COPENHAGEN BUSINESS SCHOOL

Forecasting the term structure of LIBOR yields for CCR measurement

by

Jonas Cumselius & Anton Magnusson

Supervisor: David Skovmand

A thesis submitted in partial fulfillment for the
degree of Master of Science in Advanced Economics and Finance

Presented to the
Department of Economics
Copenhagen Business School

April 2016

Number of pages (characters): 92 (209 300)

COPENHAGEN BUSINESS SCHOOL

Abstract

Department of Economics
Copenhagen Business School

Master of Science in Advanced Economics and Finance

by Jonas Cumselius & Anton Magnusson

In this thesis, we analyze the forecasting performance of three versions of the Dynamic Nelson-Siegel (DNS) model derived by Diebold and Li (2006), applied to LIBOR rates during the time period between January 2003 and March 2015. Our objective is to determine if it would be suitable to use the DNS model to forecast LIBOR rates for the Counterparty Credit Risk (CCR) measurement.

The first version represents the standard DNS AR(1) model with a fixed decay parameter (λ), where lambda governs the speed of decay for the other model factors. Small values of lambda results in a better fit at longer horizons and vice versa. Our second DNS model also has the same AR(1) factor dynamics, but with a time dependent decay parameter, i.e., (λ) varies over time. Lastly, we have a DNS model with VAR(1) factor dynamics. We compare the results of these estimates to those from benchmark models, including the random walk model, simple AR(1) and VAR(1) models, AR(1) on three principal components, and a slope regression model. Before assessing the forecasting ability we also analyse the in-sample fit and find that the DNS models show good in-sample results. The forecasting section involves out-of-sample forecasts, distribution forecasts, and backtesting of the DNS model.

First, by letting lambda vary over time in the DNS model we are able to produce slightly better out-of-sample forecasting results than the traditional DNS model with fixed lambda. However, our overall findings indicate that none of our DNS models are able to keep up with the forecasting performance of the random walk model or the simple AR(1) model. Thus, we can conclude that from our analysis there is no convincing advantage in using the more advanced and complicated Dynamic Nelson-Siegel model over a simple AR(1) or random walk model. Finally, our backtesting results support our findings of the overall poor forecasting ability of the DNS model, and indicate that further studies need to be conducted to develop a forecasting model suitable to include in CCR measurement.

Acknowledgements

The following people have contributed to the final product of this masters thesis. David Skovmand our supervisor at CBS for continuously challenging us throughout this process. Damien Bright at Nordea, Group Counterparty Credit Risk, for helpful guidance in the field of Counterparty Credit Risk, and providing us with a data set with endless opportunities. Our families for all the love. Finally, of course our girlfriends for putting up with us during, and hopefully after our time at CBS. To all of you, thank you!

Contents

List of Figures

List of Tables

Chapter 1

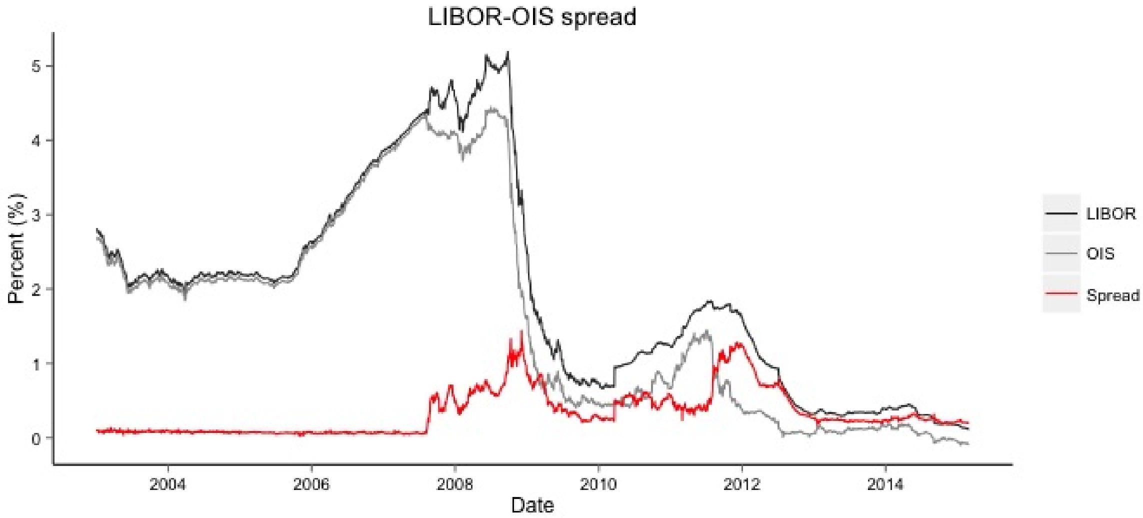
Introduction

Counterparty Credit Risk (CCR) is the credit risk related to counterparties trading over-the-counter (OTC) derivatives. In other words, CCR is the risk that the counterparty to a financial contract will default prior to the expiration of the contract so that credit losses may occur. Thus, with CCR the cause of economic loss is tied to the default (i.e., the health) of the obligor. In addition to traditional forms of credit risk, counterparty risk deals with the uncertainty of exposure as well as the bilateral nature of credit risk (Zhu and Pykhtin (2008)).

Today, the LIBOR-OIS spread is considered by many economists to be a key measure of economic health in the banking sector and an important metric for measuring CCR. It tells us a story of both risk and liquidity in the interbank money market; an indication of the relative health. An increase in the spread, often meaning that LIBOR is high, indicates a decreased willingness to lend by major banks, while a tapered spread is interpreted as an indication of higher liquidity in the market. The LIBOR rate reflects riskiness as it measures the premium demanded by a lending bank for providing an unsecured loan to another bank. On the other hand, the OIS is considered stable as both counterparties only swap fixed for floating interest rate payments. Thus, the LIBOR–OIS spread can be seen as a measure of the creditworthiness of financial institutions, reflecting counterparty risk premiums (Sengupta and Tam (2008)).

Up until the recent financial crisis of 2007-2008, the LIBOR–OIS spread did not get much attention. Historically, the difference between these two important interest rates has hovered around 10 basis points. However, as seen in Figure 1.1 the spread spiked to an all-time high level in mid 2008, and reached similar heights around 2012 at the height of the crisis in the Eurozone. Post-crisis, the spread has gradually stabilised, but it is still higher then the pre-crisis level. Notably, both LIBOR and OIS rates are now considerably below the pre-crisis level.

Figure 1.1: Historical shock decomposition, 2007Q1–2012Q4



Note: The figure above represent the LIBOR–OIS spread shown in percentage over the time period between 2003 and 2015.

Studying these trends provide important insights for how we can account for CCR in valuing derivatives. For many years, the standard practice in the industry was to value derivatives mark-to-market without taking counterparty credit risk into account. The LIBOR rate was traditionally used for discounting all cash flows, since it was considered a proxy for the risk-free rate. After the turbulence around the credit crunch, this practice has been called into question since the true market value of the derivatives must include the possibility of losses due to counterparty default (Zhu and Pykhtin (2008)). According to Hull and White (2013) many banks now suggest that OIS rate is more accurate as the proxy for the risk-free rate when collateralized portfolios are valued, but still consider LIBOR to be suitable for valuation when portfolios are non-collateralized.

There are various approaches to account for CCR when valuing derivatives. Zhu and Pykhtin (2008) focus on two main issues in this area: pricing counterparty risk and modeling credit exposure. They define credit value adjustment (CVA) as the price of counterparty credit risk. The basic idea when computing CVA is to take the difference between the risk-free portfolio value and the true portfolio value that takes into account the possibility of counterparty default, giving the market value of counterparty credit risk. In terms of modeling credit exposure, one approach is to visualize uncertain future exposure through exposure profiles, where these profiles are obtained by simulation of the underlying risk factors in order to attain a realization of future outcomes. Then, at each simulation date, certain statistics are given for the future exposure distribution.

One of the main risk factors to consider when performing these applications is predictions of the term structure of interest rates. The term structure of interest rates, also known as the yield curve, represents the relationship between interest rates and the remaining time to maturity, the so called term. To get an understanding of the credit risk associated with the banking sector, one has to accurately model and forecast the yield curve. In order to investigate the dynamics of the yield curve of interest rates, researchers and practitioners have produced a wide variety of models that can be grouped into big families. Typically, there are two common approaches to term structure modeling. First, the no-arbitrage models that focus on perfectly fitting the term structure to eliminate arbitrage opportunities, important for derivatives pricing. Second, the equilibrium models that model the dynamics of the instantaneous rate, typically using affine models, with the goal set on deriving yields at other maturities. For discussions of these models, see, e.g., Hull and White (1990), Heath et al. (1992a), and Dai and Singleton (2000). Alternatively, there exist more market driven approaches, falling under the LIBOR and Swap Market models (see, e.g., Brigo and Mercurio (2006)) and the Black-type shadow rate models, which model interest rates as options (see, e.g., Black (1995)).

In this thesis, we depart from the aforementioned models and instead focus on the Nelson-Siegel model group; a model group that is widely used by central banks and industry due to the proven benefits in terms of empirical fit. This is vital since we aim to use a model that fits well within a realistic market environment. An important feature of the model is that it has to be able to cope with negative rates. The models of interest are derived from the Nelson-Siegel (1987) model, and the most commonly used is the Svensson (1995) extension, which has gained popularity thanks to its ability to accurately capture the variability of yields. Recent innovations have brought the model into a world where it can be both arbitrage-free and affine (see, e.g., Christensen et al. (2009)). However, our main interest lies in forecasting the term structure and therefore we focus foremost on the Diebold and Li (2006) model approach, a dynamic three-factor version of the Nelson-Siegel (1987). The motivation is that the Diebold and Li (2006) model has been shown to provide superior out-of-sample forecasting, especially for a one-year-ahead horizon. Furthermore, Yu and Zivot (2011) conclude that the Dynamic Nelson-Siegel model with AR(1) factor dynamics - the Diebold and Li (2006) model - performs as well as, if not better than, other more complicated forecasting methods for long-term horizons. We cite this convincing evidence of enhanced empirical fit and smoothness, and the benefits of simplicity, as justification for focusing on this model.

We build on the the Diebold and Li (2006) framework in order to model and forecast the term structure of LIBOR rates, for the period 2003–2015, using the Dynamic Nelson Siegel model (DNS). Thus, we focus on modeling and forecasting a single risk factor, i.e., the LIBOR rate. The common set up in the literature that employs the DNS model is to fix the decay rate factor λ, although it might be possible to further improve the performance of the model by dynamic optimization of this factor in the time series. Small values of λ reduces the speed of decay and produces a better fit for longer maturities while a larger value for λ increases the speed of decay and better fit the curve at shorter maturities. Thus, it is compelling to suggest that different values for λ should be used, depending on the maturity. Previous applications fix the decay parameter mostly for reasons of simplicity. We examine how the DNS model can be extended by dynamically optimizing this parameter. Furthermore, we introduce the Svensson (1995) extension through the Dynamic Nelson Siegel Svensson (DNSS) model, and compare how well the models fit the yields for our time period. Briefly, the Svensson (1995) extension involves adding a forth parameter to cope with even further variation throughout the yield curve. From a counterparty risk perspective, we are interested in forecasting the probability distribution of the interest rates, in contrast to merely predicting the value at a particular point in time. However, we also focus on the out-of-sample forecasting performance, compared to other natural benchmark models. To evaluate the quality of the forecast, we calculate how likely it is that the realized values come from the distribution predicted by the model, so called backtesting.

Most of the studies using the Nelson-Siegel models investigate the term structure of government bond yields and simply evaluate the results in a standard fashion, without further interpreting the performance of the models in an applied sense. Thus, the main contribution of this thesis is an extension of the current literature by exploring the interbank money market - relative to solely looking at government bond yields - and assessing the forecasting ability of our model used to describe the dynamics of the single risk factors. Further, instead of using the LIBOR and Swap Market models, as in similar studies by Brigo and Mercurio (2006), we employ the well-studied Nelson-Siegel family; motivated by the good performance. Furthermore, we contribute by providing a modeling and forecasting framework for building a comprehensive Counterparty Credit Risk model which is suitable for including LIBOR rates, and for evaluating the forecasting performance of the model.

In short we aim to answer the following questions:

1. Are we able to produce the good empirical fit that the Dynamic Nelson-Siegel model has become known for?
2. Are we able to improve these fitted yields by using the Svensson (1995) extension, i.e., the Dynamic Nelson-Siegel-Svensson model?
3. Are we able to improve the forecasted yields by letting lambda vary over time?
4. How does the DNS model perform compared to other natural forecasting competitors?

5. Are we able to produce a forecasting model for the LIBOR rates that could be used in a CCR model?

The rest of the thesis is structured as follows. In Chapter 2, we summarize previous studies in the area of term structure modeling and forecasting. Here we mainly focus on the evolution of the Nelson-Siegel models and the connection to Counterparty Credit Risk. In Chapter 3, we lay out the theoretical foundation of interest rate theory in order to understand the modeling framework, as presented in Chapter 4. In Chapter 5, we proceed to an empirical analysis, describing the data, estimating the models, and examining the empirical fit of the models. In Chapter 6, we continue by examining the out-of-sample forecasting performance compared to the natural competitors, and in Chapter 7 we assess the the adequacy of the model and ability to include it in Counterparty Credit Risk measurement. In Chapter 8, we wrap up this paper by providing concluding remarks.

Chapter 2

Literature Review

In order to investigate the dynamics of the yield curve researchers and practitioners have produced an extensive literature with a wide variety of models. Also, when it comes to the applied use of these models a vast amount of research has been conducted. Our intention here is not to make an extensive survey covering all term structure models and applications but rather to understand the evolution of one of the most widely used models, namely the Nelson-Siegel (1987) model (and its extensions), and in particular how we can connect the literature to risk factor modeling for Counterparty Credit Risk (CCR) measurement. The relation to the CCR measure includes certain performance requirements for the models. To be valuable for CCR measure the model, among other criteria, has to fulfill the following: Perform well when backtesting on the historical data. It has to forecast future distributions of zero rates at all tenors and for both short-term and long-term horizons in a best possible way. Finally, its calibration has to be stable for the entire framework. We will come back to these backtesting specifications after we have examined factor analysis and yield curve models.

Diebold and Rudebusch (2012) provide a deep dive into the literature and elaborate on a particular approach to yield curve modeling and forecasting. They guide the reader from yield curve basics, introducing the early stages of the Nelson-Siegel model, to the most recent innovations in the area. Their starting point is the static Nelson-Siegel model, with a functional form suitable for fitting the cross section of yields. Moving on to the dynamic Nelson-Siegel model, which allows for time-varying parameters. Further they lead the reader to Arbitrage-Free Nelson-Siegel models, which, as the name suggests, enforce the theoretically desirable property of absence of risk-less arbitrage. Finally, they introduce the Dynamic Nelson-Siegel-Svensson model as well as the Arbitrage-Free Generalized Nelson-Siegel model. An interesting question to be answered is of course why one should use a model of the Nelson-Siegel family in the first place. Diebold and

Rudebusch (2012) also highlight some other intriguing questions throughout their book, e.g. "Why use factor models for yields?" and "Is the imposition of "No-Arbitrage" useful?". These questions are central points in the literature of yield curve modeling. Thus, the guiding answers will provide useful introduction to the area.

2.1 Why use factor models for yield curve modeling?

At any point in time a large number of yields at different maturities may be observed. But yield curves also evolve dynamically over time. This means that we have cross-sectional and temporal variability, hence, we have a three-dimensional playground. For a large set of yields the high-dimensionality becomes rather complex. So, we want a model that is able to cope with the complexity of high dimensionality. What's more, in order to make any sense of the data one would like to simplify as much as possible without taking away the ability to capture the variability in the data. Instead of using, for example, unrestricted vector autoregressions - which may be over-parametrized, and wasteful of degrees of freedom - it has been realized that yields typically conform to a certain type of restricted vector autoregression, built on factor structure (Diebold and Rudebusch (2012)). The factor structure is known for its feature of being effective where high-dimensional objects are driven by an underlying lower-dimensional set of objects, namely the factors, providing us with a tool to understand the complex set of large observations like bond yields or interest rates, which typically display low-dimensional factor structure. Falsely assuming the data follows factor structure would of course yield a misspecified model.

Factor structure is seen in a broad range of economic research, from financial markets and financial economic theory to macroeconomic fundamentals and macroeconomic theory. Particularly, the factor structure does a good job when picturing the term structure of yields, as we are to show in this paper (and as has been shown by previous research). Early studies tracing back to, e.g., Macaulay (1938) adopted a single-factor view of only describing the long rate or the level of interest rates. However, having a single factor describing the term structure clearly limits the ability of capturing the underlying dynamics. It is obvious that more than just one common factor is needed for interesting analysis, and that is why modern empirical term structure models involve multiple factors.

It is said that merely three factors, or the first three principal components, are everything that is needed to explain most yield variation. In general, the purpose of principal component analysis (PCA) is to compress the data into a few main components to facilitate data analysis. Jolliffe (2002) explains the main idea as to reduce the dimensionality of a data set comprised of a large number of interrelated variables, while retaining as much as possible of the variation present in the data set. A new set of uncorrelated variables, these being the principal components (PCs), are constructed by linear combinations of the original variables. The new factors are then ordered so that the first PC accounts for the largest variability in the data, and each succeeding component with the highest variance possible conditional on being orthogonal to the preceding components. Since the components are equivalent to the eigenvectors of the symmetric covariance matrix, the components are also orthogonal to each other.

In terms of the individual factors, Litterman and Scheinkman (1991) and many others in the literature suggest an interpretation of the first three PCs as level, slope and curvature, respectively. The first PC corresponding to the level is to be seen as the general rise or fall of the yield curve, which is relatively flat. The second PC, which is known as the slope, usually captures situations where the short end moves up while the long end moves down, or vice versa, so that we see opposite signs at both ends of the maturity spectrum. The third PC is often interpreted as the curvature of the yield curve, which is due to both the short and long end of the yield curve moving in the same direction but some region in the middle moving the opposite way. Thus, we have the same sign at both ends of the maturity spectrum but the opposite sign somewhere in the middle. Lord and Pelsser (2007) mathematically investigates whether the common interpretation of the first three PCs as level, slope and curvature are fact or artefact. They define the observed pattern by stating that if the first three factors or eigenvectors have, in order, zero, one and two sign changes, the corresponding matrix displays level, slope and curvature. Using generalisations of theorems from the mathematical study of total positivity, they find sufficient conditions under which level, slope and curvature are present.

Empirical observations by Joslin et al. (2014) tell us that in most developed countries the cross-correlations of bond yields can be well described by a low-dimensional factor model. They note that often three or four factors explain nearly all of the cross-sectional variation in yields, seen for a wide range of maturities. Diebold and Rudebusch (2012) show that close to a 100% of the variation in US government bond yields (January 1985 to December 2008) can be explained by the first three principal components. For their data set, the first PC varies the most, but it is at the same time the most predictable of the three, due to high persistence. Diebold and Rudebusch (2012) relate the reduction of the first factor to the reduction of the inflation over the period, relative to the high level in the beginning of the 80s. The second factor follows business cycle movements and is less variable, fairly persistent and predictable, but less so than the first factor. The third PC shows the least variability, persistence, and predictability among the three. Furthermore, by plotting these three PC factors against standard empirical yield measures, they show that the PC factors respectively coincide with the standardized measures of level, slope, and curvature (10Y yield, 10Y−6M spread, a 6M+10Y−2×5Y "butterfly spread", respectively), as suggested by the literature. Interesting to note that the factor's relation to business cycle movements and inflation rate imply that they are likely to have specific macroeconomic determinants, an area which has not yet been widely explored.

Diebold and Rudebusch (2012) state three key reasons for why dynamic factor models are commonly used: First of all, the factor structure are able to give an accurate empirical description of yield curve data. Nearly all information of the underlying dynamics can be summarized with only three or a few factors. Thus, yield curve models are usually structured by a small number of factors along with their associated factor loadings, relating yields of different maturities to those factors.

Second, from a statistical point of view factor models are very appealing since they provide valuable compression of information. By effectively reducing the dimensionality it provides a low-dimensional modeling environment which is more manageable to work with, allowing us to focus only on how the factors evolve through time. Consistent with the "parsimony principle", using restricted simple models (with only a few parameters or factors) often prevents a lot of data mining and helps to produce good out-of-sample forecasts, even in the case of false restrictions that may degrade in-sample fit. A more complex model (e.g., an unrestricted vector autoregression) might enhance the in-sample fit, but at the cost of reducing the value the out-of-sample forecasting due to the large number of estimated coefficients.

The final reason relates to financial economic theory and the use of factor structure. Although thousands of financial assets are seen in the markets, the interesting part is the risk premiums separating their expected returns, which are driven by a small number of components, i.e., risk factors. As an example, Diebold and Rudebusch (2012) relate to the capital asset pricing model (CAPM), a single-factor model frequently used in the equity sphere. Even extensions of the model, e.g. Fama and French (1992) that adds a few factors, seldom exceed five factors due to dimensional reasons. Yield curve factor models are thus said to be the bond market counterpart.

2.2 Term Structure Models

The original Nelson-Siegel model was first introduced by Nelson and Siegel in 1987 and belongs to the group of exponential affine three-factor term structure models. These type of models focus more on empirical fit than theoretical rigorousness. Typically, there are two common approaches to term structure modeling. The no-arbitrage models that focus on fitting the term structure in order to rule out any existence of arbitrage possibilities, an important characteristic for derivatives pricing. And the equilibrium models trying to model the dynamics of the instantaneous rate, typically using affine models. Under various assumption of the risk premia, the instantaneous rate is then used to derive yields at other maturities (Diebold and Li, 2006). For papers contributing to the affine equilibrium models see, e.g., Vasicek (1977), Cox et al. (1985), and Duffie and Kan (1996). And for papers contributing to the no-arbitrage models see, e.g., Hull and White (1990), and Heath et al. (1992a).

Some appealing features of the Nelson-Siegel model have made it very popular for curve fitting in practice, especially among financial market practitioners and central banks (Diebold and Rudebusch, 2012). The first desirable feature is that the model enforces some basic constraints from financial theory, e.g. the zero-coupon Nelson-Siegel curve satisfies

$$\lim_{\tau \to 0} y(\tau) = f(0) = r,$$

the instantaneous short rate, and $\lim_{\tau \to \infty} y(\tau) = \beta_1$, a constant. Second, the Nelson-Siegel form provides a parsimonious, yet flexible approximation. Parsimony in this sense means the use of few parameters, which promotes smoothness (yields tend to be smooth functions of maturity), protecting the model against in-sample overfitting, a valuable feature for good forecasting. Also, it promotes estimations that are empirically tractable and trustworthy. Since the yield curve take on a variety of shapes at different times a flexible approximation is desirable. The shape of the curve depends on the values of four parameters only. Third, from a mathematical perspective the model form takes on the approximation-friendly Laguerre structure: the yield curve is a constant plus a Laguerre function. Laguerre functions or Laguerre polynomials are polynomials multiplied by exponential decay terms. They are solutions to Laguerre's differential equation:

$$xy^{''} + (1-x)y^{'} + \lambda y = 0,$$

which is a second-order linear differential equation. These Laguerre polynomials are conventional approximating functions on the domain zero to infinity, which is suitable for approximations of the term structure.[1]

Theoretically, the Nelson-Siegel model does not rule out arbitrage opportunities, but in practice it has shown robust performance. Coroneo et al. (2011) use a non-parametric re-sampling technique on zero-coupon yield data from the US market to show that the Nelson-Siegel model is arbitrage free in a statistical sense. They find that, at a

[1] For deeper knowledge of Laguerre functions, see Abramowitz and Stegun (1964).

95 percent confidence level, there is no statistical difference between the Nelson-Siegel parameters and the no-arbitrage parameters. Still, the model does not have an arbitrage-free foundation. Some might argue that it is not good enough since we work in deep and well-organized markets, which are approximately arbitrage-free, and thus need models that impose this restriction. Diebold and Rudebusch (2012) argue that the imposition of no-arbitrage would have little effect if the model provides an almost exact description of an arbitrage-free reality. Then, the model would be approximately arbitrage-free, without explicitly imposing the condition of being so.

Further, Diebold and Li (2006) established a dynamic formulation of the classic Nelson-Siegel model, the Dynamic Nelson-Siegel (DNS). It has proven to be favourable, especially when it comes to out-of-sample forecasting. The dynamic structure allows for time-varying parameters. The authors show that their model produce encouraging results when forecasting term structure, especially for long horizons. The 12-month-ahead forecast shows superior results, while the 1-month-ahead forecast perform almost the same result as the random walk and other leading forecasting competitors. Also, the dynamic factor structure of the model made it possible to interpret the three factor loadings as level, slope and curvature; long-term, short-term, and medium-term respectively. The characteristics of each factor is further explained in Chapter 4. The Diebold and Li (2006) specification of the model is not arbitrage-free. Instead they focus on good forecasting abilities and easier interpretation. Although it is not obvious how extensions to an arbitrage-free framework would affect forecasting performance, Diebold (2008) suggest that parsimonious models are often better for out-of-sample forecasting. Duffee (2002) argue that the no-arbitrage characteristic alone may not provide for good forecasting abilities, and that a well specified model is also an important factor. For further elaboration of the trade-off between forecasting performance and freedom of arbitrage see, e.g., Dai and Singleton (2002).

