, \tag{B.4}$$

$$-\frac{B_3(t,T)}{T-t} = \frac{1-e^{-\lambda(T-t)}}{\lambda(T-t)} - e^{-\lambda(T-t)}. \tag{B.5}$$

Multiplaying equations B.3, B.4, B.5 by $(t-T)$ and differentiating with respect to t we get

$$\frac{dB_1(t,T)}{dt} = 1, \tag{B.6}$$

$$\frac{dB_2(t,T)}{dt} = e^{-\lambda(T-t)} = 1 + \lambda B_2(t,T), \tag{B.7}$$

$$\frac{dB_3(t,T)}{dt} = \lambda\tau e^{-\lambda(T-t)} = -\lambda B_2(t,T) + \lambda B_3(t,T). \tag{B.8}$$

Because of B.2, we need the functions $B_i(t,T)$ to satisfy final conditions $B_1(T,T) = B_2(T,T) = B_3(T,T) = 0$.

Since we want the bond prices to evolve in an arbitrage free way a price of a zero-coupon bond under risk neutral measure Q must follow

$$dP(t,T) = P(t,T)[r_t dt + \sigma_P dW^Q]. \tag{B.9}$$

In the Nelson-Siegel model, instantaneous risk-free rate is the sum of the first two factors $r_t = \beta_{1t} + \beta_{2t}$. If we can find a diffusion process for factors $\beta_{1t}, \beta_{2t}, \beta_{3t}$ such that diffusion of $P(t,T)$ calculated from B.1 with $B(t,T)$ satisfying B.6, B.7, B.8 will coincide with B.9 then the spot rates given by B.2 will have Nelson-Siegel factor loadings and the cross sectional shape and dynamics of $R(t,T)$ will be consistent with the no-arbitrage principle.

We try to find the diffusion process for betas as first order vector autoregression process with independent innovations[2].

$$\begin{pmatrix} d\beta_{1t} \\ d\beta_{2t} \\ d\beta_{3t} \end{pmatrix} = -\begin{pmatrix} K_{11} & K_{12} & K_{13} \\ K_{21} & K_{22} & K_{23} \\ K_{31} & K_{32} & K_{33} \end{pmatrix} \begin{pmatrix} \beta_{1t} \\ \beta_{2t} \\ \beta_{3t} \end{pmatrix} dt + \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{pmatrix} \begin{pmatrix} dW_t^{1,Q} \\ dW_t^{2,Q} \\ dW_t^{3,Q} \end{pmatrix} \tag{B.10}$$

[2] We restrict the innovations to be independent only for simplicity

This is a system that does belong to the group of diffusion processes described in Duffie and Kan (1996) to produce affine arbitrage free models thus the pricing equation B.1 holds. Differentiating B.1 and using B.10 as diffusion for factors β_{it} yields

$$\begin{aligned}
dP(t,T) = P(t,T)[dC &+ dB_1\beta_{1t} - B_1K_{1\cdot}\beta_t dt + \frac{1}{2}B_1^2\sigma_1^2 dt + \\
&+ dB_2\beta_{2t} - B_2K_{2\cdot}\beta_t dt + \frac{1}{2}B_2^2\sigma_2^2 dt + \\
&+ dB_3\beta_{3t} - B_3K_{3\cdot}\beta_t dt + \frac{1}{2}B_3^2\sigma_3^2 dt + \\
&+ B_1\sigma_1 dW_1^Q + B_2\sigma_2 dW_2^Q + B_3\sigma_3 dW_3^Q].
\end{aligned}$$

Arbitrage free evolution of bond prices implies also B.9. By inserting $\beta_{1t}+\beta_{2t}$ for r_t in B.9 we can now match the coefficients in front of $\beta_{it}dt$

$$\begin{aligned}
dB_1\beta_{1t} - B_1K_{11}\beta_{1t}dt - B_2K_{21}\beta_{1t}dt - B_3K_{31}\beta_{1t}dt &= \beta_{1t}dt & \text{(B.11)}\\
dB_2\beta_{2t} - B_1K_{12}\beta_{2t}dt - B_2K_{22}\beta_{2t}dt - B_3K_{32}\beta_{2t}dt &= \beta_{2t}dt & \text{(B.12)}\\
dB_3\beta_{3t} - B_1K_{13}\beta_{3t}dt - B_2K_{23}\beta_{3t}dt - B_3K_{33}\beta_{3t}dt &= 0 & \text{(B.13)}\\
dC + \frac{1}{2}B_1^2\sigma_1^2 dt + \frac{1}{2}B_2^2\sigma_2^2 dt + \frac{1}{2}B_3^2\sigma_3^2 dt &= 0 & \text{(B.14)}
\end{aligned}$$