Extensions of the three factor Nelson-Siegel model have been made in order to improve the empirical fit of yield curves. The Svensson (1995) extension is widely used by both industry and central banks. By adding a second curvature factor it allows for a better fit even at longer maturities, which is useful for a more accurate fit of the whole maturity spectrum of yields. This is also discussed in, e.g., Svensson (1995), BIS (2005), Gürkaynak et al. (2007), and Nyholm (2008). However, not even this popular extension enforce a consistent arbitrage-free environment over time. Therefore, recent innovations in this space have attacked this problem. Christensen et al. (2007) first introduce the affine arbitrage-free class of the Nelson-Siegel term structure models, without incorporating the Svensson extension. In order to obtain an arbitrage-free approximation of the Svensson extended Neslon-Siegel model, Christensen et al. (2009) add an additional slope factor to pair with the the second curvature factor. Thus, they provide a five-factor arbitrage-free generalized Nelson-Siegel (AFGNS) model. The authors show that the AFGNS model displays theoretical consistency and good in-sample fit. However, it does not provide us with any forecasting improvements. This comes back to the trade-off between the theoretical rigorous no-arbitrage consistency and forecasting performance.

2.3 Yield Curve Forecasting

As discussed by Diebold and Li (2006), little attention has been paid to the key practical problems of yield curve forecasting, although there has recently been powerful theoretical advances in yield curve modeling. The no-arbitrage model literature is mainly concerned with fitting the term structure at a point in time and thus gives us limited information about the dynamics of the yield curve or forecasting. The affine equilibrium model literature on the other hand is mainly concerned with the dynamics driven by the short rate, which could be linked to forecasting. However, most focus only on historical in-sample fit as opposed to out of-sample forecasting. Moreover, as noted by Duffee (2002), those who actually do focus on out-of-sample forecasting achieve poor results.

Diebold and Li (2006) provide relatively short-term out-of-sample forecasts (longest horizon being 12-month-ahead forecasts) compared to a more recent paper by Yu and Zivot (2011), that focus on long-term (as far as 60-month-ahead) forecasts of Treasury bonds and corporate yields. Yu and Zivot (2011) compare different forecasting approaches and their findings suggest that the one-step approach (i.e. state space approach) state is not necessarily better than the simple two-step dynamic Nelson-Siegel with the AR(1) model, which is the main model in Diebold and Rudebusch (2012).

As reported by Diebold and Li (2006) their DNS out-of-sample forecasting results improve considerably as the forecast horizon lengthens. Going from 1-month-ahead to 6-month-ahead improves the results moderately, but stretching out to 12-month-ahead retrieves results that outperform those of all the compared methods for all maturities included. When compared to the random walk or "no change" forecast explicitly, it does not come as a big surprise that random walk is not outperformed by DNS for the shortest forecast horizon Diebold and Rudebusch (2012). For the 1-month-ahead horizon the yield factor mean reversion captured by DNS may not have sufficient time to operate, whilst the random walk fails to capture the mean reversion in yield factors for the longer horizons. According to Diebold and Rudebusch (2012) the relative performance of DNS is often optimized at 6- to 12-month horizons. However, it might be interesting to examine how the model performs when further lengthen the forecast horizon to 18- to 26-month horizons.

In addition to versions of the Nelson-Siegel model, and often as a comparison, principal component analysis (PCA) is frequently used to depict yield curve behavior and decomposition. Litterman and Scheinkman (1991) apply PCA in order to explain US treasury bond returns in terms of the three factors; level, slope and curvature, similar to the factorization by Diebold and Li (2006). Knez, Litterman, and Scheinkman (1994) extend the three-factor model and provide a four-factor model by observing money market returns. They argue that the additional factor is related to private issuer credit spread. Further, Duffie and Singleton (1997) argue that the no-arbitrage and equilibrium term structure models are not applicable to the swap market since the swap contracts include default risk. Instead they propose a multi-factor model for interest rate swaps incorporation both credit ant liquidity risk. Furthermore, Blaskowitz and Herwartz (2009) provide a term structure decomposition of the EURIBOR swap by PCA and AR models for adaptive forecasting. They find that these models produce superior results in terms of directional accuracy and forecast value when compared to the benchmark models.

Diebold and Li (2006) report that although their approach may have a close relation to direct principal components regression, yet the approach and especially the results are not identical. Moreover, they state that there is reason to prefer the Diebold and Li (2006) approach both from an empirical and a theoretical perspective. Empirically, the results indicate that the forecasting performance on the specific sample of yields is superior. And theoretically, they argue that methods including regression on principal components regression often have the following unappealing features:

- cannot be used to produce yields at maturities other than those in the sample,
- do not guarantee a smooth yield curve and forward curve,
- do not guarantee positive forward rates at all horizons,
- do not guarantee that the discount function starts at 1 and approaches 0 as maturity approaches infinity.

Predictive ability of interest rate term structure models is a fundamental concern in economics. There are three main types of interest rate forecasting: point, interval and density forecasting. Diebold and Lopez (1996) reveal that during the last decades most attention has been paid to evaluating point forecasts, crucial for bond portfolio management. While little attention has been given to the evaluation of density forecasts, important for both derivatives pricing and financial risk management. Density forecasting provides an estimate of the probability distribution of future values of the variable of interest. Thus, it provides a detailed description of the uncertainty associated with the predicted values. As opposed to point forecasting, which by itself gives only a specific value at a future point in time (Tay and Wallis (2000)). According to Diebold et al. (1998) density forecasting is increasingly more important for evaluating future scenarios. The authors develop a simple and operational framework for density forecast evaluation. They illustrate the framework with a detailed application to density forecasting of asset returns in environments with time-varying volatility. Historically, analysis of density forecasts has required restrictive assumptions and computationally intensive techniques. However, improvements in computer technology and an increasing demand for density forecasting have brought attention to the area and made it possible to conduct improved density forecasts.

Tay and Wallis (2000) explains density forecasting as being implicit in the standard construction and interpretation of a symmetric prediction interval around the point forecast. Usually the interval is constructed with one or two standard errors, and the corresponding probability of 68% or 95% respectively rests on the distributional assumption: normal or Gaussian. In some cases the Student's t-distribution is also used. Bear in mind that in those cases the forecast errors have to be estimated, usually computed with models resting on normality assumptions. Therefore it is suggested to test for normality. Such tests typically rely on third and fourth moments, rejecting the null hypothesis of normality if there is significant skewness and/or excess kurtosis. As noted by Tay and Wallis (2000) many empirical studies have found non-normal higher moments in the (unconditional) distributions of stock returns, interest rates, and other financial data series.

2.4 Relation to Counterparty Credit Risk measure

Applications of density forecasting is seen both in macroeconomics and microeconomics, especially within the field of finance. According to Tay and Wallis (2000) density forecasting in finance derives from the literature that aims to model and forecast volatility, e.g. ARCH and GARCH models. Since we are dealing with uncertainty around the predicted values it is closely related to volatility measures. There are several reasons for an interest in more a complete and accurate probability statement, particularly within the financial sector, and more so in the area of risk management. In the wake of the recent financial crisis risk management has developed into an industry, where density forecasts are regularly being issued. This allows for generating density forecasts of the change in the value of a particular portfolio over a specified holding period. The special interest here is usually the n^{th} percentile of the distribution, related to the commonly known risk measure, Value-at-Risk (VaR), predicting that the portfolio is going to lose a value greater or equal to its VaR over the holding period with the probability $n/100$.

Departures from normality in the portfolio returns will distort the usefulness of VaR estimates if the assumption of normality is used inappropriately when generating the forecast (Tay and Wallis (2000)). Important to note here is that VaR is a measure of market risk as opposed to Counterparty Credit Risk. VAR typically uses 1 day and 10 day forecast horizons whereas CCR risk uses much longer horizons, as long as up until portfolio maturity.

According to Gregory (2010) the basic strategy for financial institutions in managing counterparty risk should be based on the following key elements: credit exposure, default probability, expected loss given default (or equivalently recovery rate). These components may be assessed in different ways by separate divisions within an organisation but at some point they all need to be collected and combined, usually done by a specific counterparty risk group. When considering the individual weights of the components; a counterparty with large default probability and small exposure may be considered preferable to a situation with larger exposure and smaller underlying default probability. Furthermore, a high level of collateralization and thus a reduced amount of loss given default may be considered preferable to a less risky counterparty with more limited arrangements.

Of these elements we are foremost concerned with credit exposure in this paper, and more specifically with single risk factors, e.g. interest rates, LIBOR rates in our case. When examining CCR exposure one looks closely at expected positive exposure (EPE), which is the expected average credit exposure on a future date conditional on positive market values, as well as potential future exposure (PFE), usually at the 95th confidence interval. The EE forecast affects the capital requirements (related to IMM model approval and reported capital numbers such as default risk charge and CVA risk charge ie as of CRD IV regulations) and the PFE is generally used for trading limits management (i.e. affects ability of traders to book trades against a counterparty).

Zhu and Pykhtin (2008) offer a framework for modeling credit exposure and pricing counterparty risk. In their article they provide a guide to modeling counterparty credit risk by highlighting and answering the questions – What are the steps involved in calculating credit exposure? What are the differences between counterparty and contract-level exposure? How can margin agreements be used to reduce counterparty credit risk? What is credit value adjustment and how can it be measured? – For modeling credit exposure they lay our a three-step procedure consisting of Scenario Generation, Instrument Valuation, and Portfolio Aggregation. We limit ourselves to a sort of Scenario Generation, where future market scenarios are simulated for a fixed set of simulation dates using evolution models of the risk factors. By using term structure models one is able to forecast future distributions of the yield curve.

Nowadays, most larger banks and financial institutions have permission to use internal model methods (IMM) to calculate regulatory capital for their CCR exposure. With this permission comes a requirement to carry out ongoing validation of CCR exposure models in order to demonstrate that the models are appropriate. The Basel Committee on Banking Supervision specify in their guidance paper: "Sound practices for backtesting counterparty credit risk models" that the ongoing validation is expected to be able to identify issues with the models, and also meant to reaffirm that the model assumptions are not violated, as well as that known limitations are kept appropriate. Backtesting Counterparty Credit Risk models is becoming increasingly important in the financial industry. However, there are no clear guidelines by regulators as to how to perform this backtesting, as opposed to for Market Risk models, for which the Basel Committee has set a strict set of rules from 1996, that are widely followed. The importance of backtesting arises from the recent financial crisis, and since then, both the CCR capital charge and CVA management have become more central to banks.

Backtesting is a vital part of the model validation process and the recent financial crisis has revealed that additional guidance in this area is required. Furthermore, The Basel Committee state that implementation of these sound practices most likely will improve the backtesting of internal models and, as a result, will enhance the elasticity of both individual banks and the financial system (BIS (2010)). The Basel Committee define backtesting as:

"Backtesting is part of the quantitative validation of a model that is based on the comparison of forecasts against realized values. Validation is a broader term that encompasses backtesting, but can be any process by which model performance is assessed."

As brought up by Ruiz (2012), there are two major areas where backtesting applies. The first is in the calculation of the Value at Risk (VaR), which later feeds into the Market Risk capital charge. The second is in the calculation of EPE profiles, that feed into the Counterparty Credit Risk (CCR) and CVA-VaR charge. As mentioned earlier, The Basel Committee has stated clear rules as how to perform the VaR backtest, as well as being clear about the consequences of a negative backtest for financial institutions. However, since there are only guidelines on how to perform backtesting for CCR models, financial institutions face a mixture of requirements from different national regulators. As a consequence this causes some confusion across financial institutions, and as Ruiz (2012) states, the global financial system is thus exposed to regulatory "arbitrage" in this area.

In an attempt to reduce this "arbitrage" situation Ruiz (2012) propose a methodology in the context of counterparty risk that can be related to the strict backtesting framework which is in place for market risk, with the criteria of being: scientifically sound, practical and easily used by management. According to Ruiz (2012) the general practice of CCR model backtest refers to the backtest of the models generating EPE profiles. We can decompose these models into a number of sub-models: Risk Factor Evolution (RFE) models for the stress-testing of the underlying factors (e.g., yield curves, FX rates), pricing models for each derivative, collateral models for secured portfolios, and netting and aggregation models.

Of these sub-models we focus on RFE models, where we want to stress-test the underlying factors, i.e. yield curves. Furthermore, the RFE models tend to be the most important driver of the EPE profile, since a 5% change in volatility of a risk factor tends to have a significant impact of the EPE profile, compared to a limited impact for a 5% inaccuracy in pricing.

Ruiz (2012) describes the backtesting of a RFE model as comparing the distribution of the risk factor given by the model over time with the distribution actually seen in the market. In other words, we want to check how the RFE measure and the observed "real" measure compare to each other. We are going to expand on this more technically later on. And in order to assess the performance of the backtest certain criteria must be set up.

Determining the backtesting performance criteria of IMM models is of major concern for the banks, since these criteria are a key factor of the backtesting process and should be re-considered over time. The criteria are necessary in order to determine whether or not the observed performance is appropriate. However, the Committee itself has some guidelines on how the criteria are set up and what they must fulfill. For example, the Committee requires that the forecasting system is reliable BIS (2010). The Basel Committee define a reliable forecasting system as:

"A reliable forecasting system is one for which events forecast occur with an observed relative frequency that is consistent with the forecasted values. [2]

In other words, backtesting of a risk factor model is reliable if the p^{th} percentile and the q^{th} percentile of the forecast distribution capture the actual values with the forecast probability of $(q - p)$. Furthermore, the model must not be calibrated to the observed performance of the model, instead it should be constructed objectively.

As observed in Kenyon and Stamm (2012) the choice of calibration of the risk factor model (market implied or historical) influence the assessment of the forecasting ability of the model used to describe the dynamics of the single risk factors. On the one hand, market implied calibration should be used for CVA computation, where prices

[2] BIS (2010) p.8

are central. On the other hand, forecasts of interest rates which hinge on real world environment are best suited with historical calibration.

Chapter 3

Interest Rate Theory

Before we dive deeper into the specific models that we will use in the thesis, we want to introduce the reader to some interest rate theory that we believe is necessary background knowledge in order to fully understand the methods and terminologies in the proceeding chapters. This theory overview will if not otherwise stated be based on Kani (2007).

3.1 The Money Market Account

A money market account is a risk-less account which profits accruing continuously at the risk free rate prevailing in the market at any instant.

Following Kani (2007) we define $B(t)$ to be the value the continuously compounded bank account at time $t \geq 0$. Further assume that $B(0) = 1$ and that the account follows:

$$dB(t) = r_t B(t)dt \quad , B(0) = 1 \tag{3.1}$$

Where r_t is a positive (can be stochastic) real valued process. Which means that we can write:

$$B(t) = exp\left(\int_0^t r_s \mathrm{ds}\right) \tag{3.2}$$

By these definitions we know that investing in a unit of currency at time 0 yields at time t the value given in 3.2 and that r_t is the instantaneous rate at which the account accrues. For a small interval $[t, t + \Delta t]$ we have:

$$\frac{B(t + \Delta t) - B(t)}{B(t)} = r(t)\Delta t \tag{3.3}$$

Which tells us that the bank account grows at the rate $r(t)$. By this we further know that the value of a unit of currency payable at time τ is given by $\frac{B(t)}{B(\tau)}$ and if r is a stochastic process then this is called the stochastic discount factor.

3.2 The stochastic discount factor

The stochastic discount factor $D(t,\tau)$ is used to relate amounts of money across time. So the stochastic discount factor is the price of one unit of currency at time t that is equivalent to one unit of currency at time τ. The stochastic discount factor is given by:

$$D(t,\tau) = \frac{B(t)}{B(\tau)} = exp\left(\int_t^\tau r_s \mathrm{ds}\right) \tag{3.4}$$

The stochastic discount factor leads us into the simplest form of bonds, namely the zero-coupon bond.

3.3 The zero coupon bond

A zero coupon bond is the simplest form of bond since it has no "coupon" or periodic interest payment i.e. the investor in a zero coupon bond only receives one payment, at maturity. This payment is equal to the principal plus the interest earned at a stated yield. Zero coupon bonds are in practice however not directly observed in the markets. Long maturities zero coupon bonds are not traded at all, but can be obtained by bootstrapping coupon bonds. A zero coupon bond with maturity τ is defined as in Kani (2007):

Definition 3.1. *A τ maturity zero coupon bond is a contract that guarantees its holder the payment of one unit of currency at time τ, with no intermediate payments. The contract value i.e. the price of a zero coupon bond at time $t < \tau$ is denoted by $P(t,\tau)$ and $P(\tau,\tau) = 1\ \forall\,\tau$ and is equal to the present value of the nominal amount:*

If we know that the zero-coupon bond is a contract in time t that gives us the present value of one unit of currency to be paid at time τ and if r is deterministic then D is deterministic since it only depends on r and hence $D(t,\tau) = P(t,\tau)\ \forall[t,\tau]$ and we have the price of the zero coupon bond. However if r is not deterministic but stochastic then $D(t,\tau)$ will depend on the path of r between t and τ we still however need to know $P(t,\tau)$, therefore we say that $P(t,\tau)$ is the market expectation of $D(t,\tau)$ under the risk neutral measure.

3.4 Time to maturity

The time to maturity $\tau - t$ is the amount of time (in years) from the present time t to the maturity $\tau > t$. In order to have a consistent (or fair) measure of this distance we need to define the distance between them in terms of days (or years). This definition is not unique and is called the day count convention.

3.5 The Day Count Convention and The Compounding Types

3.5.1 The day count convention

The day count convention is an agreed upon system to determine the number of days between two dates e.g. t and τ and we define it as $d(t, \tau)$. This time difference is usually defined in the fraction of a year. Different bond markets have different day count conversions for example 30 days in a month and 360 in a year. For the LIBOR market (which we are looking at in this thesis) the day count convention is Actual/360 except for the GBP where it is Actual/365 Fixed. This means that each month is treated normally (i.e. 28,29,30 or 31 days) but the year is fixed at 360 or 365 regardless if it is a leap year or not. The day count convention is important to define since it defines how much interest that has been accrued between two payments.

3.5.2 Compounding Types

The compounding can be classified into four different groups:

1. Continuously compounding
2. Simply compounding
3. k-times per year compounding
4. Annually compounding

All these can be expressed as both spot rates and as forward rates. LIBOR rates are of type two i.e the simply compounding type. We choose to, due to space limitations only discuss the continuously and simply compounded rates here. For the other types see Kani (2007).

1. Continuously compounded rates:

The continuously compounded rate is the rate that yields one unit of amount of currency at the time of maturity (τ) for an investment of $P(t,\tau)$ at time t.

The Continuously Compounded Spot Rate:
If the contracting date is equal to the start of the interval i.e. $[T = t, \tau]$ then we have the continuously compounded spot rate which we define as $R(t,\tau)$. We have that the continuously compounded rate for the period $[t,\tau]$ is defined as:

$$R(t;t,\tau) = -\frac{\ln P(t,\tau)}{d(t,\tau)} \tag{3.5}$$

From this we can derive the price of a zero coupon bond ($P(t,\tau)$) as:

$$R(t,\tau)d(t,\tau) = -\ln P(t,\tau) \;\Rightarrow$$
$$\mathrm{P(t,}\tau\mathrm{)=exp}\Big(\text{-R(t,}\tau\text{)}d(t,\tau)\Big) \tag{3.6}$$

2. The Simply-Compounded Rate: When the accruing occurs proportionally to the time of investment then we have a simply compounded rate $L(t,\tau)$. It is the rate that yields one unit of amount of currency at the time of maturity (τ) for an investment of $P(t,\tau)$ at time t, when accruing is proportional to the investment time.

The Simply-Compounded Spot Rate:

$$L(t,\tau) = \frac{1 - P(t,\tau)}{d(t,\tau)P(t,\tau)} \tag{3.7}$$

The market LIBOR rates that we will be using in this thesis are simply-compounded rates and are linked to the zero-coupon bond prices by the day count convention for computing $d(t,\tau)$.

Zero coupon bond prices in terms of $L(t,\tau)$:

$$P(t,\tau) = \frac{1}{1 + d(t,\tau)L(t,\tau)} \tag{3.8}$$

3.6 Yield curve

The yield curve, is the graph of the function that maps maturities into rates for different t. The yield curve can be thought of as the curve that describes the relationship of the

returns on bonds with the same credit risk but with different maturities.
Define the Yield Curve as:

Definition 3.2. *The yield curve at t is the graph of the function:* $\tau \to R(t,\tau)$, Where $\tau > t$

3.7 Forward rates

Forward rates have three time periods to consider, the time today (t), the expiry time T and the maturity τ, where we have that $t < T < \tau$. Thus a forward rate are interest rates that are agreed upon today for an investment in a future time period and are set consistently with the current yield of discount factors. In practice the holder of a forward contract has the obligation to buy or sell a product at a future date at a given price. We define the forward rate by following Senghore (2013):

Definition 3.3. *Given three fixed time points $t < T < \tau$, a contract at time t which allows an investment of a unit amount of currency at time T and gives a risk less deterministic rate of interest over the future interval $[T,\tau]$ is called the forward rate*

Also note that if $t = T$ then the forward rate is just the spot rate. We can also define the forward rates in terms of the Forward Rate Agreement (FRA), following Kani (2007):

In order to lead us into the simply compounded forward interest rates, we say that the FRA is a contract that gives the holder an interest rates payment for the period $[T,\tau]$ based on the spot rate $L(T,\tau)$ resetting in T with maturity τ. The seller of the contract gets a fixed payment based on the rate K. To be specific:

If we have a nominal value of the contract at N and assume the same day count convention then at time T one receives: $Nd(T,\tau)K$ units of currency and pays: $Nd(T,\tau)L(T,\tau)$.

So the value of the contract at T is:

$$
\begin{aligned}
Nd(T,\tau)K - Nd(T,\tau)L(T,\tau) \;\Rightarrow \\
Nd(T,\tau)(K - L(T,\tau))
\end{aligned}
\tag{3.9}
$$

Using 3.7 we can rewrite it to be:

$$
N\left[d(T,\tau)K - \frac{1}{P(T,\tau)} + 1\right] \tag{3.10}
$$

In order to find the fair value of this contract at time t we first need to find the present value of the expression above. Note that $P(t,\tau) = P(t,T)P(T,\tau)$ and that $d(T,\tau)K+1$ at time τ is worth $P(t,\tau)[d(T,\tau)K+1]$ at time t. First we rewrite:

$$P(t,\tau) = P(t,T)P(T,\tau) \ \Rightarrow P(T,\tau) = \frac{P(t,\tau)}{P(t,T)} \tag{3.11}$$

Plug this into 3.7:

$$\begin{aligned} &N\left[d(T,\tau)K - \frac{1}{\frac{P(t,\tau)}{P(t,T)}} + 1\right] \Rightarrow \\ &N\left[d(T,\tau)K - \frac{P(t,T)}{P(t,\tau)} + 1\right] \end{aligned} \tag{3.12}$$

Take the present value of 3.12 and we have the value for the FRA at time t:

$$\begin{aligned} \text{FRA}(t,T,\tau,d(T,\tau),N,K) &= P(t,\tau)N\left[d(T,\tau)K + 1 - \frac{P(t,T)}{P(t,\tau)}\right] \Rightarrow \\ \text{FRA}(t,T,\tau,d(T,\tau),N,K) &= N\left[P(t,\tau)d(T,\tau)K - P(t,T) + P(t,\tau)\right) \end{aligned} \tag{3.13}$$

We want to find the fair value of this contract i.e. solve for the rate K that sets the value of this contract to 0 at time t. It turns out that there is only one solution for this and that the resulting rate K is our definition for the simply compounded forward rate prevailing at time t for expiry $T > t$ and maturity $\tau > T$:

$$F(t;T,\tau) = \frac{1}{d(T,\tau)}\left[\frac{P(t,T)}{P(t,\tau)} - 1\right] \tag{3.14}$$

So in summary the simply compounded forward rate is the rate that gives a fair value to the FRA contract at time t.

3.7.1 Instantaneous Forward Interest Rate

The instantaneous forward interest rate $(f(t,\tau))$ at time t for maturity $\tau > t$ is defined as the forward rate when $T \to \tau$:

$$\begin{aligned} f(t,\tau) &= \lim_{T\to\tau} \frac{1}{\tau - T}\left[\frac{P(t,T)}{P(t,\tau)} - 1\right] \Rightarrow \\ f(t,\tau) &= \lim_{T\to\tau} \frac{1}{P(t,\tau)}\left[\frac{P(t,T) - P(t,\tau)}{\tau - T}\right] = -\frac{1}{P(t,\tau)}\frac{\partial P(t,T)}{\partial \tau} \Rightarrow \\ f(t,\tau) &= -\frac{\partial \ln P(t,T)}{\partial \tau} \end{aligned} \tag{3.15}$$

We can solve for $P(t, \tau)$ in 3.15 which gives us:

$$P(t,\tau) = exp\left(-\int_{t}^{\tau} f(t,u)\mathrm{d}u\right) \tag{3.16}$$

We need one assumption in order for this to work, namely; smoothness of the zero coupon price function $T \rightarrow P(t, \tau)$. We will come back to the instantaneous forward rate in the section about arbitrage theory.

3.8 Arbitrage Theory

An arbitrage opportunity is a trading strategy or portfolio satisfying one of either: (A) its price now is zero and it has a strictly positive payoff in the future, or (B) its price is strictly negative now and it may payout in a future state. Less formally, the absence of arbitrage requires that: (A) a portfolio cannot cost nothing and payoff later, or (B) it cannot give a payoff today without any obligations in the future. This means that the law of one price must hold i.e. the same commodity must hold the same price if no arbitrage opportunity should exist in the market. We will in this section provide an understanding of the interpretations of having an arbitrage free financial market which will lead us into the martingale approach of pricing financial derivatives. This summary is if not otherwise stated based on Lando and Poulsen (2006), Björk (2009) and Senghore (2013). First we need some assumptions (and/or definitions):

- We have the following probability space: $(\Omega, \mathcal{F}, \mathbb{P})$
  - Ω is the sample space i.e the set of possible outcomes (events), $\omega \in \Omega$.
  - $\mathbb{P}$ is the observed probabilities.
  - $\mathcal{F}$ is the full information space. It represents the set of possible events (ω) where an event is a subset of Ω.

- We also have the "filtration" at time t: $\mathcal{F}_t$ which is the information available from the past up until time t.

This means that if we know at time t that an event ω has occurred or not then $\omega \in \mathcal{F}_t$.

- We need to define the: $\mathcal{F}_t$ measurable: $\{\omega : X(\omega) \leq x\}$ belongs to $\mathcal{F}$ for all real x.

This condition means that if the outcomes of the random variable X are always in $\mathcal{F}$ then we can always find a probability for the event $X \leq x$. The simplest example being if we assume X to be a constant, then we will have one outcome of X and we can assign a probability that it is less or equal to x. For a stochastic random variable we have multiple outcomes and all of these need to be in $\mathcal{F}$ for it to be $\mathcal{F}_t$ measurable.[1]

- We say that a process $X = (X_t)\, t \geq 0$ is adapted to $\mathcal{F}$ iff the random variable X_t is $\mathcal{F}_t$ measurable for all $t \geq 0$.

So by using the definition of $\mathcal{F}_t$ measurable, adapted means that at time t when the information in $\mathcal{F}_t$ is known then we also know the value of X_t since all the possible outcomes are in $\mathcal{F}$ and we can assign a probability to them.

Now consider a stochastic process $X_{t\,,t\in\{0,1,2\}}$ that we say is adaptive to the filtration $\mathcal{F}_{t\,,t\in\{0,1,2\}}$. This means that the information generated by the stochastic process (X) is contained in the full information set i.e $\mathcal{F}_t \in \mathcal{F}$.

- We also need the definition of a Martingale. An adapted stochastic process X with $E(|X_t|) < \infty$ for all $t > 0$ is:
  - a submartingale if $\forall\, t,\ s$ with $t > s$ we have $E(X_t|F_s) \geq X_s$
  - a supermartingale if $\forall\, t,\ s$ with $t > s$ we have $E(X_t|F_s) \leq X_s$
  - a martingale if $\forall\, t,\ s$ with $t > s$ we have $E(X_t|F_s) = X_s$

So to be clear a martingale is when the value in expectation (or the best possible prediction) at time t is the previous value, this means that you cannot make any predictions at time s of the value in time t.

A familiar example of a martingale is the symmetric random walk (the random walk without a trend) i.e at time s the next step in the process is completely unknown.

- A semi-martingale is the sum of a completely unpredictable path (a martingale) and a predictable component.

Now that we have an overview of the basic notation, assume a market on the probability space $(\Omega, \mathcal{F}_t, \mathcal{F}, \mathbb{P})$, where time $\tau > 0$ and with $n+1$ dividend paying assets $S(t) =$

[1] Note that $\{\omega : X(\omega) \leq x\}$ is the definition of a Cumulative Distribution Function (CDF) i.e it defines the probability space.

$S_0(t)...S_n(t)$, where $S(t)$ are traded continuously from time 0 to time τ and $S(t)$ can be described by the following stochastic processes:

$$dS_t = r_t S_t dt \tag{3.17}$$

Under the adaption of $\mathcal{F}_t$ we say that: $S_0(t) = 1$ and r_t is the instantaneous short rate (i.e. $r(t) = f(t,t)$ and tied to the instantaneous forward rate as discussed earlier:$f(t,\tau)$). We see that the return of S is:

$$\frac{dS(t)}{S(t)dt} = r_t \tag{3.18}$$

So in time t we completely know the return of S by just observing the prevailing short rate in the market and S is thus risk free at time t (locally risk-free). In addition we assume that $S_0(t) > 0 \; \forall t > 0$. We say that $S_0(t)$ is the risk free asset and thus the discounting factor $D(t,\tau)$ is given by $\frac{1}{S_0(t)}$. A portfolio or trading strategy equivalently is defined as[2]:

Definition 3.4. A portfolio contains a range of underlying assets in the financial market, denoted by $h(t) = [h_0(t), ..., h_n(t)]$, where the components $h_i(t)$ for $i = 0, ..., n$ are locally bounded and predictable. These components $h_i(t)$ represents the number of the underlying asset held at time t of asset i.

Here we have that the portfolio weights can be both positive (long position) or negative (short position) and not constants i.e. they can change as time goes by. The value created by the portfolio $h(t)$ at any given time t is defined in the value process:

Definition 3.5. The value process $V_{t \geq 0}$ associated with portfolio $h(t)$ is defined as:

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

So the value created by the portfolio are the weight times the asset summed over all assets/weights. And the gain process is

Definition 3.6. The gain process G_t associated with portfolio $h(t)$ is given by

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

A trading strategy (portfolio) is said to be self-financing if no additional value has to be added after the initial value $V^h(0)$ was supplied to the portfolio. So the changes in

[2]In the following definitions we follow Senghore (2013)

portfolio value only comes from changes in the assets already hold in the portfolio. We can define a self-financing portfolio in three ways:

Definition 3.7. A portfolio is self-financing if its value changes only due to changes in the prices of the underlying assets, expressed as:

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

Meaning that the change in portfolio value is only due to the change in S and h. Alternatively we can define the self-financing portfolio in terms of the gain process:

Definition 3.8. A portfolio $h(t)$ is self-financing if the value process $V(t) \geq 0$ follow:

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

Note that we can scale (discount) Definition 3.8 with the discount factor defined earlier and still define it as a self-financing portfolio:

Definition 3.9. The portfolio $h(t)$ is self-financing iff:

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

Previously we have stated the portfolio $(h(t))$ as a weighted sum of all the underlying assets $S_i(t)$, but we can state it as the relative weights i.e we scale each asset (multiplied by the weight) with the total value of the portfolio. We continue to follow Senghore (2013) and define it as:

Definition 3.10. For a given trading strategy, the relative portfolio weights u_i, which in general can be both $u_i \leq 0$ and $u_i \geq 0$, is given by the fraction of the total value from asset i. This can be expressed as:

$$u_i(t) = \frac{h_i(t)S_i(t)}{V^h(t)} \text{ , Where } i = 0, ..., n \text{ and } \sum_{i=0}^{n} u_i(t) = 1$$

We can also rewrite the self-financing portfolio that we defined earlier using the relative weights:

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

Remember that we said that an arbitrage was a trading strategy (portfolio) that satisfied either: (A) its price is zero now and it has a strictly positive payoff in the future, or (B) its price is strictly negative now and it may payout in a future state. Now we can define this in terms of the self-financing portfolio and the value process:

Definition 3.11. A self-financing portfolio is an arbitrage possibility such that for every $t > 0$ it follows:

$$
\begin{aligned}
V^h(0) &= 0 \\
P(V^h(t) \geq 0) &= 1 \\
P(V^h(t) \geq 0) &> 0
\end{aligned}
$$

If there exists arbitrage opportunities in the market we have an inefficient market. Normally, we assume frictionless, efficient financial markets. In order to satisfy efficiency we need that the value process of the dynamics of a portfolio to be locally risk-free and that the return with 100% certainty (probability = 1) equals the return on the bank account. This means that in a financial market that is free of arbitrage there can be only one short rate of interest.

Definition 3.12. For a locally risk-free self-financed portfolio $h(t)$ formulated as:

$$dV^h(t) = k(t)V^h(t)dt,$$

where $k(t)$ is any $\mathcal{F}_t$-adapted process

it must hold that the probability that the adapted process $k(t)$ is equal to the risk-free interest rate $r(t)$, is equal to one, formulated as:

$$P(k(t) = r(t)) = 1$$

which need to be true in order to attain an efficient financial market.

Furthermore, efficiency in the market can be stated by the relation between absence of arbitrage and the existence of a risk neutral probability measure, $\mathbb{Q}$, known as the equivalent martingale measure (EMM). This relation tells us that a market is arbitrage-free if, and only if, there exists at least one risk neutral probability measure $\mathbb{Q}$ that is equivalent to the original probability measure, $\mathbb{P}$. This is referred to as the: **First Fundamental Theorem of Asset Pricing**. We proceed by defining an EMM below

Definition 3.13. An EMM is a probability measure, $\mathbb{Q}$, on the space $(\Omega, \mathcal{F})$ such that:

1. The probability measure $\mathbb{P}$ and $\mathbb{Q}$ are equivalent, if and only if:
   $\mathbb{P}(A) = 0$ and
   $\mathbb{Q}(A) = 0$, for all $A \in \mathcal{F}$

2. The discounted price process $\frac{S_i}{S_0}$ are $\mathbb{Q}$-martingales, for all $i \in (0, \ldots, n)$ shown as:

   $$\mathbb{E}^{\mathbb{Q}}\left[D(o,t)S_i(t)|\mathcal{F}_u\right] = D(o,u)S_i(u)$$

   for all $i \in (0, \ldots, n)$ and all $0 \leq u \leq t \leq T$,
   where $\mathbb{E}^{\mathbb{Q}}$ denotes the expectation under $\mathbb{Q}$

3. The Randon–Nikodym derivative $\frac{d\mathbb{Q}}{d\mathbb{P}}$ is quadratically integrable with respect to $\mathbb{P}$
   The Randon-Nikodym derivative, referred to as the **likelihood ratio (L)** between the equivalent measure $\mathbb{Q}$ and the probablity measure $\mathbb{P}$, is generally used to enable movement to and from these two measures. That $L = \frac{d\mathbb{Q}}{d\mathbb{P}}$ is quadratically integrable implies that the following integral is finite:

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

The **Second Fundamental Theorem of Asset Pricing** tells us that, if and only if, there exists a unique risk-neutral measure (an EMM, $\mathbb{Q}$) that is equivalent to the probability measure $\mathbb{P}$ we also have a **complete** market.

Now we have defined the fundamentals of an arbitrage free market, so we continue by considering the financial derivatives or **contingent claims**, which are the traded assets in that market. The objective here is to price these assets. A contingent claim can be seen as a contract where the holder of the contract receives a deterministic payoff, at a pre-specified date in the future, the exercise date. Note that the payoff can be positive, negative or zero. Contingent claims are completely defined in terms of their underlying assets, in our case the interest rates but they can also be written on other underlying assets, e.g. stocks, bonds, or other financial assets.

Lets assume we have a financial market as defined above with the vector price process:

$$S = [S_0, \ldots, S_n]^T$$

where T is the exercise date of a contingent claim, any stochastic variable $\mathcal{X}$ such that $\mathcal{X} \in \mathcal{F}_S^T$. This expression means that it is possible to derive the price of $\mathcal{X}$ at time T. In other words we can price the contingent claim at the exercise date.

T-claim is just another word for a contingent claim, or a simple claim if it is of the below form (with Φ being the contract function):

$$\mathcal{X} = \Phi(S(T))$$

. If there exists a self-financing portfolio h with a value at the maturity T that is equal to a contingent claim $\mathcal{X}$, we say that it is attainable:

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

Another word for these attainable claims is hedgeable claims. If we have an attainable claim and there is a self-financing portfolio h, it implies that this claim can be traded in the financial markets. If all claims $\mathcal{X}$ in the market are attainable it is said that the market is complete. As mentioned earlier market completeness means that there exist a unique equivalent measure.

If we assume an efficient market, the contingent claim at time t, $\Pi(t, \mathcal{X}$ can be determined by either: demand consistency of the price of the underlying asset and the price of the contingent claim – or recognize that the value process at time t, generated by the self-financing portfolio, must be equal to the price of the contingent claim at time t. So if we have an efficient market, then:

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

Determining the price of a contingent claim by demanding consistency of the price of the claim with that of the underlying implies the existence of a martingale measure $\mathbb{Q}$ for the extended market $\Pi(; \mathcal{X}), S_0, \ldots, S_n)$. This guides us into the martingale approach for derivatives pricing, an effective method for pricing financial instruments. By applying the definition of a martingale measure under the $\mathbb{Q}$ we are able to obtain the general pricing formula for $\mathbb{Q}$, where the general arbitrage free pricing formula is defined as:

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

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

If we assume that $S_0(t)$ above is the risk less money account, we can re-write it as:

$$S_0(t) = S_0(0) \times \exp^{\int_0^t r(s)ds}$$

where r represents the short rate.

Given the above assumption we end up with the Risk Neutral valuation Formula **RNVF**, which is a martingale approach to derivative pricing..

3.9 Risk Neutral valuation Formula

Using this martingale approach we can price a contingent claim by taking the expectation, under the $\mathbb{Q}$-martingale measure, of the discounted claim with the money account as a numeraire, given an adapted filtration. This gives us the following:

$$\Pi(t, \mathcal{X}) = \mathbb{E}^{\mathbb{Q}} \left[\frac{\mathcal{X} B(t)}{B(T)} | \mathcal{F}_t \right] = \mathbb{E}_{\mathbb{Q}} \left[\exp^{\int_0^t r(s)ds} \times \mathcal{X} | \mathcal{F}_t \right].$$

The result yielded above implies that the discounted price process is a $\mathbb{Q}$-martingale measure and the existence of a $\mathbb{Q}$-martingale measure implies absence of arbitrage in the market. Once again this is evidence that the martingale property can be interpreted as being the same as the existence of a self-financing portfolio.

For stochastic underlying assets though, the bank account is no longer a suitable discounting factor. The joint distribution $(B(T), \Phi(S))$, needed in order to solve for the expectation under the $\mathbb{Q}$-measure of the RNVF is difficult to compute, because it involves solving of a double integral. Therefore, it is necessary to change the numeraire.

Chapter 4

The Nelson-Siegel Models

4.1 The Static Nelson-Siegel Model

Nelson and Siegel (1987) introduced a three-factor yield curve model for fitting to static yield curves. The Nelson-Siegel model originates from a constant and the solution to a second order differential equation with constant coefficients, when the roots of the polynomial are real and equal. When this is true we can write the instantaneous forward rate as:

$$f(t,\tau) = \beta_1 + \beta_2 e^{-\lambda\tau} + \beta_3 \lambda e^{-\lambda\tau} \tag{4.1}$$

To find the relationship between the instantaneous forward rate and the continuous rate with maturity τ consider the growing factor $[\tau, \tau + \delta\tau]$:

$$\frac{P(0,\tau)}{P(0,\tau+\Delta\tau)}, \Delta\tau > 0 \tag{4.2}$$

Where $P(t,\tau)$ is the price at time t of one (1) amount of money that will be paid at time τ. In the case $\Delta\tau \to 0$ we have that:

$$1 + f(\tau)\Delta\tau \approx \frac{P(0,\tau)}{P(0,\tau+\Delta\tau)} \Rightarrow f(\tau) = \lim_{\Delta\tau\to 0} \frac{P(0,\tau) - P(0,\tau+\Delta\tau)}{\Delta\tau P(0,\tau+\Delta\tau)}$$

$$f(\tau) = \lim_{\Delta\tau\to 0} -\left(\frac{P(0,\tau+\Delta\tau) - P(0,\tau)}{\Delta\tau}\right)\frac{1}{P(0,\tau+\Delta\tau)}$$

$$f(\tau) = -\frac{\delta P(0,\tau)}{\delta\tau}\frac{1}{P(0,\tau)} \tag{4.3}$$

We know that for continuous compounding with unit amount is $A(0,t) = e^{-tR(t)}$ and using this we have:

$$P(0,\tau) = e^{-\tau F(\tau)} \Rightarrow \frac{\delta P(0,\tau)}{\delta\tau} = -e^{-\tau F(\tau)}F'(\tau)\tau$$

Plug this into 4.3 and we end up with:

$$\begin{aligned} f(\tau) = F'(\tau)\tau e^{-\tau F(\tau)}\frac{1}{e^{-\tau F(\tau)}} \\ f(\tau) = F'(\tau)\tau \end{aligned}$$

$$F'(\tau) = \frac{1}{\tau}f(\tau) \tag{4.4}$$

Aggregate this and and set $F(\tau) = y(\tau)$ we have:

$$y(\tau) = \frac{1}{\tau}\int_0^\tau f(x)\mathrm{d}x \tag{4.5}$$

This implies that the zero-coupon yield is just a weighted average of the forward rates. Plug in the original Nelson and Siegel equation (from equation 4.1) and we have:

$$y(\tau) = \frac{1}{\tau}\int_0^\tau \beta_1 + \beta_2 e^{-\lambda\tau} + \beta_3\lambda e^{-\lambda\tau}\mathrm{d}\tau \tag{4.6}$$

Make the following substitutions, $x = \lambda\tau, d\tau = \frac{1}{\lambda}dx$ and we get.

$$\begin{aligned} y(\tau) &= \beta_1 + \frac{\beta_2}{\lambda\tau}\int_0^{\lambda\tau} e^{-x}\mathrm{d}x + \frac{\beta_3}{\lambda\tau}\int_0^{\lambda\tau} xe^{-x}\mathrm{d}x \\ y(\tau) &= \beta_1 + \frac{\beta_2}{\lambda\tau}\left[-e^{-x}\right]_0^{\lambda\tau} + \frac{\beta_3}{\lambda\tau}\left[-xe^{-x} - e^{-x}\right]_0^{\lambda\tau} \\ y(\tau) &= \beta_1 + \frac{\beta_2}{\lambda\tau}\left(1 - e^{-\lambda\tau}\right) + \frac{\beta_3}{\lambda\tau}\left(1 - \lambda\tau e^{-\lambda\tau} - e^{-\lambda\tau}\right) \end{aligned}$$

$$y(\tau) = \beta_1 + \beta_2\left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau}\right) + \beta_3\left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right) \tag{4.7}$$

4.2 Dynamic Nelson-Siegel Model (DNS)

In a dynamic setting we let β vary over time by adding subscript time factor t on both the rate process and the factors. Thus, we get the final Dynamic Nelson-Siegel three factor model represented as in Diebold and Li (2006):

$$y_t(\tau) = \beta_{1t} + \beta_{2t}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) + \beta_{3t}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right) \tag{4.8}$$

Note that this simple model is linear in the parameters β_{it}, when keeping λ constant. Below we, also present the equation for when the model uses a time dependent λ_t. The only difference in the equation being the subscript t on lambda, referring to a time varying decay parameter.

$$y_t(\tau) = \beta_{1t} + \beta_{2t}\left(\frac{1-e^{-\lambda_t\tau}}{\lambda_t\tau}\right) + \beta_{3t}\left(\frac{1-e^{-\lambda_t\tau}}{\lambda_t\tau} - e^{-\lambda_t\tau}\right) \tag{4.9}$$

The DNS model has one dependent variable $y_t(\tau)$ and four independent variables ($\beta_{1t}, \beta_{2t}, \beta_{3t}, \lambda_t$). The majority of the literature does however support the use of λ as a constant, and therefore fit the model by solving the three β:s keeping λ fixed, which allows for linear regression analysis. We are running the model both with fixed λ and with lambda calibrated for each regression. Thus, we can investigate the difference in performance. Of course, there is a trade-off between keeping it simple and reaching for the results, as we will discuss subsequently.

In order to have a successful model for the yield curve it is necessary for it to represent the characteristics and facts of the historical yield curve behavior as well as being able to adapt to sudden shocks to the yield curve, without being over complicated. Nelson-Siegel provides an elegant and easily interpreted solution to the dynamics of the yield curve.

Diebold and Li (2006) specifically mention that a good model for the yield curve should be able to reproduce the historical stylized facts about (among others):

1. The average shape of the yield curve.
2. The variety of shapes assumed at different times.
3. The strong persistence of yields and weak persistence of spreads.

Below we will go through the dynamics of the DNS model. First, the model is fitted for a given t where we observe multiple rates with different maturity τ. These rates are captured in $y_t(\tau)$. In the model framework the rates will be more present at the shorter end and not as much at longer time frames. This is a behavior which is also observed in the markets, where most trading are done in shorter maturities.

Further, the model should also be able to reproduce the average historical yield curve, which is increasing and concave. Meaning that short term money is cheaper than long term money. The average yield curve in the Nelson-Siegel model is determined by the average of the parameters $\beta_{1t}, \beta_{2t}, \beta_{3t}$, which allow for a shape of that sort. However, under some circumstances the yield curve may be of a different shape, it could even flip i.e short term money is more expensive than long term money. This can occur for some shocks to the market expectations of the mid-term rates in the economy and the model need to be able to represent such shapes as well. Since the Nelson-Siegel model comes from a second order differential equation, such a variety of shapes are certainly possible. Persistent yield dynamics should according to Diebold and Li (2006) be represented by a strong persistence of β_{1t} and a weaker persistence of β_{2t}.

In the literature the three factors of the model (corresponding to $\beta_{1t}, \beta_{2t}, \beta_{3t}$) has been described as representing the level, slope and curvature of the yield curve. Moving to a state-space representation and changing the notation to highlight the interpretation of the DNS factors, we follow Diebold et al. (2006), and depict the DNS measurement equation in matrix form:

$$\begin{pmatrix} y_t(\tau_1) \\ y_t(\tau_2) \\ \vdots \\ y_t(\tau_N) \end{pmatrix} = \begin{pmatrix} 1 & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} - e^{-\lambda\tau_1} \\ 1 & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} - e^{-\lambda\tau_2} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} - e^{-\lambda\tau_N} \end{pmatrix} \begin{pmatrix} L_t \\ S_t \\ C_t \end{pmatrix} + \begin{pmatrix} \varepsilon_\tau(\tau_1) \\ \varepsilon_\tau(\tau_2) \\ \vdots \\ \varepsilon_\tau(\tau_N) \end{pmatrix} \tag{4.10}$$

where the measurement errors $\varepsilon_\tau(\tau_i)$ are assumed to be independently and identically distributed (i.i.d.) white noise. When collapsing to matrix notation we get the equation:

$$\underset{N\times 1}{Y_t} = \underset{N\times 3}{Z_t} \underset{3\times 1}{\begin{pmatrix} L_t \\ S_t \\ C_t \end{pmatrix}} + \underset{N\times 3}{e_t} \tag{4.11}$$

Examining the matrix notation, the N dimensions of Z_t change over time depending on number of market contracts observed (for different maturities) at time t. Changes in L_t are constant across all maturities and the loading on L_t determines the level of the curve. The loading on S_t determines the overall shape. More specifically, the factor loading for β_{1t} is 1 and is therefore said to represent the long term behavior (more specific,

$y(\infty) = \beta_{1t}$) and referred to as the level. The second term β_{2t} is associated with $\frac{1-e^{-\lambda\tau}}{\lambda\tau}$ which starts at 1 for $\tau = 0$ and decays to 0 as τ increases, hence it is a short term factor or slope as Diebold and Li (2006) frame it. Finally the last term β_{3t} is associated with $\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}$ which starts at 0 and then increases before decaying for the longest maturities, reflecting the medium term behaviour.

The term λ is called the decay parameter and determines the speed of the decay of the loadings for β_{2t} and β_{3t}. Small values of λ will reduce the speed of decay and produce a better fit for longer maturities while a larger value for λ will increase the speed of decay and better fit the curve at shorter maturities. In addition λ also governs where the factor loadings for β_{3t} achieves its maximum. To illustrate this Figure 4.1 shows the factor loadings of the model using different values of λ.