Since we know that $B_1(t,T)$, $B_2(t,T)$, $B_3(t,T)$ are not identical (B.3, B.4, B.5) and we want them to follow B.6, B.7, B.8 we can identify the unique values of K_{ij} so that our conditions are satisied, namely

$K_{11} = 0, K_{12} = 0, K_{13} = 0, K_{11} = 0, K_{22} = \lambda, K_{23} = -\lambda, K_{31} = 0, K_{32} = 0, K_{33} = \lambda.$

Solving equation B.14 with final condition $C(T,T) = 0$ gives us an expression for the yield adjustment term $C(t,T)$. The value of this term does not evolve over time, thus if we denote $\tau = T - t$ a time to maturity, we get

$$\begin{aligned}
\frac{C(\tau)}{\tau} &= \sigma_1^2\frac{\tau^2}{6} + \sigma_2^2\left[\frac{1}{2\lambda^2} - \frac{1-e^{-\lambda\tau}}{\lambda^3\tau} + \frac{1-e^{-2\lambda\tau}}{4\lambda^3\tau}\right]\\
&+ \sigma_3^2\left[\frac{1}{2\lambda^2} + \frac{e^{-\lambda\tau}}{\lambda^2} - \frac{\tau e^{-2\lambda\tau}}{4\lambda} - \frac{3e^{-2\lambda\tau}}{4\lambda^2} - \frac{2(1-e^{-\lambda\tau})}{\lambda^3\tau} + \frac{5(1-e^{-2\lambda\tau})}{8\lambda^3\tau}\right].
\end{aligned}$$

The conclusion is that if the diffusion process for factors β_{it} under risk neutral measure Q is specified as

$$\begin{pmatrix} d\beta_{1t} \\ d\beta_{2t} \\ d\beta_{3t} \end{pmatrix} = -\begin{pmatrix} 0 & 0 & 0 \\ 0 & \lambda & -\lambda \\ 0 & 0 & \lambda \end{pmatrix} \begin{pmatrix} \beta_{1t} \\ \beta_{2t} \\ \beta_{3t} \end{pmatrix} dt + \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{pmatrix} \begin{pmatrix} dW_t^{1,Q} \\ dW_t^{2,Q} \\ dW_t^{3,Q} \end{pmatrix} \quad \text{(B.15)}$$

then the term structure model derived by using no-arbitrage bond pricing formula B.9 with $r_t = \beta_{1t} + \beta_{2t}$ recovers the factor loadings identical to those of Nelson and Siegel and adds the adjustment term $-\frac{C(\tau)}{\tau}$ to the yield curve to ensure cross sectional absence of arbitrage (with regard to the underlying short rate model) as well as arbitrage free evolution of the yield curve in time.

So far, we have only considerated the diffusion process of the Nelson-Siegel factors under the risk neutral measure Q. This was sufficient to derive the formula for spot rates through all maturities. However, to capture the observed factor dynamics we need to specify also a diffusion process under the real world probability P. Following Christensen et al. (2007) and Tunc et al. (2009) we restrict our analysis to an affine specification of the real world factor dynamics[3] , which allows us to model the observed time series of betas as an autoregresive process[4].

More precisely, we assume the three state variables to be independant and evolve as first-order univariate autoregressive processes.

$$\begin{pmatrix} d\beta_{1t} \\ d\beta_{2t} \\ d\beta_{3t} \end{pmatrix} = \begin{pmatrix} K_1 & 0 & 0 \\ 0 & K_2 & 0 \\ 0 & 0 & K_3 \end{pmatrix} \begin{pmatrix} \theta_1 - \beta_{1t} \\ \theta_2 - \beta_{2t} \\ \theta_3 - \beta_{3t} \end{pmatrix} dt + \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{pmatrix} \begin{pmatrix} dW_t^{1,P} \\ dW_t^{2,P} \\ dW_t^{3,P} \end{pmatrix} \quad \text{(B.18)}$$

By integrating B.18 and computing the expected value and variance of β_{it} for $i = 1, 2, 3$ we get

[3]Factor dynamics under the real world measure is tied with its equivalent under the risk neutral measure linearily via a parameter of the market price of risk

$$dW^Q = dW^R + \gamma_t dt \quad \text{(B.16)}$$

where γ denotes the risk premium parameter. Since we want the diffusion process for betas to be affine functions of factors betas both in risk neutral and real world probability measure the risk premium parameter must be also an affine function of the three beta factors

$$\begin{pmatrix} \gamma_{1t} \\ \gamma_{2t} \\ \gamma_{3t} \end{pmatrix} = \begin{pmatrix} \gamma_1^0 \\ \gamma_2^0 \\ \gamma_3^0 \end{pmatrix} + \begin{pmatrix} \gamma_{11}^1 & \gamma_{12}^1 & \gamma_{13}^1 \\ \gamma_{21}^1 & \gamma_{22}^1 & \gamma_{23}^1 \\ \gamma_{31}^1 & \gamma_{32}^1 & \gamma_{33}^1 \end{pmatrix} \begin{pmatrix} \beta_{1t} \\ \beta_{2t} \\ \beta_{3t} \end{pmatrix}. \quad \text{(B.17)}$$

[4]Note, that it is correct to use AR(1) model for parameters beta only if these are stationary.

$$
\begin{aligned}
E(\beta_{iT}|F_t) &= (1 - e^{-K_i(T-t)})\theta_i + e^{-K_i(T-t)}\beta_{it}, && \text{(B.19)} \\
Var(\beta_{iT}|F_t) &= \int_0^{T-t} \sigma_i^2 e^{-2K_i \cdot s} ds && \text{(B.20)}
\end{aligned}
$$

or after discretizing

$$
\begin{aligned}
\beta_{i,t+dt} &= (1 - e^{-K_i dt})\theta_i + e^{-K_i dt}\beta_{it}, && \text{(B.21)} \\
Var(\beta_{it}) &= \frac{\sigma_i^2}{2K_i}(1 - e^{-2K_i dt}). && \text{(B.22)}
\end{aligned}
$$

Appendix C

Kalman Filter

Kalman filter is a technique used to filter a signal observed over time from a measurement noise and inaccuracies of a measuring device that are always a part of the observed values. The resulting value of the signal processed by Kalman filter is usually closer to the true value of the signal and it is computed as a weighted average of a figure calculated from previous observations and an actual measured value. Particular weights of contribution of the predicted and the measured value to the estimate of the true value depend on the accuracy of the measuring device and intensity of the measurement noise. The more accurate device we have the bigger is the weight by which actual measurents influence the filtered value. In the following, we will use the notation as in Bishop and Welch (2006) which was our primary source of information for writing this section.

Kalman filter is usually used to estimate a true value of a state variable $x \in \mathbb{R}^n$ which is assumed to be driven by the following linear stochastic difference equation

$$x_k = Ax_{k-1} + Bu_{k-1} + w_{k-1}. \tag{C.1}$$

This equation is called a state equation because it defines the value of the state variable x. The matrix A reflects the dependance of the current value of x on its value in the previous time step. Variable $u \in \mathbb{R}^l$ is an optional control variable and the matrix B tells us how the input u influences the change in the value of the state variable x. Noise of this process is captured by a random variable w which is assumed to be normally distributed with zero mean and covariance matrix Q.

Typically, we cannot measure directly the values of the state variable x. Instead, we measure a signal $z \in \mathbb{R}^m$ which is assumed to be a linear transformation of the state vector x,

$$z_k = Hx_k + v_k. \tag{C.2}$$

Our observation of z is subject to measurement errors caused by imperfections of the measurement technique. Random vector v captures this measurement noise and is assumed to be normally distributed with zero mean, covariance matrix R and independent of the process noise u.

We denote $\hat{x}_k^-$ the predicted value of the true value x_k using the information available at time $k-1$. Value of x_{k-1} estimated by Kalman filter so that we use all available information up to this time point (i.e. predicted value $\hat{x}_{k-1}^-$ and measured signal z_{k-1}) will be denoted $\hat{x}_{k-1}$. Since we believe that x is governed by the difference equation C.1, being at time $k-1$ we can predict future value of x_k as

$$\hat{x}_k^- = A\hat{x}_{k-1} + Bu_{k-1}. \tag{C.3}$$

Measurement equation C.2 is then used to adjust the prediction of state variable for error between the measured and the predicted signal, i.e.

$$\hat{x}_k = \hat{x}_k^- + K_k(z_k - H\hat{x}_k^-) \tag{C.4}$$

where K_k is the weighting matrix called Kalman gain. The value $\hat{x}_k^-$ is called an a priori estimate (because we do not use all information available at time k) of state variable x_k and $\hat{x}_k$ is called an a posteriori estimate (because we account for observed signal) of x_k. However, due to error terms in equations C.1, C.2, neither $\hat{x}_k^-$ nor $\hat{x}_k$ are granted to be equal to the true value of x_k. We denote P_k^- the covariance matrix of a priori error term $(x_k - \hat{x}_k^-)$, i.e. $P_k^- = E[(x_k - \hat{x}_k^-)(x_k - \hat{x}_k^-)^T]$ and P_k the covariance matrix of a posteriori error term $(x_k - \hat{x}_k)$, i.e. $P_k = E[(x_k - \hat{x}_k)(x_k - \hat{x}_k)^T]$. Racall that our goal is to estimate the true value of x_k, using all information available. This is equivalent to minimizing the a posteriori error term covariance matrix. Thus, the matrix K_k is set such that produces minimal P_k. Resulting optimal algorithm to find the true value of x_k is as follows.