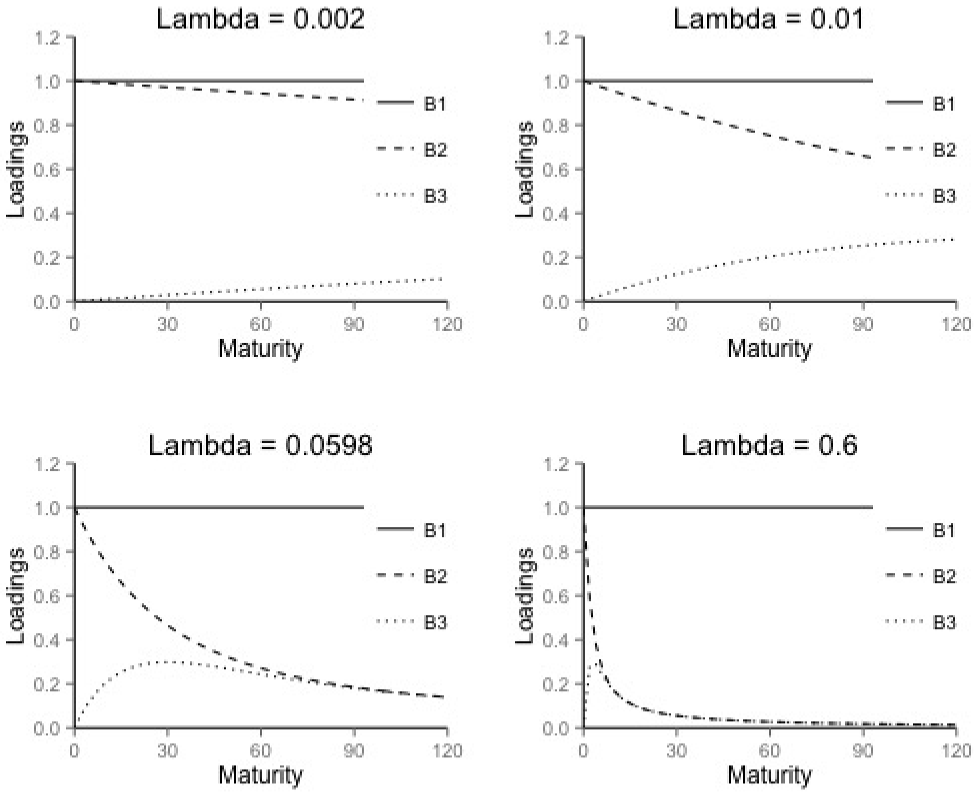


FIGURE 4.1: Factor loadings Nelson-Siegel Model

As seen in Figure 4.1 variations of λ cause large variations in the factor loadings connected to the three factors. It is therefor important to carefully specify an optimal value for λ. Diebold and Li (2006) argue that, since λ governs where the mid-term factor or curvature has its maximum and that the consensus for this is around the 2-3 year point. Diebold and Li (2006) therefore picks the average in this span i.e. $\tau = 30$ months and solve for λ.

Recall that the curvature factor is specified as:

$$F_{\beta_3} = \frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}$$

Take the first derivative with respect to λ:

$$\frac{\partial F_{\beta_3}}{\partial \lambda} = \frac{e^{-\lambda\tau}(1 + \lambda\tau) - 1}{\lambda^2\tau} + \tau e^{-\lambda\tau} \tag{4.12}$$

Plug in $\tau = 30$ and put to zero:

$$0 = \frac{e^{-30\lambda}(1 + 30\lambda) - 1}{30\lambda^2} + 30e^{-30\lambda} \tag{4.13}$$

Which we then solve for using Matlab's solve function and get $\lambda \approx 0.0598$. The resulting λ does not match Diebold and Li (2006) result of $\lambda \approx 0.0609$. However, if we plug in 0.0609 and 0.0598 in 4.13 we get:

$$\frac{e^{-30*0.0609}(1 + 30*0.0609)}{30*0.0609^2} + 30e^{-30*0.0609} = -0.07271402$$
$$\frac{e^{-30*0.0598}(1 + 30*0.0598)}{30*0.0598^2} + 30e^{-30*0.0598} = -0.001584571$$

Which support our result. In any case the use of 0.0609 or 0.0598 will have minimal impact on the result. We will proceed with $\lambda = 0.0598$ and later come back to the choice of a constant value versus a time dependent value of lambda.

Further, as previously mentioned the model will, for very long maturities reduce to β_{1t} and thus we can proxy the level by the ten year rate. To illustrate that β_{2t} is attached to the yield curve slope we follow Diebold and Li (2006) and proxy the slope factor by considering the difference between the yield of the ten year rate and that of the three month rate $y_t(120) - y_t(3) = -0.78\beta_{2t} + 0.06\beta_{3t}$ where we have set $\lambda = 0.0598$. Note that the slope depends heavily on the value for β_{2t}.

The last factor to be discussed is the one we call the curvature factor, which is associated to β_{3t}. By adding a curvature factor to the model we can allow for significant yield curve variation. We can for example allow for a hump in the curve at medium term rates, which would be the case if the medium term rates are significantly higher than the short and long term rates. We can again proxy for this medium term factor (or curvature) by finding a similar representation for the curvature as we did for the slope. Yet again following Diebold and Li (2006) we define it as twice the two-year yield minus the sum of the ten-year and three-month yields i.e. $2y_t(24) - y_t(3) - y_t(120) = 0.007\beta_{2t} + 0.368\beta_{3t}$.

4.3 Dynamic Nelson-Siegel Svensson model (DNSS)

The Nelson-Siegel Svensson model is an extended version of the original three factor Nelson-Siegel model, by adding a second curvature term, making it a four-factor model. This additional curvature term allows for more flexibility and often better in-sample fit at long maturities as compared to the three-factor Nelson-Siegel model (Diebold and Rudebusch (2012). The additional gain in flexibility is one of the reasons for the popularity of the Svensson extended model amongst Central Banks when it comes to modeling, estimating and forecasting the term structure of interest rates (BIS (2005)). Letting $y(\tau)$ be the zero rate for maturity τ, we have the following representation:

$$y_t(\tau) = \beta_{1t} + \beta_{2t}\left(\frac{1-e^{-\lambda_1\tau}}{\lambda_1\tau}\right) + \beta_{3t}\left(\frac{1-e^{-\lambda_1\tau}}{\lambda_1\tau} - e^{-\lambda_1\tau}\right) + \beta_{4t}\left(\frac{1-e^{-\lambda_2\tau}}{\lambda_2\tau} - e^{-\lambda_2\tau}\right), \tag{4.14}$$

where $\beta_1, \beta_2, \beta_3$, and β_4 are the model factors and λ_1 governs the rate of exponential decay of β_2 and β_3, while λ_2 governs the exponential growth and decay rate of the second curvature term β_4. This is easy to see when stating the equation in matrix form:

$$\begin{pmatrix} y_t(\tau_1) \\ y_t(\tau_2) \\ \vdots \\ y_t(\tau_N) \end{pmatrix} = \begin{pmatrix} 1 & \frac{1-e^{-\lambda_1\tau_1}}{\lambda_1\tau_1} & \frac{1-e^{-\lambda_1\tau_1}}{\lambda_1\tau_1} - e^{-\lambda_1\tau_1} & \frac{1-e^{-\lambda_2\tau_1}}{\lambda_2\tau_1} - e^{-\lambda_2\tau_1} \\ 1 & \frac{1-e^{-\lambda_1\tau_2}}{\lambda_1\tau_2} & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} - e^{-\lambda_1\tau_2} & \frac{1-e^{-\lambda_2\tau_1}}{\lambda_2\tau_1} - e^{-\lambda_2\tau_1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\lambda_1\tau_N}}{\lambda_1\tau_N} & \frac{1-e^{-\lambda_1\tau_N}}{\lambda_1\tau_N} - e^{-\lambda_1\tau_N} & \frac{1-e^{-\lambda_2\tau_1}}{\lambda_2\tau_1} - e^{-\lambda_2\tau_1} \end{pmatrix} \begin{pmatrix} L_t \\ S_t \\ C_t^1 \\ C_t^2 \end{pmatrix} + \begin{pmatrix} \varepsilon_\tau(\tau_1) \\ \varepsilon_\tau(\tau_2) \\ \vdots \\ \varepsilon_\tau(\tau_N) \end{pmatrix} \tag{4.15}$$

where the measurement errors $\varepsilon_\tau(\tau_i)$ are assumed to be (i.i.d.) white noise.

As seen in Christensen et al. (2009) the two curvature factors take on very different roles in fitting the yield curve. In their paper λ_1 is estimated at 0.838, implying that the factor loading of the first curvature factor peaks around the 2-year maturity. The factor loading of the second curvature factor peaks around the 19-year maturity, since λ_2 is estimated at a much smaller value, 0.097. Thus, two curvature factors make it possible to fit yields with more than one local minimum/maximum along various maturities.

Furthermore, Christensen et al. (2009) note that adding the second curvature factor only affects the level factor in the model. They explain it as, without the second curvature factor, corresponding to the three factor DNS model, only the level factor is able to fit yields for long-term maturities. However, when including the second curvature factor, which is able to fit yields with maturities in the range of 10 to 30 years, the level factor is now allowed to fit other areas of the yield curve.

Since the DNSS model has two decay parameters, λ_1 and λ_2, the model is non-linear. This non-linearity is highlighted by a high degree of multi-collinearity between factors, especially when λ_1 takes on a similar value as λ_2, thereby making it impossible to separately identify the curvature factors, β_3 and β_4. Thus the difficulty in the estimation of model parameters.

4.4 Constant versus time dependent lambda

The literature, and especially Diebold and Li (2006), argue for keeping lambda constant. This in order to retain simplicity in the estimation technique used to fit and forecast yields, by allowing for linear regression analysis. They also provide some supporting evidence for doing so. As mentioned earlier, the decay parameter, lambda, determines the speed of the decay of the loadings for β_{2t} and β_{3t}. Figure 4.1 shows that small values of λ will reduce the speed of decay and produce a better fit for longer maturities while a larger value of λ will increase the speed of decay and better fit the curve at shorter maturities. In addition λ also governs where the factor loadings for β_{3t} achieves its maximum. So, there are a collection of areas affected by the size of lambda.

Diebold and Li (2006) found that letting lambda be time dependent causes some variation over time in the estimated value of lambda. However, they claim that the variation is small relative to the standard error. Nelson and Siegel (1987) also state that the sum-of squares function is not very sensitive to variations in lambda. Diebold and Li (2006) state that these findings together suggest that fixing lambda would have little, to no, drawback. However, they admit that allowing lambda to vary over time can improve the fit significantly when the short end of the yield curve is steep, which tend to happen from time to time.

According to Koopman et al. (2010) most empirical studies assume lambda to be fixed at some constant known value or estimated as constant through time. However, when investigating the matter Koopman et al. (2010) found that when estimating lambda the data can be highly informative about the shape of the factor loadings, and by applying a simple step function and a spline function they show considerable evidence that lambda in fact is a time-varying parameter. Further, they suggest that if possible, one should also treat lambda that way. Especially, since keeping lambda fixed over the entire sample period may be too restrictive as the characteristic of the yield curve may change over a longer time period and lambda have a great impact of the shape of the yield curve.

Diebold and Li (2006) state that even though a time varying lambda might result in a slightly better in-sample fit, it does not necessarily produce a better out-of-sample forecasting. In fact little attention has been paid the forecasting effects of a time-varying decay parameter. Koopman et al. (2010) show that forecasting results are affected by these modifications of the baseline model, however, they provide ambiguous results. Some forecasting results are improved and some are not, but to what extent is not clear. In the next two chapters we test and compare both the in-sample fit and the out-of-sample forecasting performance for when lambda is pre-specified and when lambda is allowed to vary over time. These are also compared with other benchmark competitors.

4.5 Why are the Nelson-Siegel models not arbitrage-free?

Despite its proven empirical performance and ease of use the Nelson-Siegel model fails on one important theoretical dimension. It does not theoretically enforce absence of arbitrage. The reason for this is the same as for affine term structure family of models, i.e. the estimated yield curve does not match the actual yield curve. Heath et al. (1992b) indicate that most of the affine term structure models price zero-coupon bonds according to the parameters obtained from the endogenously derived yield curve, which results in that it could differ from the actual market curve. Consequently, the resulting bond prices could be different than the actual market prices, and thus leaving space for arbitrage opportunities.The technical details of this matter is shown in e.g. Björk and Christensen (1999).

However, according to Diebold and Rudebusch (2012) the Dynamic Nelson-Siegel model, illustrated in Equation 4.8, is almost arbitrage-free as it only requires an additional term that accounts for Jensen's inequality to make it entirely arbitrage-free. So, in order to answer the question why the Nelson-Siegel class models are not arbitrage-free, we provide what is required in the model specification and how it is possible to actually make it arbitrage-free. Although, we do not include the no-arbitrage version of the Nelson-Siegel model in our forecasting performance evaluation, we emphasize this section for completeness and understanding of empirical yield curve fitting.

The deterministic foundation of the Nelson-Siegel model leads to the breakdown of arbitrage-freeness, since it does not account for the convexity arising from Jensen's inequality effects when having stochastic yield factors. In other words, the Nelson-Siegel models cannot theoretically enforce no-arbitrage because there is a missing link between the parameters of the state equation and the parameters of the measurement equation, meaning that the parameters that determine the dynamic evolution of the yield curve factors are not linked to the parameters that determine the shape and location of the yield curve. Therefore it does not ensure consistency between the dynamic evolution of yields over time, and the shape of the yield curve at a given point in time (Diebold and

Rudebusch (2012)). This crucial piece needed is a time-invariant yield-adjustment term, which only depends on maturity, and takes the form:

$$\frac{C(t,T)}{T-t}$$

This term connects the state equations to the measurements equations, which assures that discounted bond prices are semi-martingales and hence arbitrage-free.

As shown in Christensen et al. (2009) it is possible to derive an Arbitrage Free Nelson–Siegel (AFNS) model that corresponds to the DNS model specification, using Proposition 2.1. (See Appendix A for more details.) The AFNS model is formulated in continuous time and the relationship between the real-world dynamics represented by the P-measure and the risk-neutral dynamics represented by the Q-measure is given by the measure change:

$$dW_t^Q = dW_t^P + \Gamma_t d_t \tag{4.16}$$

where Γ_t represent specification of the risk premium. In order to preserve affine dynamics under the P-measure we follow Christensen et al. (2009), and limit the focus to essentially affine risk premium specification a la Duffee (2002). Therefore, Γ_t is formulated as:

$$\Gamma_t = \begin{pmatrix} \gamma_1^0 \\ \gamma_2^0 \\ \gamma_3^0 \end{pmatrix} + \begin{pmatrix} \gamma_{11}^1 & \gamma_{12}^1 & \gamma_{13}^1 \\ \gamma_{21}^1 & \gamma_{22}^1 & \gamma_{23}^1 \\ \gamma_{31}^1 & \gamma_{32}^1 & \gamma_{33}^1 \end{pmatrix} \begin{pmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{pmatrix}.$$

Using this specification of the risk premium, the stochastic differential equation (SDE) for the state variables under the P-measure remains affine, as follows from equation,

$$dX_t = K^P \left[\theta^P - X_t\right] dt + \Sigma dW_t^P. \tag{4.17}$$

As a result of the flexible specification of Γ_t, one can choose any mean vector θ^P and mean-reversion matrix K^P under the P-measure, still keeping the required risk-neutral structure, as described in Proposition 2.1 (Appendix A). Assuming all three factors are independent under the P-measure the following AFNS model corresponds to the DNS model introduced in section 4.3:

$$\begin{pmatrix} dX_t^1 \\ dX_t^2 \\ dX_t^3 \end{pmatrix} = \begin{pmatrix} K_{11}^P & 0 & 0 \\ 0 & K_{22}^P & 0 \\ 0 & 0 & K_{33}^P \end{pmatrix} \left[\begin{pmatrix} \theta_1^P \\ \theta_2^P \\ \theta_3^P \end{pmatrix} - \begin{pmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{pmatrix} \right] dt + \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{pmatrix} \begin{pmatrix} dW_t^{1,P} \\ dW_t^{2,P} \\ dW_t^{3,P} \end{pmatrix}.$$

Under these specifications the measurement equation for the AFNS model takes the form:

$$\begin{pmatrix} y_t(\tau_1) \\ y_t(\tau_2) \\ \vdots \\ y_t(\tau_N) \end{pmatrix} = \begin{pmatrix} 1 & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} - e^{-\lambda\tau_1} \\ 1 & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} - e^{-\lambda\tau_2} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} - e^{-\lambda\tau_N} \end{pmatrix} \begin{pmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{pmatrix} - \begin{pmatrix} \frac{C(\tau_1)}{\tau^1} \\ \frac{C(\tau_2)}{\tau^2} \\ \vdots \\ \frac{C(\tau_N)}{\tau^N} \end{pmatrix} + \begin{pmatrix} \varepsilon_\tau(\tau_1) \\ \varepsilon_\tau(\tau_2) \\ \vdots \\ \varepsilon_\tau(\tau_N) \end{pmatrix}, \tag{4.18}$$

where the measurement errors $\varepsilon_t(\tau_i)$ are assumed to be i.i.d. white noise.

Now, comparing the specification of the measurement equation for the AFNS model above, to the DNS model (Equation 4.8 and 4.9) it is evident that the only difference is the yield-adjustment term. Given this additional term we get a model that not only in practice, but also in theory enforce the no-arbitrage condition. Note, that referring to the Nelson-Siegel model as arbitrage free in practice refers to studies conducted by Coroneo et al. (2011) and Diebold and Rudebusch (2012).

Chapter 5

Data and Estimation

5.1 The Data

In our paper we use daily quotes of LIBOR rates with maturities of $1, 2, 3, 6, 9, 12, 24, 36, 48, 60, 72$ and 120 months, provided by Nordea. For longer maturities the swap rate has been used, but throughout the thesis we will refer to all the rates as LIBOR rates. The data ranges from January 1, 2003 through March 3, 2015, resulting in approximately 36 000 observations, depending on, which of the 37 different currencies to employ. Table 5.1 presents descriptive statistics for the yields of the different LIBOR rates for each currency at 24 months maturity. In order to proceed, we will due to space and time constraint provide a deeper analysis of two currencies, and for forecasting purpose we proceed with only one. A natural first choice being the local currency of Danish Krone (DKK) and the second falling on US Dollars for the purpose of being a large reference currency. Both of the above mentioned having data for the full period. The reference currency is chosen to fully test the liability and quality of our models. Notable in Table 5.1 is that the more stable group of currencies has a mean of around 2-3% and a standard deviation of around 1-2 %. The Japanese Yen (JPY) with a mean of only 0.462 and a standard deviation of 0.319 really stands out from the rest and could be seen as the most stable currency because of the low volatility, bear in mind though that fixing JPY was a major part of the LIBOR scandal, which in turn could have influenced the low standard deviation. USD has a mean of 2.173% and a standard deviation of 1.696%. Observing the less stable currencies the Argentine Peso stands out with a standard deviation of 5.536% and also among the highest mean of 11.346%. Comparing DKK and USD makes them seem fairly similar, however note that DKK has got minimum values of below zero, which is another way of testing the model specification on how it handles negative rates. The DKK LIBOR rate shows a steadily declining curve since the peak in late 2008. In our data set the 1-month DKK LIBOR rate displays a minimum of -0.7 in February 2015.

Table 5.1: Descriptive statistics for LIBOR yield for all currencies at a maturity of 24 months

Currency	Mean	Standard deviation	Minimum	Maximum
AED	2.639	1.506	0.774	5.831
ARS	11.346	5.536	0.051	14.929
AUD	4.842	1.424	2.125	8.161
BRL	10.671	1.647	7.165	16.545
CAD	2.474	1.198	0.759	4.995
CHF	1.085	0.978	-1.006	3.496
CNY	2.147	1.390	-1.788	4.657
COP	9.539	0.007	9.529	9.562
CZK	2.226	1.100	0.283	4.742
DKK	2.395	1.350	-0.146	5.737
EUR	2.159	1.326	0.095	5.351
GBP	3.070	1.884	0.557	6.412
HKD	1.936	1.459	0.368	5.032
HUF	6.582	1.982	1.785	11.365
IDR	7.736	1.837	4.055	14.446
ILS	10.239	0.823	6.335	10.470
INR	6.989	3.861	2.506	16.770
ISK	8.491	3.043	4.949	14.521
JPY	0.462	0.319	0.087	1.418
KWD	2.526	1.738	0.703	9.177
LTL	3.365	2.423	0.403	11.513
LVL	4.559	4.129	0.422	17.690
MXN	6.179	1.820	0.000	10.969
MYR	3.255	0.430	1.985	4.769
NOK	3.324	1.238	0.933	6.659
NZD	5.152	1.821	2.362	8.594
PHP	2.760	1.560	0.161	7.459
PLN	4.663	1.188	1.527	7.650
RON	4.956	4.967	-0.135	100.862
RUB	7.718	2.536	5.064	16.647
SAR	2.397	1.763	-0.097	5.696
SEK	2.523	1.205	-0.015	5.559
SGD	1.581	1.012	0.318	3.815
THB	3.246	1.183	1.634	6.308
TRY	11.830	3.779	5.560	19.728
USD	2.173	1.696	0.342	5.640
ZAR	7.536	1.671	4.852	12.409

In Table 5.2 we present some descriptive statistics for the yield curves of the USD LIBOR for different maturities. Some results are worth mentioning. We see that the curve is upward sloping, that the long rates are less volatile and more persistent compared to the short rates, that the level (defined as the yield curve for $\tau = 120$ months) is highly persistent but varies the least in relationship to its mean. Further we note that the slope factor is less persistent than the level but that the curvature factor is the least persistent among the factors. We also note that the pairwise correlations between the level,slope are not high, the highest being ≈ 0.5.

TABLE 5.2: Descriptive statistics for LIBOR yield for USD

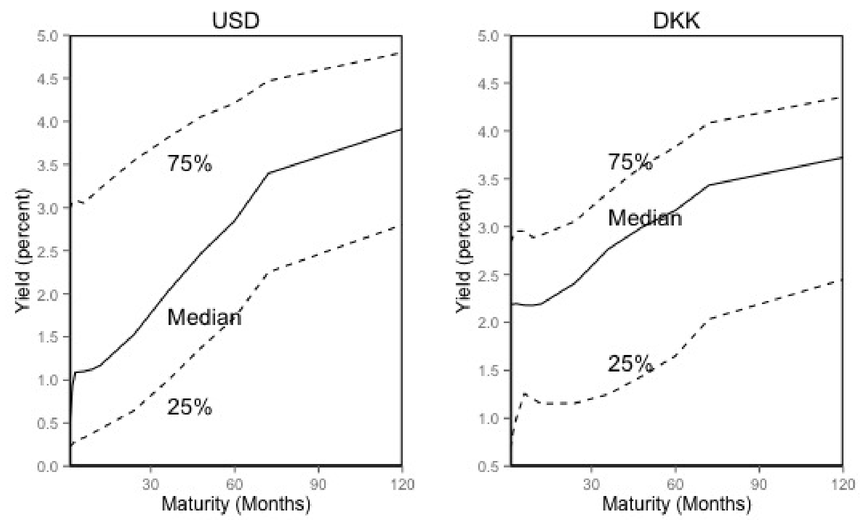
Maturity	Mean	Std	Minimum	Maximum	$\hat{\rho}(1)$	$\hat{\rho}(12)$	$\hat{\rho}(30)$
1	1.699	1.896	0.143	5.867	0.979	0.510	-0.229
2	1.766	1.887	0.185	5.795	0.980	0.525	-0.234
3	1.814	1.875	0.219	5.754	0.981	0.529	-0.235
6	1.848	1.859	0.224	5.616	0.986	0.544	-0.222
9	1.883	1.837	0.233	5.655	0.986	0.558	-0.208
12	1.926	1.810	0.249	5.667	0.986	0.574	-0.192
24	2.173	1.696	0.342	5.640	0.983	0.636	-0.117
36	2.468	1.598	0.424	5.636	0.978	0.676	-0.064
48	2.747	1.505	0.554	5.651	0.972	0.689	-0.026
60	2.998	1.418	0.737	5.672	0.965	0.686	0.004
72	3.400	1.277	1.144	5.707	0.953	0.661	0.051
120 (Level)	3.799	1.156	1.570	5.801	0.940	0.614	0.099
Slope	1.985	1.220	-0.755	4.176	0.932	0.154	-0.482
Curvature	-1.265	0.753	-3.316	0.310	0.880	0.194	-0.126

In Table 5.3 we present descriptive statistics for DKK yield for all maturities along with slope and curvature factors. Similar to the USD yields see that the curve is upward sloping, that the long rates are less volatile and more persistent compared to the short rates and the same tendency regarding level. Here the short rates is less persistent and also more volatile compared to the long rates. The level (defined as the yield curve for $\tau = 120$ months) is still highly persistent among the factors and and varies only moderately relative to its mean, while the slope factor varies the most relative to its mean. Highest pairwise correlation being 0.32. Note that we see negative minimum rates for all maturities up to 36 months. The lowest rate of -0.7 was noted on February 12, 2015 for the shortest one month rate.

TABLE 5.3: Descriptive statistics for LIBOR yield for DKK

Maturity	Mean	Std	Minimum	Maximum	$\hat{\rho}(1)$	$\hat{\rho}(12)$	$\hat{\rho}(30)$
1	1.985	1.552	-0.700	6.133	0.973	0.488	-0.076
2	2.067	1.552	-0.627	6.302	0.974	0.477	-0.091
3	2.128	1.536	-0.581	6.169	0.974	0.468	-0.102
6	2.224	1.432	-0.393	5.804	0.975	0.458	-0.101
9	2.238	1.413	-0.342	5.655	0.974	0.474	-0.097
12	2.259	1.398	-0.291	5.717	0.972	0.489	-0.092
24	2.395	1.350	-0.146	5.737	0.965	0.529	-0.053
36	2.578	1.322	-0.035	5.641	0.962	0.555	-0.038
48	2.737	1.275	0.080	5.493	0.958	0.554	-0.016
60	2.887	1.226	0.197	5.361	0.954	0.547	0.000
72	3.152	1.144	0.430	5.219	0.948	0.530	0.022
120 (Level)	3.456	1.060	0.710	5.133	0.941	0.510	0.036
Slope	1.328	0.877	-2.294	2.538	0.947	0.123	-0.303
Curvature	-0.794	0.553	-2.564	1.218	0.834	0.071	-0.164

FIGURE 5.1: Median Yield for LIBOR USD and DKK along with 25th and 75th percentiles



In Figure 5.1 we have plotted the median yield together with the 25th and 75th percentile for both USD and DKK. We can see the upward-sloping and concave curve (to the left) that we mentioned in Chapter 4, as well as the long rates being less volatile than the short rates mentioned earlier. One can also see that the distribution of yields around the median is slightly asymmetric, with a longer right tail for the LIBOR USD whereas we see a longer left tail for DKK. The tables above in combination with Figure 5.1 displays evidence of a normal yield curves.