set x_0, P_0, Q, R

. for $k = 1$ to $k = T$

$$\begin{aligned}
\hat{x}_k^- &= A\hat{x}_{k-1} + Bu_{k-1} \\
P_k^- &= AP_{k-1}A^T + Q \\
K_k &= P_k^- H^T (HP_k^- H^T + R)^{-1} \\
\hat{x}_k &= \hat{x}_k^- + K_k(z_k - H\hat{x}_k^-) \\
P_k &= (I - K_k H)P_k^-
\end{aligned}$$

next k

Appendix D

Algorithm

First we estimate the static Nelson-Siegel model for the whole data set (we measure time in years, so we can use e.g. $\lambda = 0.7173$ as suggested in Diebold and Li (2006)) and assign θ_i the mean value of β_i throughout this time horizont

$$\theta_i = E(\beta_{it}) \qquad \text{for } i = 1,\, 2,\, 3.$$

We can compute also $Var(\beta_i)$ the variance of the state variables estimated above and initial values of K_i from the AR(1) coefficients of the time series of betas.

As described in Appendix C, to start the Kalman filter estimation, we need to provide input values of Q, R and initialize state variables β_{it} and covariance matrix P_t. In our case, matrix Q is defined as in Appendix B

$$Q = \frac{\sigma_i^2}{2K_i}(1 - e^{-2K_i dt})$$

and we assume the measurement errors to have standard deviation 0.1% as in Tunc et al. (2009). Following Christensen et al. (2007) and Diebold et al. (2006) we start the estimation of the state variables β_i at their unconditional mean

$$\beta_{i0} = \theta_i \qquad \text{for } i = 1,\, 2,\, 3, \tag{D.1}$$

and unconditional variance

$$P_0 = Var(\beta_i) = \frac{\sigma_1^2}{2K_i} \qquad \text{for } i = 1,\, 2,\, 3. \tag{D.2}$$

Initial values of σ_i can be calculated from D.2.

Now we can use the estimation procedure of Kalman filter described in Appendix C. Matching the common Kalman filter notation with variables in our model, we get the state vector $x_t = [\beta_{1t},\, \beta_{2t},\, \beta_{3t}]^T$, matrix

$$A = \begin{pmatrix} e^{-K_1 dt} & 0 & 0 \\ 0 & e^{-K_2 dt} & 0 \\ 0 & 0 & e^{-K_3 dt} \end{pmatrix},$$

control vector $u = [\theta_1,\, \theta_2,\, \theta_3]$,

$$B = \begin{pmatrix} 1 - e^{-K_1 dt} & 0 & 0 \\ 0 & 1 - e^{-K_2 dt} & 0 \\ 0 & 0 & 1 - e^{-K_3 dt} \end{pmatrix}$$

and the matrix of linear transformation of state variables to measurements is given by the Nelson-Siegel specification

$$H = \begin{pmatrix} 1 & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} - e^{-\lambda\tau_1} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\lambda\tau_n}}{\lambda\tau_n} & \frac{1-e^{-\lambda\tau_n}}{\lambda\tau_n} - e^{-\lambda\tau_n} \end{pmatrix}.$$

The vector of measurements z_t is equal to the vector of spot rates that were observed at date t (in the no-arbitrage model we must add the adjustment term $\frac{C(T-t)}{T-t}$ to the observed yields). We run the Kalman filter for the data set we want to fit with our model.

After this procedure we maximize the loglikelihood function. This is constructed under the assumption that the prediction errors $e_k = z_k - Hx_k^-$ are normally distributed with covariance matrix $\Omega_k = (HP_k^- H^T + R)$. Using this assumption yields the following loglikelihood function

$$\mathfrak{L} = \sum_{k=1}^{T} \left[-\frac{n}{2}\ln(2\pi) - \frac{1}{2}\det(\Omega_k) - \frac{1}{2} e_k^T \Omega_k^{-1} e_k \right]$$

where n is the number of yields estimated at one date and T the number of dates for which we fit the curve.

List of Tables

List of Tables

List of Figures