5.2 Fitting the DNS model with fixed lambda

As discussed in Chapter 4 we fit the yield curve using the Dynamic Nelson-Siegel model shown in Equation 4.8,

$$y_t(\tau) = \beta_{1t} + \beta_{2t}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) + \beta_{3t}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right).$$

By fixing $\lambda = 0.0598$, as computed in Chapter 4, we are able to compute the two factor loadings and use ordinary least squares to estimate the beta factors β_{1t}, β_{2t} and β_{3t}, for each t. This enhances simplicity and computational trustworthiness. Still, the question of the optimal or appropriate value of λ arises. Therefore, we follow up with another estimation methodology further on. As mentioned in Chapter 4, λ determines the maturity where the loading of β_{3t}, the curvature, and mid-term, factor reaches its maximum. Bear in mind that the maturities are not equally spaced. Thus, we implicitly put more weight on the most active region of the yield curve, that being the region where we have smaller distance, when fitting the model.

There are many ways to go in order to evaluate the fit of the model. For our other models, in subsequent sections, we also present some important aspects through tables and additional figures. However, since the forecasting performance is our primary focus, for fitting DNS yields with fixed λ we only include the two tables below 5.4 and 5.5 that show some descriptive statistics of the yield curve residuals. The last three columns contain residual sample autocorrelation at 1, 12, and 30 months displacements. These indicate that the yield curve residuals are persistent.

In the next section we compare the in-sample performance of this DNS model with fixed lambda and its counterpart with time dependent lambda.

Table 5.4: Descriptive statistics for DKK LIBOR yield curve residuals using $\lambda = 0.0598$. Jan, 2003 to March, 2015

Maturity	Mean	Std	Min	Max	MAE	RMSE	$\hat{\rho}(1)$	$\hat{\rho}(12)$	$\hat{\rho}(30)$
1	0.049	0.066	-0.086	0.341	0.064	0.082	0.920	0.493	0.043
2	0.018	0.039	-0.110	0.152	0.035	0.043	0.892	0.323	-0.016
3	-0.006	0.023	-0.126	0.067	0.017	0.024	0.706	-0.180	-0.009
6	-0.042	0.054	-0.218	0.041	0.048	0.069	0.945	0.579	0.022
9	-0.028	0.048	-0.214	0.099	0.045	0.056	0.927	0.425	-0.012
12	-0.016	0.044	-0.209	0.128	0.040	0.047	0.877	0.215	-0.013
24	0.008	0.029	-0.109	0.114	0.023	0.030	0.746	-0.103	0.049
36	0.012	0.022	-0.040	0.069	0.019	0.025	0.874	0.403	0.002
48	0.014	0.019	-0.031	0.079	0.018	0.024	0.938	0.483	-0.068
60	0.014	0.020	-0.053	0.120	0.020	0.024	0.835	0.162	0.111
72	-0.024	0.028	-0.086	0.048	0.031	0.037	0.791	0.153	-0.019
120	0.002	0.012	-0.058	0.044	0.010	0.012	0.755	-0.074	0.050

Table 5.5: Descriptive statistics for USD LIBOR yield curve residuals using $\lambda = 0.0598$. Jan, 2003 to March, 2015

Maturity	Mean	Std	Min	Max	MAE	RMSE	$\hat{\rho}(1)$	$\hat{\rho}(12)$	$\hat{\rho}(30)$
1	0.011	0.060	-0.123	0.337	0.043	0.061	0.780	-0.170	0.197
2	-0.013	0.030	-0.164	0.068	0.026	0.033	0.569	-0.107	-0.159
3	-0.027	0.031	-0.222	0.031	0.031	0.041	0.664	0.236	-0.023
6	-0.009	0.026	-0.119	0.094	0.021	0.027	0.717	0.062	0.048
9	0.006	0.035	-0.095	0.171	0.029	0.036	0.716	-0.094	-0.057
12	0.016	0.041	-0.088	0.199	0.035	0.044	0.733	-0.127	-0.056
24	0.023	0.043	-0.087	0.153	0.039	0.049	0.820	0.043	-0.053
36	0.014	0.033	-0.055	0.104	0.027	0.036	0.887	0.248	-0.148
48	0.007	0.017	-0.049	0.054	0.014	0.019	0.848	0.034	-0.142
60	0.004	0.020	-0.079	0.067	0.017	0.021	0.754	-0.245	-0.127
72	-0.046	0.053	-0.179	0.052	0.053	0.070	0.891	0.355	-0.069
120	0.013	0.020	-0.035	0.064	0.019	0.024	0.829	0.140	-0.039

5.3 Fitting the DNS model with time dependent lambda

As in the previous section, we fit the yield curves using the three factor Dynamic Nelson-Siegel model. However, now we let λ be estimated as shown in Equation 4.9,

$$y_t(\tau) = \beta_{1t} + \beta_{2t}\left(\frac{1 - e^{-\lambda_t \tau}}{\lambda_t \tau}\right) + \beta_{3t}\left(\frac{1 - e^{-\lambda_t \tau}}{\lambda_t \tau} - e^{-\lambda_t \tau}\right)$$

Now we estimate all the four parameters $\beta_{1t}, \beta_{2t}, \beta_{3t}$ and λ_t using non-linear least squares, for each t. So, instead of fixing $\lambda = 0.0598$, we estimate λ_t for every separate regression, using the fminsearch function in Matlab. Note that for USD LIBOR the λ_t in-sample mean is 0.8876, which is substantially larger than the previously fixed lambda value of 0.0598. In Table 5.7 and Table 5.6 we present statistics describing the yield curve residuals, the in-sample fit of the DNS model, of DKK LIBOR and USD LIBOR respectively.

When comparing these tables to the in-sample performance of the DNS model with fixed lambda, presented in the previous section, we see that the DNS model with time dependent lambda produce RMSE that is slightly lower than that of the model with fixed lambda, for all maturities except for the ten-year yield of DKK. Even though these results show some minor advancement they might not be strong enough to support the claim of improved in-sample fit by letting lambda vary over time.

TABLE 5.6: Descriptive statistics for DKK LIBOR yield curve residuals DNS opt λ, Jan, 2003 to March, 2015

Maturity	Mean	Std	Min	Max	MAE	RMSE	$\hat{\rho}(1)$	$\hat{\rho}(12)$	$\hat{\rho}(30)$
1	0.049	0.054	-0.011	0.221	0.050	0.073	0.937	0.535	0.061
2	0.018	0.027	-0.040	0.073	0.022	0.033	0.925	0.552	-0.063
3	-0.007	0.016	-0.084	0.045	0.011	0.018	0.535	-0.015	-0.053
6	-0.044	0.053	-0.187	0.011	0.045	0.069	0.951	0.617	0.017
9	-0.031	0.039	-0.126	0.019	0.033	0.050	0.953	0.584	-0.018
12	-0.019	0.027	-0.093	0.028	0.023	0.033	0.915	0.500	-0.048
24	0.009	0.015	-0.071	0.121	0.013	0.017	0.783	0.485	0.013
36	0.017	0.025	-0.028	0.147	0.022	0.030	0.927	0.517	-0.069
48	0.020	0.021	-0.014	0.114	0.021	0.029	0.922	0.522	-0.064
60	0.018	0.015	-0.012	0.065	0.018	0.023	0.873	0.375	-0.027
72	-0.024	0.020	-0.085	0.015	0.025	0.031	0.915	0.565	-0.124
120	-0.007	0.015	-0.198	0.022	0.012	0.017	0.641	-0.054	-0.064

Table 5.7: Descriptive statistics for USD LIBOR yield curve residuals DNS opt λ, Jan, 2003 to March, 2015

Maturity	Mean	Std	Min	Max	MAE	RMSE	$\hat{\rho}(1)$	$\hat{\rho}(12)$	$\hat{\rho}(30)$
1	0.034	0.042	-0.031	0.290	0.035	0.054	0.803	0.089	0.129
2	0.002	0.017	-0.103	0.053	0.009	0.017	0.438	0.073	-0.190
3	-0.020	0.027	-0.227	0.017	0.020	0.033	0.659	0.058	0.108
6	-0.018	0.019	-0.110	0.020	0.019	0.026	0.748	0.257	0.046
9	-0.012	0.018	-0.092	0.045	0.014	0.021	0.774	0.199	-0.096
12	-0.005	0.016	-0.080	0.044	0.012	0.017	0.778	0.069	-0.188
24	0.011	0.016	-0.017	0.074	0.014	0.020	0.802	0.091	-0.016
36	0.015	0.019	-0.018	0.078	0.016	0.024	0.846	0.315	-0.100
48	0.013	0.013	-0.011	0.060	0.014	0.019	0.857	0.179	-0.218
60	0.010	0.008	-0.014	0.039	0.011	0.013	0.771	-0.015	-0.270
72	-0.044	0.033	-0.123	-0.000	0.044	0.055	0.958	0.642	-0.060
120	0.012	0.017	-0.053	0.051	0.018	0.021	0.870	0.368	0.085

In Figure 5.3 and Figure 5.2 below we plot the raw yield curve, represented as dots and the DNS fitted yield curve in solid line, for four selected dates. We use the same dates for both DKK and USD for DNS, as well as for DNSS, in order to be able to compare the ability to fit the yields. From these figures it is evident that the three-factor DNS model is capable of replicating some different yield curve shapes: upward-sloping, downward-sloping, humped and inverted humped.

However, we see that the DNS model has a hard time fitting the whole yield curve when the data points are too non-linear as displayed in the top right corner of Table 5.2 at date 09/03/2009. Since the DNS model only has on curvature factor, it cannot handle situations where the yield curve display a double hump, as for this date. Bear this in mind for the coming section where we display the fit for the DNSS model.

Figure 5.2: Selected fitted yield curves. DNS fitted yield curves (DKK) for selected dates, together with actual yields

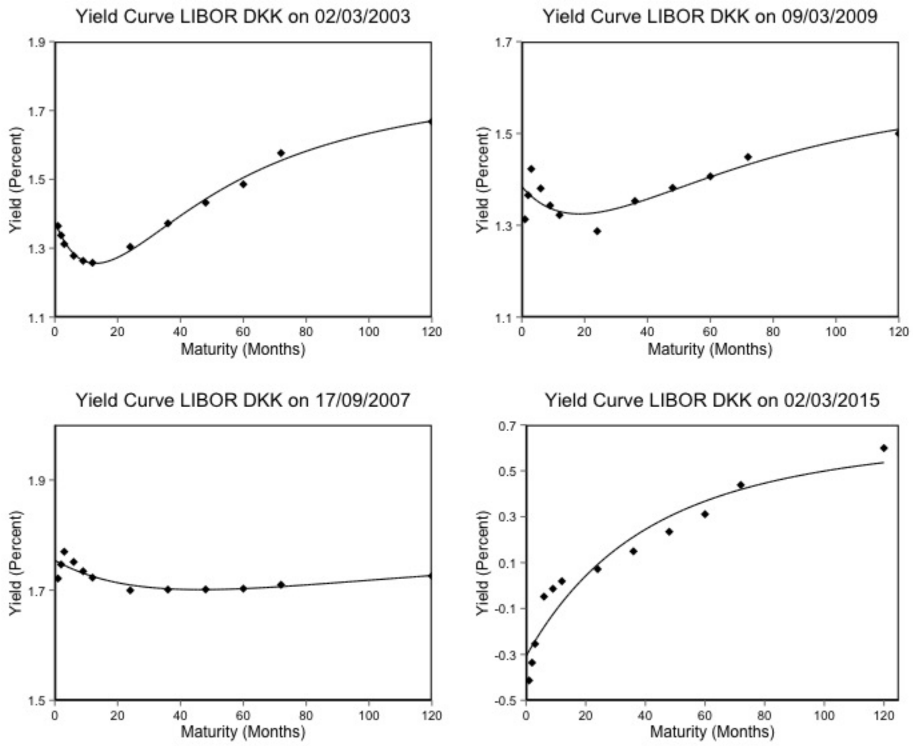
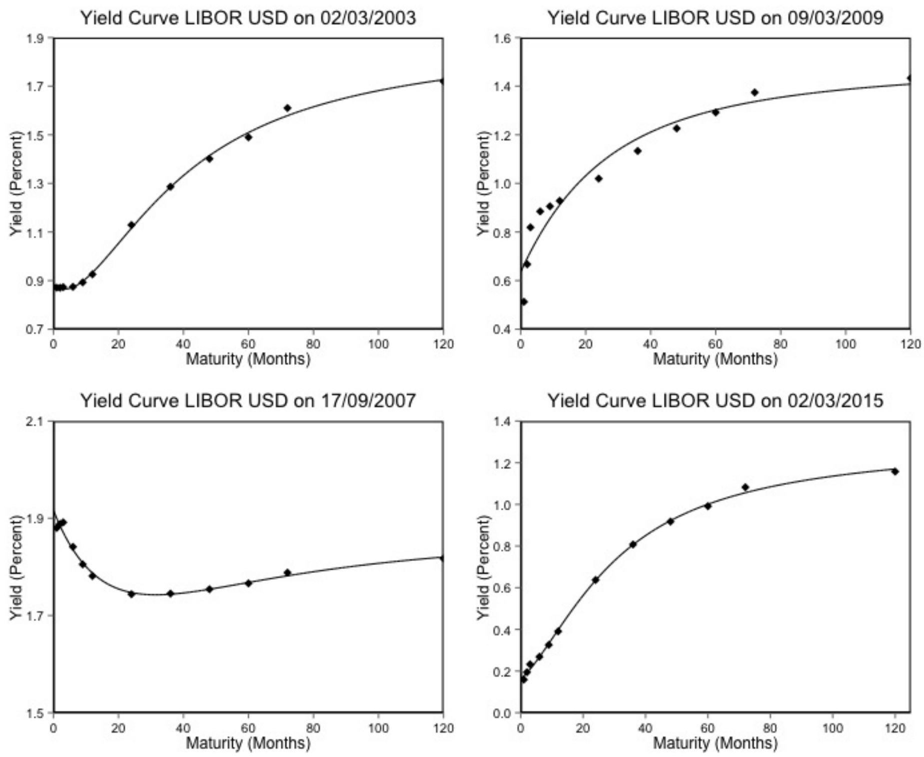


Figure 5.3: Selected fitted yield curves. DNS fitted yield curves (USD) for selected dates, together with actual yields



In the three figures below Figure 5.4, Figure 5.5 and Figure 5.6 we plot the estimated factors, $\beta_{1t}, \beta_{2t}, \beta_{3t}$, representing the level, slope and curvature factors in grey, along with the the empirical level, slope and curvature in black, defined in Chapter 4. These figures validates that the three factors in our DNS model actually correspond to level slope and curvature, which supports the findings from Diebold and Li (2006). As shown numerically in Chapter 4, we now also graphically show in Figure 5.5 that the model-based slope factor, $-\hat{\beta_{2t}}$, coincide with the empirical slope factor, and in 5.6 that the model-based curvature factor, $0.3\hat{\beta_{3t}}$, coincide with the empirical curvature factor.

FIGURE 5.4: Model-based level factor vs. empirical level factor

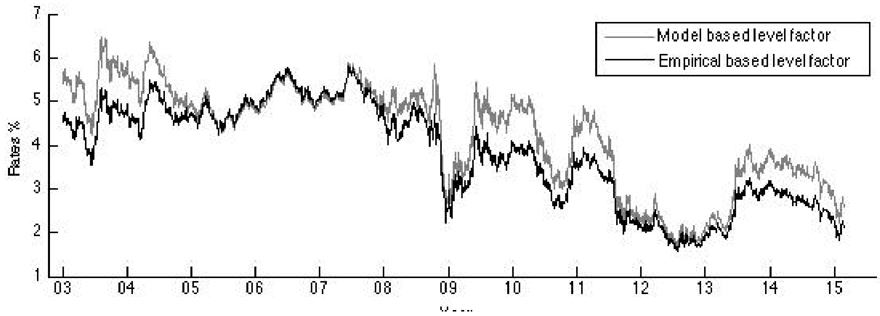


FIGURE 5.5: Model-based slope factor vs. empirical slope factor

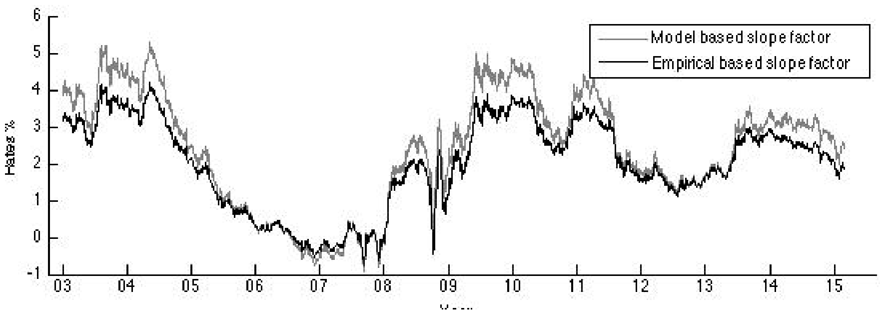
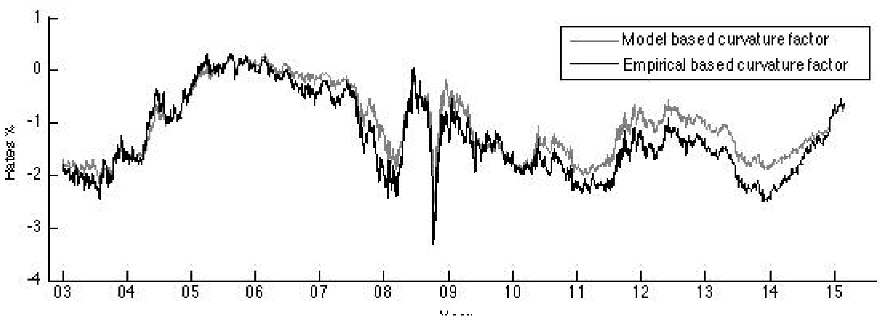


FIGURE 5.6: Model-based curvature factor vs. empirical curvature factor



In Table 5.8 we present descriptive statistics for the estimated factors. The last column (to the right) contains augmented Dickey-Fuller (ADF) unit root test statistics, which suggest that the factors may have unit roots. The three columns to the left of the ADF test statistics show sample autocorrelation at different displacements. The autocorrelations of the factors show that the first factor is the most persistent, and that the second factor is more persistent than the third one.

Table 5.8: Descriptive statistics for estimated factors

Factor	Mean	Std	Min	Max	$\hat{\rho}(1)$	$\hat{\rho}(12)$	$\hat{\rho}(30)$	ADF
$\hat{\beta}_{1t}$	4.245	1.170	1.659	6.470	0.907	0.473	0.111	-1.078
$\hat{\beta}_{2t}$	-2.374	1.533	-5.278	0.947	0.933	0.134	-0.501	-1.179
$\hat{\beta}_{3t}$	-3.443	2.195	-8.421	1.061	0.898	0.153	-0.207	-1.470

5.4 Fitting the Dynamic Nelson-Siegel-Svensson model

We proceed by fitting the yield curve using the Dynamic Nelson-Siegel-Svensson model shown in Equation 4.14,

$$y_t(\tau) = \beta_{1t} + \beta_{2t}\left(\frac{1-e^{-\lambda_1\tau}}{\lambda_1\tau}\right) + \beta_{3t}\left(\frac{1-e^{-\lambda_1\tau}}{\lambda_1\tau} - e^{-\lambda_1\tau}\right) + \beta_{4t}\left(\frac{1-e^{-\lambda_2\tau}}{\lambda_2\tau} - e^{-\lambda_2\tau}\right),$$

In this equation we actually have five parameters, $\beta_{1t}, \beta_{2t}, \beta_{3t}, \lambda_1$ and λ_2 ,to be estimated by non-linear least squares, for each t. Again, λ, and this time, both of them are optimized for each t, using the fminsearch function in Matlab. In Table 5.9 and Table 5.10 we present some descriptive statistics for the yield curve residuals that describe the in-sample fit. Comparing these tables to the DNS counterpart, Table 5.6 and Table 5.6 we see that the DNSS model has an advantage over both of the DNS models when it comes to fitting the yield curves, for all maturities. The DNSS fitted yields follow the observed data points very closely, which is to be seen in the subsequent figures as well.

Table 5.9: Descriptive statistics for DKK LIBOR yield curve residuals, DNSS, Jan, 2003 to March, 2015

Maturity	Mean	Std	Min	Max	MAE	RMSE	$\hat{\rho}(1)$	$\hat{\rho}(12)$	$\hat{\rho}(30)$
1	-0.002	0.008	-0.030	0.065	0.005	0.008	0.432	0.135	-0.065
2	0.003	0.008	-0.031	0.026	0.006	0.009	0.366	0.164	-0.050
3	0.002	0.009	-0.070	0.043	0.006	0.009	0.464	0.199	-0.035
6	-0.008	0.013	-0.061	0.021	0.011	0.015	0.858	0.564	0.015
9	0.003	0.004	-0.011	0.019	0.004	0.005	0.772	0.456	-0.041
12	0.004	0.007	-0.015	0.030	0.006	0.009	0.731	0.443	-0.035
24	-0.005	0.009	-0.051	0.045	0.008	0.010	0.799	0.302	-0.229
36	-0.003	0.006	-0.029	0.027	0.004	0.006	0.704	-0.053	-0.165
48	0.008	0.007	-0.008	0.037	0.008	0.010	0.870	0.461	-0.168
60	0.014	0.009	-0.008	0.050	0.015	0.017	0.877	0.427	-0.042
72	-0.021	0.016	-0.058	0.015	0.022	0.026	0.941	0.567	-0.180
120	0.004	0.006	-0.026	0.022	0.006	0.007	0.678	0.363	-0.230

Table 5.10: Descriptive statistics for USD LIBOR yield curve residuals, DNSS, Jan, 2003 to March, 2015

Maturity	Mean	Std	Min	Max	MAE	RMSE	$\hat{\rho}(1)$	$\hat{\rho}(12)$	$\hat{\rho}(30)$
1	0.003	0.007	-0.020	0.036	0.005	0.008	0.410	0.073	0.003
2	-0.000	0.007	-0.048	0.033	0.004	0.007	-0.086	0.002	-0.004
3	-0.007	0.012	-0.045	0.035	0.009	0.014	0.527	0.125	-0.023
6	0.004	0.009	-0.022	0.049	0.006	0.010	0.557	0.125	0.082
9	0.002	0.005	-0.020	0.020	0.004	0.006	0.589	0.166	-0.168
12	-0.001	0.008	-0.037	0.022	0.005	0.008	0.543	0.077	0.005
24	-0.007	0.009	-0.038	0.032	0.008	0.011	0.774	0.237	-0.237
36	0.002	0.007	-0.020	0.031	0.005	0.007	0.592	0.190	-0.024
48	0.012	0.010	-0.005	0.045	0.012	0.015	0.914	0.517	-0.166
60	0.016	0.011	-0.008	0.044	0.016	0.020	0.894	0.577	-0.122
72	-0.035	0.025	-0.091	0.001	0.035	0.043	0.942	0.605	-0.101
120	0.010	0.009	-0.020	0.035	0.010	0.014	0.802	0.431	-0.011

As for DNS we plot the raw yield curve, represented as dots, and the DNSS fitted yield curve in solid lines for the same selected dates. This is presented in Figure 5.7 and Figure 5.8. It is evident that the four-factor DNSS model is very much capable of replicating some different yield curve shapes: upward-sloping, downward-sloping, humped and inverted humped. Moreover, it does a great job in capturing the non-linearity displayed in the top right corner of Figure 5.8, representing date 09/03/2009.

As we can see, it is able to fit a curve with more than one hump. If we remember the outsmoothed fit by DNS we now instead see a close fit to the raw yield curve in all areas of the yield curve. So, in terms of in-sample fit DNSS performs way better than the DNS for both rates respectively. This is evident also in Table 5.10 and Table 5.9. As mentioned earlier, both of the tables present significantly lower RMSE than the DNS counterpart.

FIGURE 5.7: Selected fitted yield curves. DNSS fitted yield curves (USD) for selected dates, together with actual yields

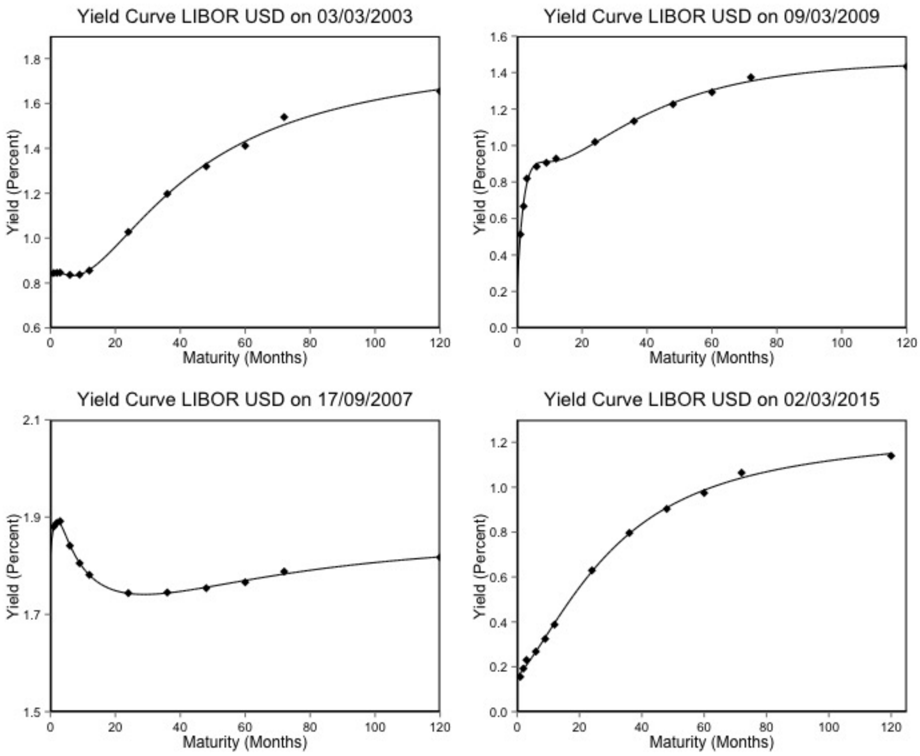
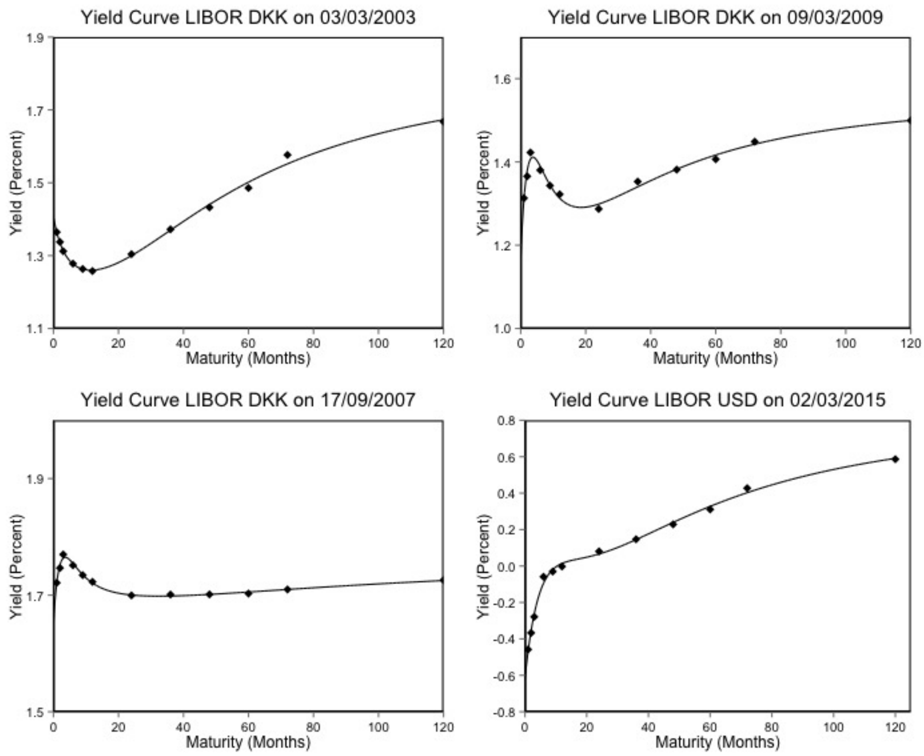


Figure 5.8: Selected fitted yield curves. DNSS fitted yield curves (DKK) for selected dates, together with actual yields



In Figure 5.9, Figure 5.10 and Figure 5.11 below, we compare the estimated level, slope and first curvature factor from the DNSS model (in grey) to the corresponding empirical factors (in black). Surprisingly, none of the three factors change notably, when introducing the second curvature factor. We would expect the level factor to change to some extent, since the second curvature factor can fit yields with maturities in the long end of the yield curve, and thus allowing the level factor to fit other areas of the yield curve. But in this case the level factor still follows the 10-year yield quite closely, correlation coefficient only drops from 0.9531 to 0.9466.

Figure 5.9: Model-based level factor vs. empirical level factor

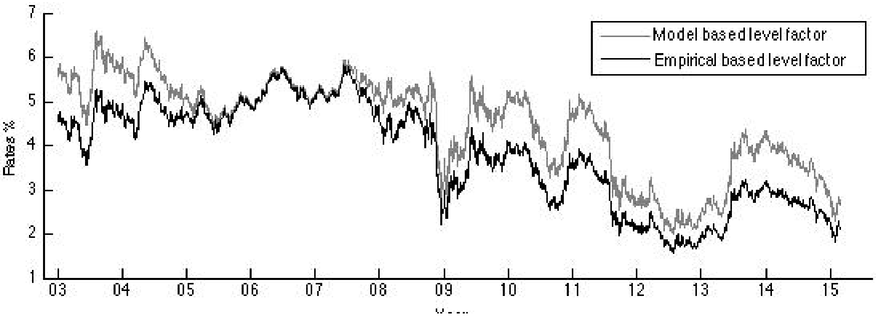


FIGURE 5.10: Model-based slope factor vs. empirical slope factor

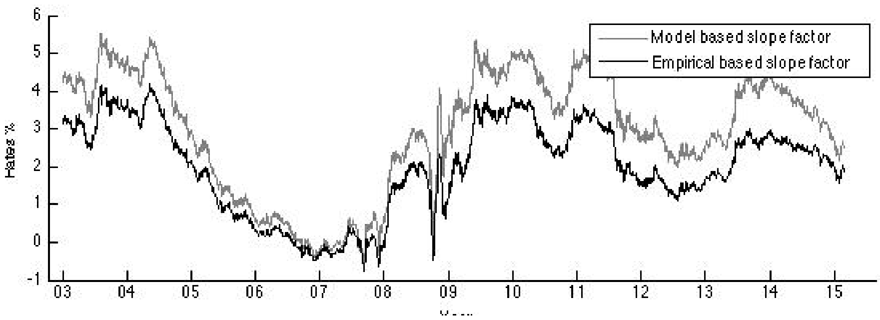


FIGURE 5.11: Model-based curvature factor vs. empirical curvature factor

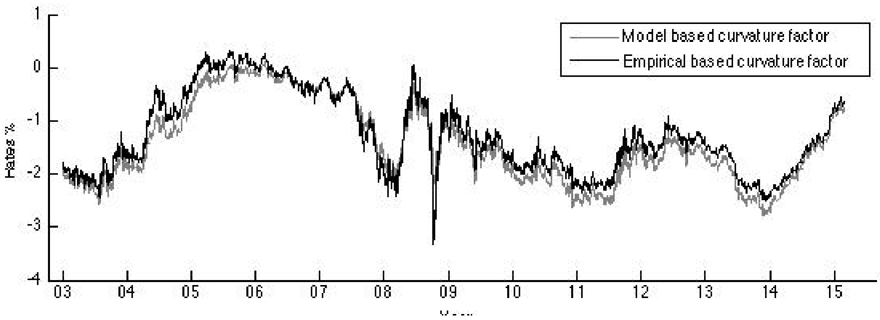
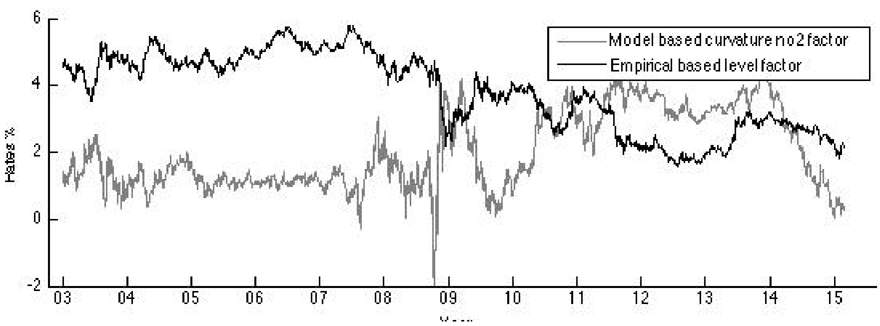


Figure 5.12 shows the estimated path of the second curvature factor from the DNSS model (in grey), along with the 120 month (10-year yield) (in black). As brought up in Chapter 4 and mentioned above, the purpose of the second curvature factor is to improve the fit of long-term yields for the DNSS model. We include the 10-year yield for comparison in order to see how well the second curvature factor does just that, fit the long-term yields. The curvature no2 factor does strongly move with the ten-year yield, however, in the opposite direction, with a correlation coefficient of -0.7021.

FIGURE 5.12: Model-based curvature no2 factor vs. 10-year yield



In Table 5.11 we present descriptive statistics for the estimated factors. The last column (to the right) contains augmented Dickey-Fuller (ADF) unit root test statistics, which suggest that the factors may have unit roots. The three columns to the left of the ADF test statistics show sample autocorrelation at different displacements, 1, 12 and 30 months. The autocorrelations of the factors show that the level factor is still the most most persistent. The first curvature factor is more persistent than the second curvature factor.

Table 5.11: Descriptive statistics for estimated factors

Factor	Mean	Std	Min	Max	$\hat{\rho}(1)$	$\hat{\rho}(12)$	$\hat{\rho}(30)$	ADF
$\hat{\beta}_{1t}$	4.463	1.085	1.983	6.593	0.901	0.428	0.118	-1.072
$\hat{\beta}_{2t}$	-2.928	1.589	-5.518	0.740	0.939	0.196	-0.484	-1.081
$\hat{\beta}_{3t}$	-4.786	2.545	-9.296	0.343	0.928	0.284	-0.178	-1.275
$\hat{\beta}_{4t}$	2.002	1.146	-1.981	4.566	0.836	0.417	-0.085	-1.753

Chapter 6

Forecasting

6.1 Forecasting yield curve level, slope and curvature

Accurate prediction of yields is of crucial matter in this thesis. Both short-term and long-term decision making is often based on forecasts. Thus, it is of utmost importance that the yield curve models perform well both in-sample and out-of-sample. A model performing well in-sample does not necessarily provide good out-of sample results, which may be due to over-parametrization.

In order to focus more in depth on the forecasting performance of our models, hereafter we exclusively work with USD LIBOR. However, the modeling framework is applicable for any other LIBOR currency. Moreover, we leave the DNSS model behind and continue by solely forecasting yields for the DNS model versions with fixed and time dependent lambda. The reason for not including the DNSS model in our forecasting analysis is that to any possible extent avoid or even eliminate the optimization problematics associated with forecasting the DNSS model.

Forecasting the DNS model yield curve is equivalent to forecasting the underlying factors, $\hat{\beta_{1t}}, \hat{\beta_{2t}}, \hat{\beta_{3t}}$, since the yield curve only depends on just those factors. When performing our forecasting exercise we use end of month data, in order to be comparable to forecast exercises made by others as well as limit the computational time for our model with time dependent lambda.

Following Diebold and Li (2006) we estimate recursively, using the data we have from the beginning of our time period starting in January, 2003, to the start of the forecasting period, July, 2010, and extending through March, 2015, which is the end date of our time period. Consequently, the sample period is divided into a training period, representing $\frac{2}{3}$ of the available data, and a test period, representing the last $\frac{1}{3}$ of the sample period.

As the wording reveals, the training period is used to practise the model to discover potentially predictive relationships, while the test period is used to assess the strength and utility of that predictive relationship.

As mentioned above, we only have to forecast the factors in order to get a forecasted DNS yield curve. We forecast the DNS factors as univariate first order autoregressive AR(1) processes as well as we produce yield forecasts based on an underlying multivariate VAR(1) specification. These two forecasting specifications make out our baseline model. For further performance evaluation of the model, we forecast yields based on natural competitors described in the next section. But first we specify how the DNS model forecasts are generated.

6.1.1 DNS model forecasting specifications

The univariate AR(1) factor forecast is specified as:

$$\hat{y}_{t+h/t}(\tau) = \hat{\beta}_{1,t+h/t} + \hat{\beta}_{2,t+h/t}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) + \hat{\beta}_{3,t+h/t}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right), \tag{6.1}$$

where

$$\hat{\beta}_{1,t+h/t} = \hat{c}_i + \hat{\gamma}_i\hat{\beta}_{it}, \qquad i = 1,2,3,$$

and $\hat{c}_i$ and $\hat{\gamma}_i$ are obtained by regressing $\hat{\beta}_{it}$ on an intercept and $\hat{\beta}_{i,t-h}$.

AR-processes are basic time series regressions, where the output variable depends linearly on its own past values. The generalized representation is a AR(p) process, where p determines the number of lagged previous values. For forecasting yield curves the most commonly used AR-process is the AR(1). An AR(1) process is given by:

$$r_t = \phi_0 + \phi_1 r_{t-1} + a_t,$$

where a_t is assumed to be white noise, mean zero and variance σ_a^2. An AR(1) model implies that, conditional on its own previous value r_{t-1}, we have:

$$E(r_t|r_{t-1}) = \phi_0 + \phi_1 r_{t-1}, \qquad Var(r_t|r_{t-1}) = Var(a_t) = \sigma_a^2$$

That is, given its past value r_{t-1}, the current value centers around $\phi_0 + \phi_1 r_{t-1}$ with standard deviation σ_a. This satisfies the Markov property that conditional on r_{t-1}, r_t is uncorrelated with r_{t-i} for $i > 1$ (for an AR(1)).

The multivariate VAR(1) factor forecast specifications are:

$$\hat{y}_{t+h/t}(\tau) = \hat{\beta}_{1,t+h/t} + \hat{\beta}_{2,t+h/t}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) + \hat{\beta}_{3,t+h/t}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right), \tag{6.2}$$

where

$$\hat{\beta}_{1,t+h/t} = \hat{c} + \hat{\Gamma}\hat{\beta}_t. \tag{6.3}$$

A VAR-process is basically an extension of the AR process, by allowing for more than one evolving variable, we now have a vector of AR-processes. Formally, a VAR(1) process is defined as:

$$\boldsymbol{r}_t = \boldsymbol{\phi}_0 + \boldsymbol{\Phi}\boldsymbol{r}_{t-1} + \boldsymbol{a}_t,$$

where $\boldsymbol{\phi}_0$ is a n-dimensional vector, $\boldsymbol{\Phi}$ is a $n \times n$ matrix, and $\boldsymbol{a}_t$ is a sequence of serially uncorrelated random vectors, mean zero and covariance matrix $\boldsymbol{\Sigma}$. When used in application, the covariance matrix is required to be positive definite; otherwise, the dimension of r_t can be reduced. In the literature, it is often assumed that $\boldsymbol{a}_t$ is multivariate normal.

Note, that the VAR process sometimes is too restrictive to properly picture the main characteristics of the data. Therefore, additional deterministic terms (such as linear trends) and/or external variables might be needed to represent the data. In this case the extended VAR process may look like:

$$\boldsymbol{r}_t = \boldsymbol{\phi}_0 + \boldsymbol{\Phi}\boldsymbol{r}_{t-1} + \boldsymbol{\Pi}\boldsymbol{D}_t + \boldsymbol{G}\boldsymbol{Y}_t + \boldsymbol{a}_t,$$

where $\boldsymbol{D}_t$ is an $(l \times 1)$ deterministic matrix, $\boldsymbol{Y}_t$ represents the $(m \times 1)$ matrix of external variables and $\boldsymbol{\Pi}$ and $\boldsymbol{G}$ are parameter matrices.

For both these forecasting specifications the forecast errors at time $t+h$ are defined as

$$y_{t+h/t}(\tau) - \hat{y}_{t+h/t}(\tau). \tag{6.4}$$

6.2 Benchmark competitors

In order to evaluate the forecasting performance of our models we also include the best practice competitors. Below we describe these in terms of how their forecasts are generated:

6.2.1 Random Walk

Random Walk (RW) hereafter, is a process commonly used when modeling unit-root non-stationary times series, e.g. interest rates or exchange rates. The RW process can be described by the following formula:

$$y_t = y_{t-1} + \epsilon_t$$

where y_t and y_{t-1} represent two consecutive values of a process and ϵ_t is a white noise series, symmetrically distributed around zero. Since ϵ_t follows that last criteria, there is a 50/50 probability of y_t to go either up or down in relation to y_{t-1}. Thus, for any forecast horizon, $h > 0$, the RW forecast follows:

$$\hat{y}_{t+h/t}(\tau) = y_t(\tau) \tag{6.5}$$

The above relationship states that the RW forecast is always "no change", meaning that for any forecast horizon, $h > 0$, the predicted yield value is always equal to the value at the forecast origin. Consequently, the RW process fails to pick up the mean reversion of the yields. The literature promotes the RW forecast for short forecasting horizons.

6.2.2 Slope regression

The slope regression forecast is specified according to the following equation:

$$\hat{y}_{t+h/t}(\tau) - y_t(\tau) = \hat{c}(\tau) + \hat{\gamma}(\tau)(y_t(\tau) - y_t(3)) \tag{6.6}$$

This forecast is obtained from a regression of historical yield changes on yield curve slopes. The changes is of the yields is given by, $\hat{y}_{t+h/t}(\tau) - y_t(\tau)$, and we estimate $\hat{c}(\tau)$ and $\hat{\gamma}(\tau)$ by OLS.

6.2.3 AR(1) on yield levels

This simple first order autoregressive forecasting model can be employed to predict the yields, given that the yield time series show a high lag-1 autocorrelation. The process is described as:

$$y_t = c + \gamma y_{t-1} + \epsilon_t$$

where y_t and y_{t-1} are two consecutive yields, and c and γ being constants, and ϵ_t is white noise, with mean zero and variance σ_ϵ^2. In order to predict the yields with a h step ahead forecast, given maturity τ, we regress $\hat{y}_{t+h/t}(\tau)$ on itself, with the h step ahead difference, where each time series regression can be expressed as follows:

$$\hat{y}_{t+h/t}(\tau) = \hat{c}(\tau) + \hat{\gamma}(\tau)(y_t(\tau) \qquad (6.7)$$

6.2.4 VAR(1) on yield levels

In short, the VAR(1) forecast model is like simultaneously running all time series equations in a system, instead of each at a time, as for the AR(1) forecast. Consequently, for a n-dimensional time series the system of equations can be expressed as follows:

$$\boldsymbol{y_t} = \boldsymbol{c} + \boldsymbol{\Gamma} \boldsymbol{y_t} + \boldsymbol{\epsilon_t}$$

where y_t is a multivariate time series, c is a n-dimensional vector of constants, Σ is a $n \times n$ matrix of parameters and ϵ_t is a sequence of k-dimensional, randomly and serially uncorrelated vectors with mean zero and positive definite covariance matrix. The vector of constants ,c, along with the matrix of parameters, Σ respectively are estimated by least squares. The resulting h-step ahead forecast is represented by the following formula:

$$\hat{y}_{t+h/t} = \hat{c} + \hat{\Gamma}\hat{y}_t \qquad (6.8)$$

6.2.5 Regression on 3 AR(1) principal components

As described in Chapter 2, principal component analysis (PCA), is a factor model structure used in order to reduce the dimensionality of a large data set, while retaining as much as possible of the variation present in the data set. From a higher dimensional set of variables, we project the data onto a lower dimensional linear space.

First, we perform principal components analysis on the full set of twelve yields y_t, one for each maturity. This procedure effectively decomposes the yield covariance matrix as $Q\Lambda Q^T$, where the diagonal elements of Λ are the eigenvalues and the columns of Ω are the associated eigenvectors. The three largest eigenvalues are denoted by λ_1, λ_2 and λ_3 respectively, and the three associated eigenvectors by q_1, q_2 and q_3. Then, the first three principal components $x_t = [x_{1t}, x_{2t}, x_{3t}]$ are defined by $x_{it} = q_i', i = 1, 2, 3$. In order to produce h-step ahead forecasts of the three principal components we use a AR(1) model, represented as the following:

$$\hat{x}_{i,t+h/t} = \hat{c}_i + \hat{\gamma}_i x_i, i = 1, 2, 3, \qquad (6.9)$$

and then we produce forecasts for the yields:

$$y_t \equiv [y_t(3), y_t(6), y_t(12), y_t(24), y_t(36), y_t(60), y_t(120)]'$$

by using the forecast principal components and the eigenvectors as:

$$\hat{y}_{t+h/t}(\tau) = q_1(\tau)\hat{x}_{1,t+h/t} + q_2(\tau)\hat{x}_{2,t+h/t} + q_3(\tau)\hat{x}_{3,t+h/t} \tag{6.10}$$

where $q_i(\tau)$ is the element in the eigenvector q_i that corresponds to given maturity τ.

For all these reference models the forecast errors at time $t + h$ are defined as

$$y_{t+h/t}(\tau) - \hat{y}_{t+h/t}(\tau). \tag{6.11}$$

Important to note here is that the forecast object in these cases are future yields, $y_{t+h/t}(\tau)$, and not future Nelson Siegel fitted yields, as for our baseline DNS models.

To examine the forecasting performance we will in the next section present one table for each forecast horizon (1-,6-,12-months) containing our DNS models along with the benchmark competitors. Since we have eight models all together to consider each table span over two pages, where we have four models on each page. In all tables we display some descriptive statistics, including mean, standard deviation, root mean squared error (RMSE) and autocorrelations using various displacements, depending on the forecast horizon.

6.3 Out-of-sample Forecasting Performance

Table 6.1: Out-of-sample 1-month-ahead forecasting for USD LIBOR yield curve

Maturity	Mean	Std	RMSE	$\hat{\rho}(1)$	$\hat{\rho}(6)$
DNS with AR(1) and fixed λ					
3	-0.060	0.061	0.085	0.870	-0.123
6	-0.086	0.054	0.101	0.851	-0.176
12	-0.128	0.054	0.139	0.736	-0.143
24	-0.157	0.080	0.176	0.606	-0.011
36	-0.139	0.101	0.171	0.579	0.122
60	-0.084	0.087	0.120	0.338	0.166
120	-0.093	0.071	0.116	-0.004	0.005
DNS with AR(1) and opt λ					
3	0.019	0.057	0.060	0.337	-0.053
6	0.056	0.070	0.089	0.266	-0.056
12	0.139	0.088	0.165	0.139	0.016
24	0.256	0.160	0.301	0.662	0.152
36	0.306	0.219	0.375	0.828	0.193
60	0.304	0.237	0.384	0.881	0.200
120	0.238	0.189	0.303	0.843	0.202
Random Walk					
3	-0.001	0.021	0.020	0.674	-0.201
6	-0.001	0.023	0.023	0.403	-0.133
12	0.001	0.033	0.033	0.083	-0.018
24	-0.000	0.053	0.052	-0.118	0.327
36	-0.002	0.071	0.070	-0.070	0.413
60	-0.005	0.078	0.078	0.015	0.301
120	-0.009	0.072	0.071	-0.003	0.014
Slope regression					
3	0.010	0.020	0.023	0.687	-0.198
6	-0.000	0.024	0.024	0.406	-0.127
12	0.003	0.034	0.034	0.062	-0.002
24	0.005	0.053	0.053	-0.119	0.330
36	0.003	0.071	0.070	-0.070	0.412
60	-0.001	0.078	0.077	0.014	0.301
120	-0.008	0.072	0.071	-0.002	0.014

Maturity	Mean	Std	RMSE	$\hat{\rho}(1)$	$\hat{\rho}(6)$
DNS with	VAR(1)	factor	dynamics		
3	0.161	0.062	0.172	0.859	-0.118
6	0.121	0.070	0.139	0.831	-0.044
12	0.053	0.079	0.094	0.750	-0.070
24	-0.021	0.068	0.070	0.271	-0.022
36	-0.041	0.075	0.085	0.112	0.296
60	-0.039	0.079	0.087	0.107	0.252
120	-0.080	0.067	0.104	-0.019	-0.002
AR(1) on yield levels					
3	0.003	0.021	0.021	0.675	-0.200
6	0.005	0.023	0.023	0.403	-0.133
12	0.006	0.033	0.033	0.083	-0.018
24	0.003	0.053	0.052	-0.119	0.327
36	-0.005	0.070	0.069	-0.074	0.414
60	-0.022	0.077	0.079	0.004	0.307
120	-0.041	0.069	0.079	-0.005	0.006
VAR(1) on yield levels					
3	-0.179	0.333	0.374	0.173	-0.367
6	-0.221	0.359	0.418	0.183	-0.374
12	-0.243	0.366	0.435	0.207	-0.383
24	-0.235	0.296	0.375	0.182	-0.333
36	-0.220	0.230	0.316	0.070	-0.258
60	-0.206	0.152	0.255	-0.160	-0.194
120	-0.176	0.113	0.208	-0.184	-0.246
Regression on 3 AR(1) principal components					
3	0.183	0.019	0.184	0.649	-0.335
6	0.212	0.030	0.214	0.544	-0.298
12	0.081	0.039	0.090	0.295	-0.027
24	-0.240	0.060	0.247	0.111	0.108
36	-0.285	0.079	0.296	0.193	0.271
60	-0.058	0.083	0.101	0.287	0.163
120	0.333	0.078	0.341	0.388	-0.196

Table 6.2: Out-of-sample 6-month-ahead forecasting for USD LIBOR yield curve

Maturity	Mean	Std	RMSE	$\hat{\rho}(6)$	$\hat{\rho}(18)$
DNS with AR(1) and fixed λ					
3	-0.461	0.224	0.511	0.227	-0.464
6	-0.483	0.219	0.529	0.229	-0.432
12	-0.515	0.206	0.554	0.240	-0.397
24	-0.529	0.190	0.561	0.322	-0.369
36	-0.495	0.192	0.530	0.435	-0.345
60	-0.401	0.171	0.435	0.458	-0.337
120	-0.317	0.121	0.338	0.288	-0.249
DNS with AR(1) and opt λ					
3	0.121	0.179	0.214	0.135	0.062
6	0.172	0.173	0.243	0.103	0.038
12	0.269	0.146	0.306	0.013	0.028
24	0.387	0.137	0.410	0.162	-0.210
36	0.426	0.185	0.464	0.370	-0.297
60	0.398	0.210	0.449	0.442	-0.322
120	0.297	0.179	0.346	0.412	-0.266
Random walk					
3	-0.004	0.083	0.082	0.011	-0.125
6	-0.007	0.076	0.075	0.005	-0.087
12	-0.004	0.076	0.075	-0.082	-0.055
24	-0.000	0.102	0.101	-0.043	-0.246
36	-0.004	0.156	0.154	0.037	-0.339
60	-0.021	0.200	0.199	0.105	-0.415
120	-0.043	0.189	0.192	0.146	-0.436
Slope regression					
3	0.089	0.067	0.111	0.300	-0.255
6	0.026	0.077	0.080	0.001	-0.086
12	0.018	0.082	0.083	-0.090	-0.043
24	0.032	0.104	0.108	-0.057	-0.251
36	0.032	0.155	0.156	0.042	-0.338
60	0.015	0.197	0.195	0.111	-0.412
120	-0.016	0.188	0.186	0.147	-0.435

Maturity	Mean	Std	RMSE	$\hat{\rho}(6)$	$\hat{\rho}(18)$
DNS with VAR(1) factor dynamics					
3	0.603	0.220	0.641	0.157	-0.441
6	0.509	0.200	0.546	0.153	-0.451
12	0.343	0.168	0.381	0.061	-0.430
24	0.097	0.127	0.159	-0.220	-0.295
36	-0.058	0.145	0.154	0.076	-0.295
60	-0.221	0.163	0.273	0.307	-0.378
120	-0.220	0.136	0.258	0.238	-0.373
AR(1) on yield levels					
3	-0.028	0.081	0.084	0.029	-0.123
6	-0.032	0.074	0.080	0.021	-0.084
12	-0.022	0.074	0.076	-0.080	-0.053
24	0.003	0.100	0.099	-0.042	-0.245
36	0.006	0.154	0.152	0.042	-0.337
60	-0.044	0.192	0.194	0.116	-0.409
120	-0.135	0.162	0.210	0.169	-0.419
VAR(1) on yield levels					
3	-0.179	0.333	0.374	0.173	-0.367
6	-0.221	0.359	0.418	0.183	-0.374
12	-0.243	0.366	0.435	0.207	-0.383
24	-0.235	0.296	0.375	0.182	-0.333
36	-0.220	0.230	0.316	0.070	-0.258
60	-0.206	0.152	0.255	-0.160	-0.194
120	-0.176	0.113	0.208	-0.184	-0.246
Regression on 3 AR(1) principal components					
3	0.398	0.057	0.402	-0.113	0.060
6	0.280	0.055	0.286	-0.145	0.140
12	-0.026	0.060	0.065	-0.222	0.093
24	-0.459	0.097	0.469	0.159	-0.241
36	-0.491	0.147	0.512	0.399	-0.340
60	-0.288	0.168	0.332	0.429	-0.396
120	0.028	0.134	0.135	0.303	-0.371

Table 6.3: Out-of-sample 12-month-ahead forecasting for USD LIBOR yield curve

Maturity	Mean	Std	RMSE	$\hat{\rho}(12)$	$\hat{\rho}(24)$
DNS with AR(1) and fixed λ					
3	-0.933	0.269	0.970	-0.133	-0.047
6	-0.952	0.259	0.985	-0.116	-0.049
12	-0.969	0.245	0.998	-0.108	-0.028
24	-0.939	0.218	0.963	-0.111	0.045
36	-0.866	0.203	0.889	-0.026	-0.004
60	-0.714	0.180	0.735	0.011	-0.082
120	-0.496	0.138	0.514	-0.279	0.056
DNS with AR(1) and opt λ					
3	0.368	0.300	0.473	0.080	-0.091
6	0.421	0.278	0.503	0.087	-0.079
12	0.516	0.242	0.569	0.075	-0.025
24	0.613	0.195	0.643	0.052	0.068
36	0.621	0.193	0.649	0.113	-0.085
60	0.536	0.182	0.565	0.078	-0.274
120	0.370	0.143	0.396	-0.099	-0.213
Random walk					
3	-0.036	0.104	0.109	-0.368	0.031
6	-0.044	0.093	0.101	-0.322	-0.004
12	-0.040	0.090	0.097	-0.231	-0.053
24	-0.004	0.132	0.130	-0.029	-0.306
36	0.021	0.211	0.209	-0.093	-0.304
60	0.013	0.290	0.285	-0.304	-0.132
120	-0.021	0.277	0.274	-0.456	0.022
Slope regression					
3	0.057	0.092	0.107	-0.334	-0.001
6	-0.041	0.138	0.142	-0.394	0.042
12	-0.032	0.133	0.135	-0.274	-0.072
24	0.042	0.157	0.160	-0.009	-0.302
36	0.068	0.229	0.235	-0.101	-0.281
60	0.050	0.298	0.298	-0.307	-0.124
120	0.008	0.280	0.276	-0.456	0.023

Maturity	Mean	Std	RMSE	$\hat{\rho}(12)$	$\hat{\rho}(24)$
DNS with	VAR(1)	factor	dynamics		
3	1.135	0.488	1.232	-0.229	0.039
6	1.005	0.462	1.103	-0.242	0.042
12	0.779	0.407	0.876	-0.253	0.041
24	0.442	0.333	0.550	-0.233	-0.038
36	0.212	0.315	0.375	-0.243	-0.121
60	-0.071	0.277	0.282	-0.320	-0.122
120	-0.130	0.204	0.239	-0.436	-0.020
AR(1) on yield levels					
3	-0.257	0.087	0.271	-0.310	-0.037
6	-0.262	0.078	0.273	-0.264	-0.061
12	-0.235	0.076	0.246	-0.203	-0.056
24	-0.125	0.119	0.171	-0.038	-0.303
36	-0.045	0.197	0.199	-0.090	-0.319
60	-0.033	0.270	0.268	-0.298	-0.146
120	-0.126	0.232	0.261	-0.460	0.012
VAR(1) on yield levels					
3	-0.106	0.330	0.341	-0.101	-0.053
6	-0.159	0.349	0.379	-0.129	-0.050
12	-0.189	0.357	0.399	-0.175	-0.043
24	-0.192	0.301	0.353	-0.196	-0.026
36	-0.203	0.254	0.322	-0.146	-0.068
60	-0.186	0.219	0.284	-0.171	-0.097
120	-0.092	0.199	0.216	-0.351	0.042
Regression on 3 AR(1) principal components					
3	0.205	0.048	0.210	-0.163	-0.169
6	0.064	0.045	0.078	-0.113	-0.248
12	-0.235	0.045	0.239	-0.051	-0.129
24	-0.600	0.100	0.608	-0.081	-0.217
36	-0.587	0.156	0.607	-0.067	-0.289
60	-0.381	0.183	0.421	-0.141	-0.236
120	-0.082	0.141	0.161	-0.334	-0.047

6.4 Discussion of Forecasting Results

From an overall perspective we can conclude that the benchmark forecasting models included in this analysis outperform our versions of the Dynamic Nelson–Siegel (DNS) models, including DNS AR(1) with fixed lambda (DNS fixed), DNS AR(1) with time dependent lambda (DNS opt), and DNS VAR(1). For all forecast horizons and maturities the DNS models falls behind its competitors at the RMSE measurement. Whats more, the difference in performance is quite substantial. We will get back to possible reasons to this outcome. During the discussion below we see the DNS models as one group of forecasting models fighting a battle against their main competitors (Random Walk, slope regression, AR(1) on yields, VAR(1) on yields and PCA). Therefore, we compare the results across and within groups.

For the shortest forecast horizon of 1-month presented in Table 6.1, Random Walk (RW) comes across as the clear favourite amongst the different models, producing the lowest RMSE for both short and long maturities. Notably, slope regression provides an almost identical results as RW, just slightly behind in performance. When it comes to the DNS models our DNS opt model performs better than both DNS fixed and DNS VAR(1) for the short maturities. But in the long end of the curve DNS VAR(1) produce better results than the two other DNS models, where also DNS fixed passes DNS opt in performance. PCA provides very unstable results, but nevertheless beats the DNS family for the 12 months maturity.

Moving on to forecast horizon of 6-months presented in Table 6.2 it is possible to make out a similar pattern as the one just described. RW together with AR(1) on yields (AR(1)) hereafter, is still in favor of the DNS group. PCA is still a wobbly competitor, but interestingly enough beats the others for not only 12 month maturity but also for the 120 month maturity. Comparing our baseline models, once again, DNS opt performs better than the other two for the shorter maturities, but approaching the long end of the curve DNS VAR(1) outperforms both DNS fixed and DNS opt. DNS fixed falls behind DNS opt throughout all maturities on the 6 month-ahead forecast, except the very long end.

In Table 6.3 we present the forecast results for 12-month-ahead horizon. The previous picture has not changed very much. We still see that RW and AR(1) have an advantage for all maturities except for 6 and 120 months, where PCA again manage to place itself on top of the others. The DNS opt keeps the trend and produce stable results and outweigh the other two DNS models for shorter maturities. From providing bad results for the short end of the curve, DNS VAR(1) picks up and outperform DNS opt after 24 months maturity and longer. Now, the DNS opt model, with time dependent lambda performs better than the DNS AR(1) model with fixed lambda for for all maturities. It is interesting that now for 12 month ahead forecast PCA really catches up and passes the DNS group. It provides strong results for all maturities, especially for the short and the long end of the curve.

When we examine the two DNS AR(1) models[1] head-to-head it is possible to make out a clear trend, where our DNS opt model slowly but surely outperforms the DNS fixed model as the forecast horizon lengthens. Initially, for the shortest forecast horizon of 1 month-ahead DNS opt only produce lower RMSE than the DNS fixed for the two shortest maturities. As we stretch the forecast horizon to 6 month-ahead DNS opt advance to be the winner for five maturities, leaving the two longest maturities for DNS fixed to take home. It is not until we reach the 12 month-ahead forecast that the DNS opt model really proves its strength and produce a better out-of-sample result than DNS fixed for all maturities.

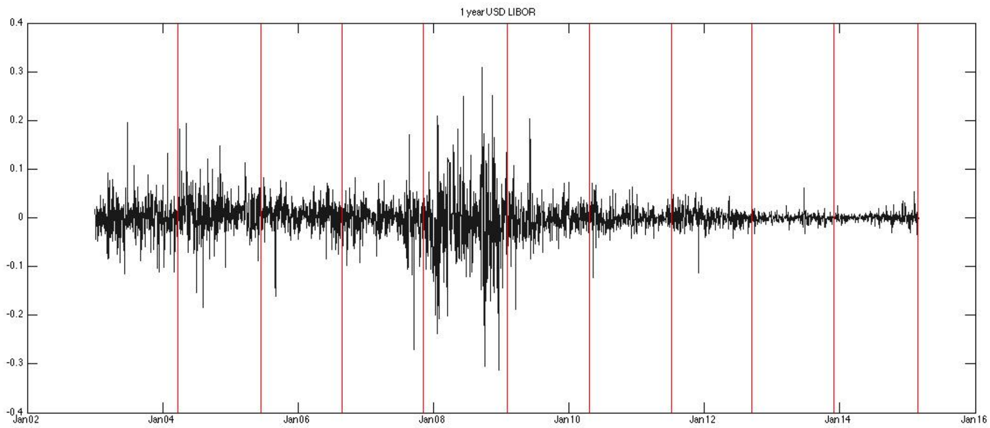
Consequently, we see that in 14 out of 21 possible cases included in this analysis the DNS opt model performs better than the DNS fixed counterpart. These results provide some evidence that in contrast to what Diebold and Li (2006) suggest, it is possible to not only improve the in-sample fit, but also the out-of-sample performance by letting the decay parameter, lambda, vary over time as together with the other factors $(\beta_{1t}, \beta_{2t}, \beta_{3t})$ along with using a more sophisticated non-linear estimation technique. However, it is important to bear in mind that this is only true for the specific time period included in our sample, and that the results might not be convincing enough to justify the complexity of including a more advanced non-linear estimation technique instead of using the standard OLS estimation, which is both faster and less computational intensive to work with.

Furthermore, our results indicate that there is no advantage in using the more advanced Dynamic Nelson-Siegel model over a simple AR(1) or random walk model, even for longer horizons, where we initially would think that the DNS model could have some advantage. So why are the random walk model superior to the DNS model? We suspect that it is a combination of our sampling technique and the time series. To illustrate this we take the first difference of the 1 year USD LIBOR and as can be seen in figure 6.1 it looks like an i.i.d sequence in each of the 10% sampling/calibration windows. This means that it is equally likely that the rate will move up and down after the sample period and hence the latest value will correspond to the mean of the realized value. When the time series have this property the random walk model will perform well against a mean

[1]DNS opt = DNS AR(1) with time dependent lambda,
DNS fixed = DNS AR(1) with fixed lambda

model since it is always re-anchored at the latest value whilst the mean model uses the full time series to estimate the mean.

Figure 6.1: USD LIBOR 1 year



Adding to this, the more standard factor analysis, i.e. PCA, also performs better than the DNS models for longer horizons. In total, our results suggest that one might be better off using one of the natural competitors for the purpose of forecasting LIBOR yields. In the next section we will further investigate the performance of the DNS model through backtesting forecasting distributions at various horizons and initialization points.

6.4.1 Possible reasons for poor forecasting performance of DNS models

Below we have tried to find an explanation for the poor relative performance we see in the forecasting results for the DNS models.

The first reason we can suspect is the we simply have a bad model for this purpose and studying the literature we do see that other authors come up with the same conclusion i.e. that the DNS models are not good when it comes to forecasting the yield curve. The second reason for the bad forecasting results could be that we have made a bad model construction. It is of course possible that we have made a manual error when coding up the models in Matlab. We have however used the same code as we used in the in-sample tests, where it produced really good results but it is of course still possible that the forecasting code contains manual errors.

Furthermore another reason for the results could be the choice of calibration period. Our sample consists of time periods where the yields has been at extreme levels e.g. during the 2008 financial crisis and it could therefore be possible that the forecasting model cannot cope with this specific sample data. We have tried to minimize this by using many different initialization points and a long sample size and we do not think this is a valid conclusion of the bad performance (and in any way the model should be able to pick up rates at all values to be considered a good model). Finally a reason could simply be that the LIBOR rates are frankly just unpredictable using this model, and consequently that other models should be considered.

Chapter 7

Backtesting

7.1 Backtesting for the CCR Measurement

As mentioned earlier, our focus within the Counterparty Credit Risk measure drills down to Risk Factor Evolution models, which is a subset of models generating the Expected Positive Exposure profiles. In this case the Risk Factor Evolution model is our interest rate model, namely the Dynamic Nelson-Siegel model. In this section we focus on assessing the performance of the forecasting model, through backtesting. In a backtest, forecasts at some confidence level are compared with what actually happened in the market. Backtests are used to test whether the forecasting model and its setup is appropriate to use or not. Within the Basel regulatory capital framework backtesting is defined as the quantitative comparison of the IMM model's forecasts against realized values.

In the previous chapter we used a number of different versions of the Dynamic Nelson-Siegel model to forecast yields for a whole range of maturities, making up the term structure of LIBOR rates. In this section we solely focus on forecasts generated by the DNS model using AR(1) factor dynamics, and we use the setup where we fix the decay parameter lambda, $\lambda = 0.0598$, in order to keep the model as simple as possible, and decrease computational complexity. Furthermore, we have chosen to limit the backtesting section to test for four tenors, that is evenly spread, representing vital part of the yield curve.

In order to test how good our model is, it is best to focus on reliability: i.e. the aim of our backtest analysis is to test that forecasts have correct confidence levels (i.e. the reliability). For some historic date there will be one rate for a particular tenor (multiple tenors in what we call the term structure) quoted in the market. In our case we have 12 different tenors, leading to 12 actual rates per historic date. The question then becomes whether the frequency of exceptions, i.e exceeds, is consistent with the frequency of expected exceptions. We will define success criteria of the model later on in this section. We will not study how close the forecasts levels are to the true outcome (the tightness of prediction intervals).

In order to get a better understanding of how backtesting for CCR measurement usually is performed we start by formulating some general backtest terminology and basic steps included in the backtesting methodology.

7.1.1 Key Backtest Terminology

In our backtesting analysis we are forecasting the whole distribution of the yields. All forecasts are initialised at a particular point in time, the **initialization date**. From that date on we are looking into the future.

How far into the future are we looking? The time between initialization and the realization of the forecast, is called the **time horizon**. If the initialization date is on 1st of January and the realized date of the forecast is on 31st of January we have a 30 day time horizon. Forecasts with different time horizons can of course have the same initialization date, i.e. two week and one month forecasts that is realized on 15th and 31st of January respectively would both have been initialized on the same date, 1st of January. The time between two initialization dates is referred to as the **time step**.

The meaning and importance of the backtesting result is highly dependent on the amount and quality of the data used in the sample. The more data and the better data we have access to, the more trustworthy results do we get. Not necessarily better, but more significant. A **backtesting data set** consists of a set of forecasts and the corresponding realizations of those forecasts, i.e. the true values related to the forecasts. This data set forms the statistical sample and can be constructed in a number of ways. A backtesting data set might consist of e.g. a) exposure forecasts and the corresponding realizations of exposure, or as in our case b) the forecasts of a risk factor and the corresponding realizations of that risk factor. In order to increase the amount of data one can use a number of risk factors to aggregate over, or aggregate the data across a number of dimensions, e.g. time, trades, or risk factors and counterparties.

The time period from the start date, i.e the first initialization date and the end date, i.e. the last realization date of the backtest, is referred to as the **observation window**. Backtests using very short observation window may not produce meaningful results, which mean that we might not be able say anything about the quality of the model we are assessing. Furhtermore, using a short observation window, would mean that the results are highly dependent on the specific initialization dates on which the data are collected. In those cases one have to make sure that good initialization dates do not mask bad behavior elsewhere during the period, or vice versa.

There are a variety of methods for generating a backtesting data set over a given observation window. Usually we are talking about non-overlapping (independent) or overlapping (non-independent) data sets. An example of a non-overlapping data set is if we have a time step of one week and a time horizon of the forecast of one week as well. Then we are simultaneously moving the initialization date and the realization date one week ahead, resulting in a non-overlapping data set. For this methodology it is important to note that as the time horizon increases the observation window must also increase in order to maintain the same number of data points and produce statistically significant results. Since non-overlapping windows generate data that can be considered independent, it gives the advantage that one can use standard statistical tests to determine the performance.

It is common to use exceptions as the basis for the backtesting assessment. When the realized quantity of the risk factor exceeds a predetermined risk measure generated by the model, we say that an exception has occurred. The gathered exceptions are then used as the basis for assessing model performance. It is basically a pass or fail assessment. Alternatively, backtesting can be carried out through determining the probability of observing an exposure that is greater than the realized exposure.

When using exposure profiles generated from simulations of market risk factors one have to recognize that these exposure profiles are dependent on the definition and calibration of the stochastic processes driving the underlying risk factor dynamics. It is recommended by the Basel Committee to backtest short and long time horizon both on exposure profiles and on the risk factor model output. The predicted risk factor distributions are then compared to the realized risk factor values at different time horizons. This way we are able to assess whether or not the assumptions of the modelled risk factor dynamics remain valid. In order to challenge the assumptions of the model it is important to include time horizons that are typical margin periods of risk, also known as the liquidation period. This is defined as the time period from the most recent exchange of collateral covering a netting set of financial instruments with a defaulting counterparty until the financial instruments are closed out and the resulting market risk is re-hedged (BIS (2010)).

7.1.2 Backtesting Methodology

In a more detailed way Ruiz (2012) suggest that CCR backtesting should be done by testing the realized path of our risk factor, i.e. the time series of yields. This realized path is given by a collection of yield rates for different time points, which get the value x_{ti}. We choose a point in time within the time series where the backtest starts, t_start, and a corresponding end time, t_end. The resulting backtest time window T is then $T = t_{end} - t_{start}$. We define the time horizon over which we want to test our model as Δ, and proceed as follows:

1. The first time point of measurement is $t_1 = t_{start}$. At t_1, we calculate the risk factor distribution at a point $t_1 + \Delta$ subject to the realization of x_{t1}. We then take the realized value $x_{t1+\Delta}$ of the time series at $t_1 + \Delta$ and observe where that value falls in the risk factor cumulative distribution calculated previously. This results in a value F_1, where $F_i \in (0, 1) \forall i$.
2. We then jump to $t_2 = t_1 + \delta$, and calculate the risk factor distribution at $t_2 + \Delta$ subject to the realization of x_{t2}, and proceed as above: we take $x_{t2+\Delta}$ in the model distribution and from that obtain F_2.
3. Repeat step 1. and 2. continuously until $t_i + \Delta = t_{end}$

This procedure results in a collection of $\{F_i\}_{i=1}^N$, where N is the number of steps taken.

The desired outcome from this procedure is that the empirical distribution from the time series is the same as the predicted distribution from the model. In that case we have a "perfect" model, and then $\{F_i\}_{i=1}^N$ is uniformly distributed. Now, we can define a metric of the difference between the empirical and model distribution, where D represent a distance in the set. If $D = 0$, meaning that we have the same distribution empirically as the one obtain from the model, i.e. a "perfect" model.

There are various typical metrics for D, of which we mention three common ones below. Let F denote the theoretical cumulative distribution function given by the model and F_e denote the empirical cumulative distribution function obtained from the outcome of the previous exercise, $\{F_i\}_{i=1}^N$. We can then use one of the metrics below:

Anderson–Darling metric:

$$D_{AD} = \int_{\infty}^{-\infty} (F_e(x) - F(x))^2 w(F(x)) dF, \tag{7.1}$$

where $w(F) = \frac{1}{F(1-F)}$.

Cramer–von Mises metric:

$$D_{CM} = \int_{\infty}^{-\infty} (F_e(x) - F(x))^2 w(F(x)) dF \tag{7.2}$$

where $w(F) = 1$.

Kolmogorov–Smirnov metric:

$$D_{KS} = \underset{x}{sup} |F_e(x) - F(x)| \tag{7.3}$$

The three metrics described above will generate a different measurement of the distance D. What metric to choose depends on the application of the model assessed. Anderson–Darling is suitable for risk management, where the quality of the models in the tails of the distribution is of interest. Cramer–von Mises is useful in capital calculations, since we then care about the whole distribution function. Kolmogorov–Smirnov is appropriate in situations where small general deviations can be accepted, but large deviations would be unacceptable. In the CCR framework, the goal is to produce a forecast distribution, that is accurate in the whole distribution. So there is no particular interest in either the tail or the middle of the distribution. Therefore, it is suggested to use the Cramer–von Mises metric for CCR backtesting.

After choosing a suitable metric for the model, the next step is to compute $\tilde{D}$, which is the measure of the quality of the model. The question is now; how small does $\tilde{D}$ need to be to reflect a good model, or equivalently, how large must $\tilde{D}$ be to be considered a bad model? Also, since N is a finite number, $\tilde{D}$ will never be exactly zero although we have a perfect model, so how can the validity of $\tilde{D}$ be assessed?

To cope with these questions Ruiz (2012) constructs an artificial time series using the examined model, and applying the earlier mentioned procedure, which yields a value D. Despite the fact that the artificial time series, by definition, follows the model perfectly, D will still not be exactly zero. By repeating this exercise M number of times, constructing M number of artificial time series, where each of them correspond to a perfect model, resulting in a collection of $\{D_k\}_{k=1}^{M}$, that will follow a probability distribution $\psi(D)$. Finally, by making M sufficiently large $\{D_k\}_{k=1}^{M}$ will approximate $\psi(D)$, which allows us to assess the validity of $\tilde{D}$. If $\tilde{D}$ is in a range with high probability in $\psi(D)$, this corresponds to a high probability of an accurate model.

7.1.3 Backtest success criteria

In order to easily score any model, Ruiz (2012) define three colored bands, where the color of the band determines the performance of the model. When letting D_y and D_r be defined as the 95th and 99.99th percentile respectively, the resulting bands are defined as follows:

- Green band if $\tilde{D} \in [0, D_y)$
- Yellow band if $\tilde{D} \in [D_y, D_r)$
- Red band if $\tilde{D} \in [D_r, \infty)$

In words this means that a model in the green band has a 95% probability of being correct, the yellow band means that probability is 4.99% and a model in the red band means that that probability that the model is correct is only 0.01%. This exercise is a sophisticated way of performing reliability and precision tests.

The setup of colored bands comes from the Basel Committees Traffic Light Coverage Test. In the 1996 Amendment the Basel Committee imposed a capital charge on banks for market risk, and therefore specified a methodology for backtesting proprietary VaR measures. The Green light represented a zone where the VaR measures were fine and raised no particular concern to the Committee. On the other hand, measures falling in the yellow zone required monitoring. VaR measures falling in the red zone were presumed flawed and had to be improved. Essentially, the traffic light represents the performance of the VaR measures. For CCR backtesting the traffic light symbolizes the goodness of the model, in terms of probability of being correct, as described above.

However, for simplicity reasons we take a slightly different approach. Reliability tests can also be summarized as a counting exercise, where the the observed number of exceptions are compared with the theoretical level for the percentiles and horizons considered. This is going to be the main focus in our backtesting analysis. We get confidence intervals reaching from 90% to 99% for each forecast horizon and tenors considered. Then in order to evaluate the model we test whether the frequency of exceptions, i.e. exceeds is consistent with the frequency of expected exceptions for different quantiles. Precision tests aim to reveal the sizes and distributions of the exceptions. The results of the corresponding reliability test should be related to these distributions.

7.1.4 Backtesting Setup:

- Time period: start date: 03/03/2008, end date: 03/03/2015

- Time step: 1 week/2 weeks/1 month
- Time Horizons: 1 week, 2 weeks, 1 months, 3 months, 6 months
- Re-calibration frequency: monthly
- Sampling used: Dependent / Independent
- Tenors included in backtest: 3 months, 1 years, 5 years, 10 years
- Currencies in backtest: USD
- Data source: Nordea, Group Counterparty Credit Risk

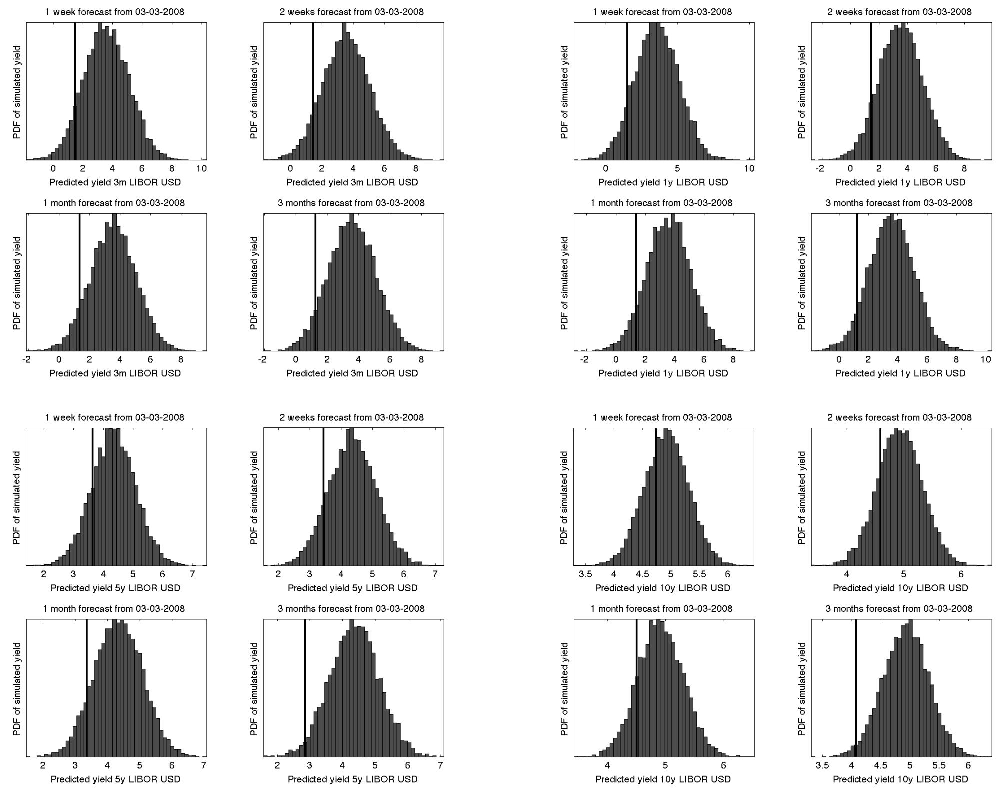
In our backtesting analysis we use the above stated setup. The historic start date is 3rd March, 2008 and the end date is 3rd March, 2015. For our shorter time horizons of 1 week and 2 weeks, we use a time step of 1 week and 2 weeks respectively, yielding non-overlapping (independent) data. For forecasting time horizons of 1 month and above we use a monthly time step, which results in non-overlapping data for the 1 month horizon, but overlapping (dependent) data for the longer time horizons. We have chosen to include 4 different maturities for each horizon, in order to test our model for different points of the yield curve.

7.2 Backtesting results

Since we have 37 different LIBOR curves we have decided to use the USD LIBOR as an example in the backtesting section. Our code can of course easily be applied to any currency in the data set.

First, we forecasted the LIBOR rates at maturities of three months, one year, five years and ten years using an AR model. We used a 1 week, 2 weeks, 1 month and 3 months forecasting horizons starting from the 3 March 2008. To get the forecasting distribution we assumed a normal errors and used the mean and variance from the predictions at a given horizon to get a Monte Carlo simulation to estimate the yields. In Figure 7.1 below we present the probability density functions of the USD LIBOR forecasts. The forecast are mapped to a probability density for each horizon. It answers the question "How likely was the realized outcome according to our model?" We also added a line as representing the realized value of the rate. As seen, the results vary quite a lot, but the overall performance is rather poor, especially for longer maturities and longer forecasting horizons. Two exceptions though are the 1 week and 2 week horizon forecast for the 10 year maturity. These results does however not tell us anything about the overall model performance since we have only forecasted it starting from one initialization point. In order to evaluate the model performance in more detail we proceed with a more thorough analysis below using boostrapping and multiple initialization points.

FIGURE 7.1: Probability Density Functions of USD LIBOR forecasts for different maturities and horizons



Next we tried to improve our results by bootstrapping, we used boostrapping since it is a simple way to get the a simulation of the yield distribution. However, the performance is not good. The methodology we applied was to estimate the parameters using our AR(1) model and then bootstrapping the model using the in sample residuals. We used 10 percent of the sample as our calibration window and let this window vary over multiple initialing points. At each initialization point we calibrated our model (however kept lambda fixed) and forecasted the parameters to get the yield curve h step into the future. To get a more realistic performance we also scaled the variance to match the observed time series better. However, the results we obtained still suggest that our model does not perform well.

The graphs on the following pages illustrate the performance of our backtest. The bars illustrates the share of times the actual yield was outside our specified percentile and the horizontal line show the sought after percentile. So in a perfect model the share of times the actual yield was greater than e.g. the 95th percentile would be 0.05 and the hence the height of the bars should be level with the horizontal line. The results in Figure 7.2 to 7.5 show that the model does not perform well and that the actual occurrence of LIBOR rates outside the specified percentiles are more common than what we would like for it to be a "good" model performance. In general the model seems to be better at the shorter end of the curve and for a shorter forecasting period. Also for the 5 year LIBOR curve we see that all actual rates are within even the 90th percentile which would suggest that the model performs well. However we see that the variance in the AR parameters are much higher for this part of the curve making the sample really wide and hence letting all actual rates be within the specified percentiles. So the conclusion of the back testing results must be that our model does not perform well for all horizons and all parts of the curve. We are not very surprised by this result since it is consistent with the result we got in the forecasting section and the result is also consistent with what other authors find. We got similar results for other currencies (than the USD). To improve our backtest a model for the variance could be applied to give us a better fit.

Figure 7.2: Backtest of 3 month USD LIBOR at different forecasting horizons.

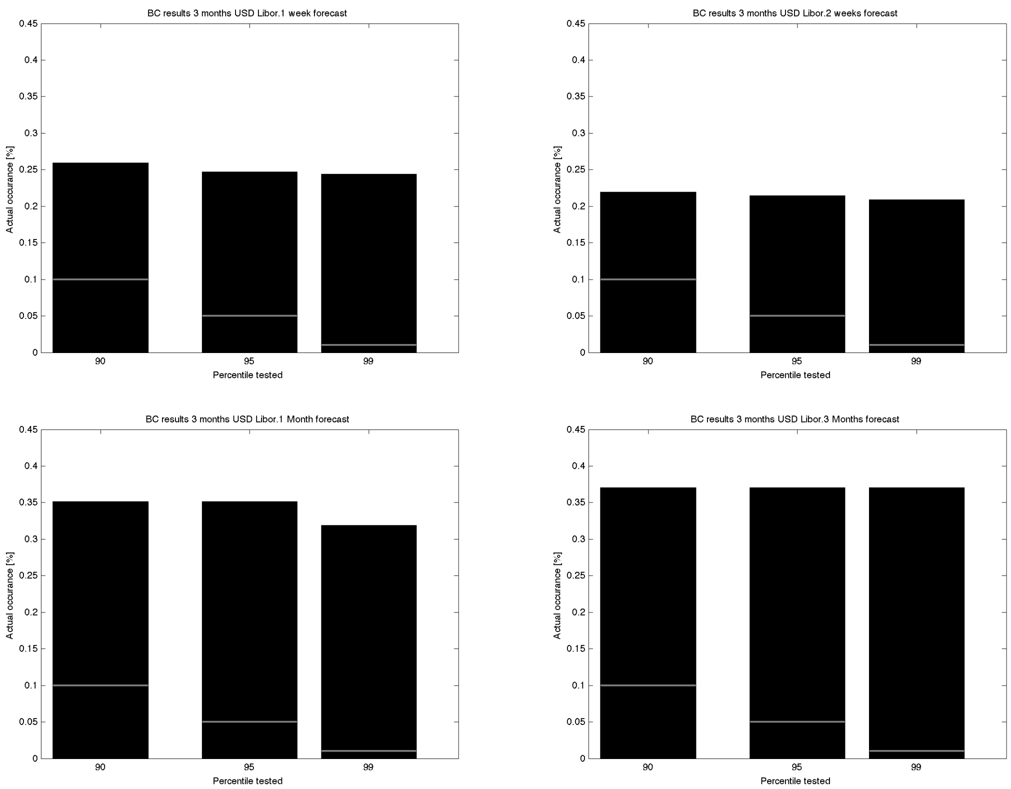


Figure 7.3: Backtest of 1 month LIBOR at different forecasting horizons.

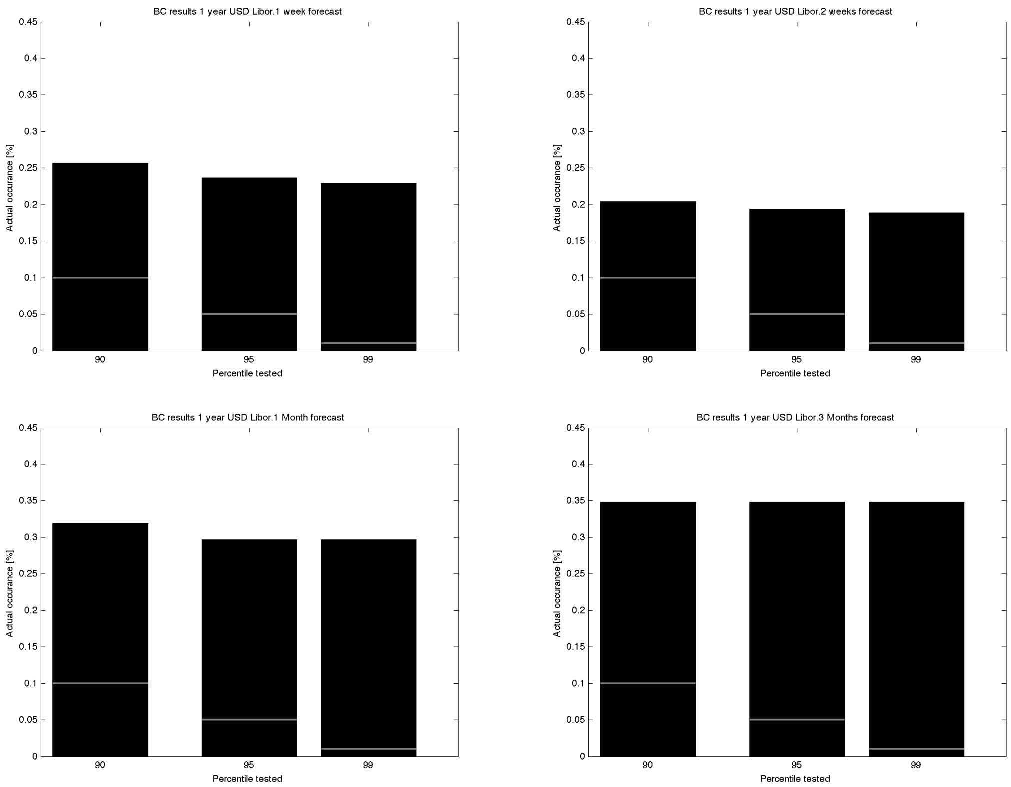


Figure 7.4: Backtest of 5 year LIBOR at different forecasting horizons.

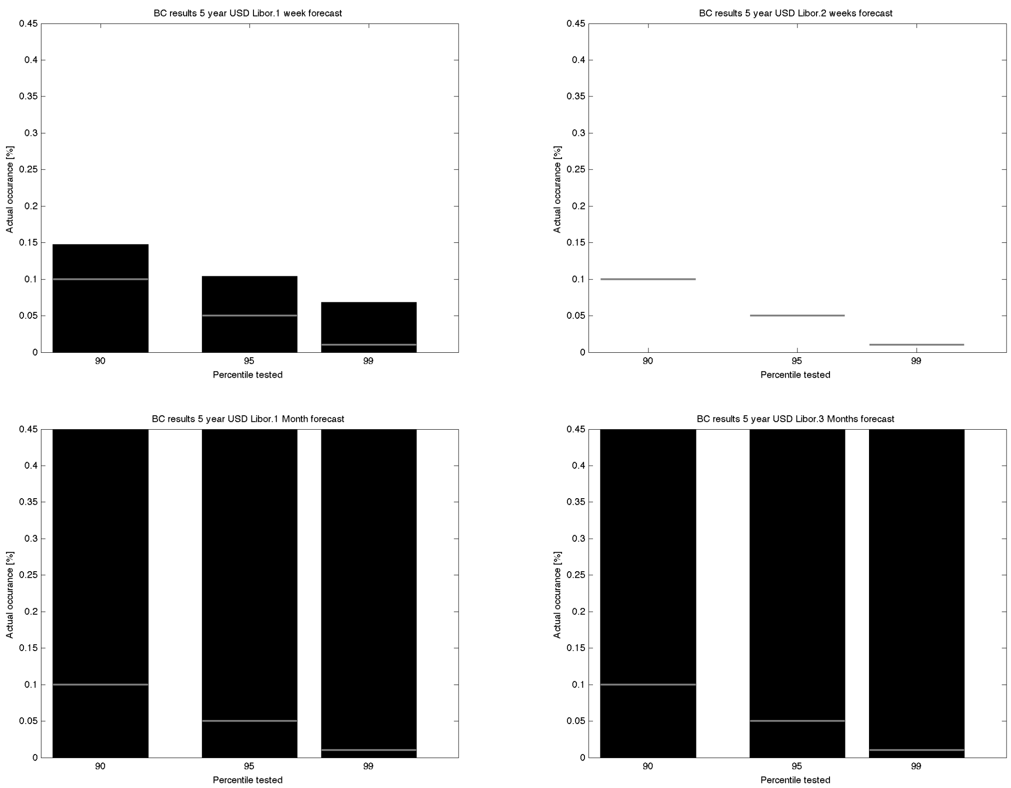
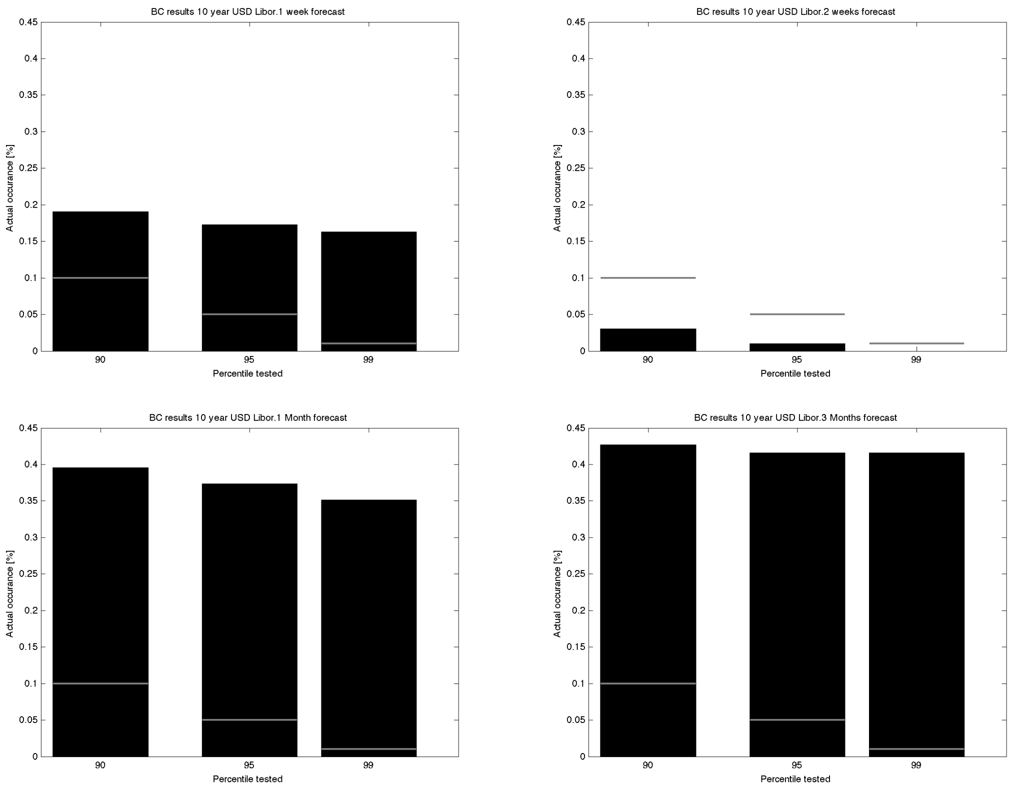


Figure 7.5: Backtest of 10 year LIBOR at different forecasting horizons.



To further illustrate our backtesting results and give numbers to the figures above we present in the following tables the number and share of exceptions (i.e. the number (share) of times the actual yield was outside the specified percentile) for each USD LIBOR rate maturity and forecasting horizon.

Table 7.1: Backtesting result 3 month USD LIBOR

Horizon	1w forecast		2w forecast		1m forecast		3m forecast		6m forecast	
Percentile	Nr	%	Nr	%	Nr	%	Nr	%	Nr	%
99th	99	25.19%	33	16.84%	32	35.16%	32	35.96%	33	38.37%
95th	102	25.95%	35	17.86%	33	36.26%	32	35.96%	33	38.37%
90th	106	26.97%	38	19.39%	33	36.26%	33	37.08%	33	38.37%

Table 7.2: Backtesting result 1 year USD LIBOR

Horizon	1w forecast		2w forecast		1m forecast		3m forecast		6m forecast	
Percentile	Nr	%	Nr	%	Nr	%	Nr	%	Nr	%
99th	96	24.43%	21	10.71%	31	34.07%	29	32.58%	53	61.63%
95th	101	25.70%	27	13.78%	31	34.07%	30	33.71%	60	69.77%
90th	103	26.21%	30	15.31%	31	34.07%	31	34.83%	65	75.58%

TABLE 7.3: Backtesting result 5 year USD LIBOR

Horizon	1w forecast		2w forecast		1m forecast		3m forecast		6m forecast	
Percentile	Nr	%	Nr	%	Nr	%	Nr	%	Nr	%
99th	62	15.78%	0	0.00%	53	58.24%	48	53.93%	55	63.95%
95th	72	18.32%	0	0.00%	54	59.34%	53	59.55%	58	67.44%
90th	89	22.65%	0	0.00%	54	59.34%	54	60.67%	61	70.93%

TABLE 7.4: Backtesting result 10 year USD LIBOR

Horizon	1w forecast		2w forecast		1m forecast		3m forecast		6m forecast	
Percentile	Nr	%	Nr	%	Nr	%	Nr	%	Nr	%
99th	69	17.56%	3	1.53%	38	41.76%	40	44.94%	44	51.16%
95th	75	19.08%	6	3.06%	41	45.05%	41	46.07%	46	53.49%
90th	84	21.37%	14	7.14%	43	47.25%	41	46.07%	47	54.65%

In order to provide more insights from our results Figure 7.6 to 7.8 below display plots showing forecasts at different confidence levels versus the realized 10 year USD LIBOR rate. In each figure the forecasting has been computed for all horizons up to 3 months for a specific initialization point. The forecast curves are computed using market data and calibration data available at the initialization point, meaning that the forecasts done at each initialization point are computed using different market data. The confidence levels seen in the figures are the 90th (green), 95th (yellow) and 99th (red) boostrapped confidence intervals. As illustrated below we can see that the intervals are wide and the actual rate is often outside the confidence levels, which indicates a poor model fit:

FIGURE 7.6: Forecasts confidence levels versus realized USD LIBOR 10 year, 7th July 2012

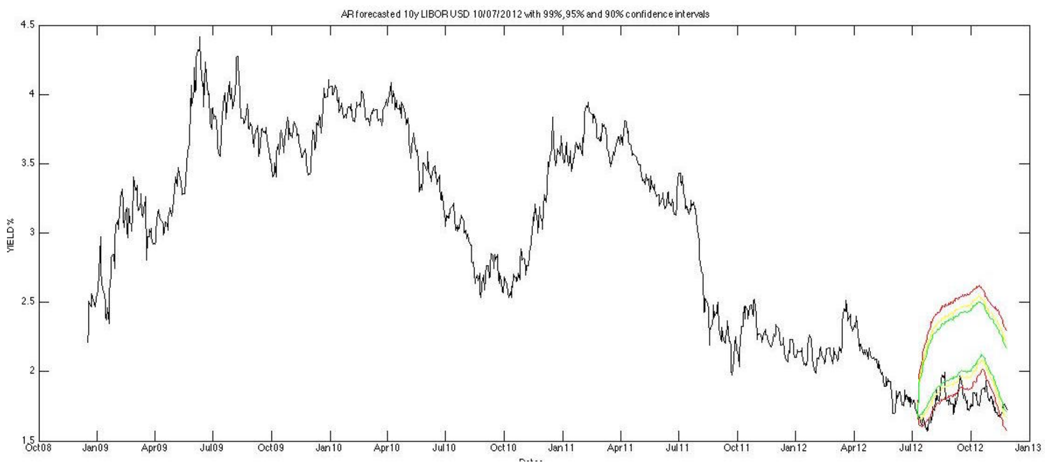


Figure 7.7: Forecast confidence levels versus realized USD LIBOR 10 year, 27th November 2012

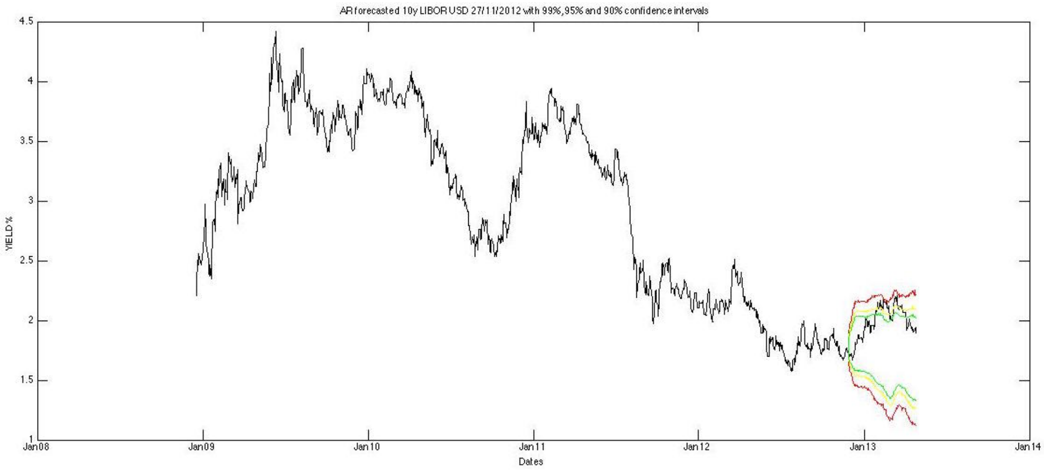
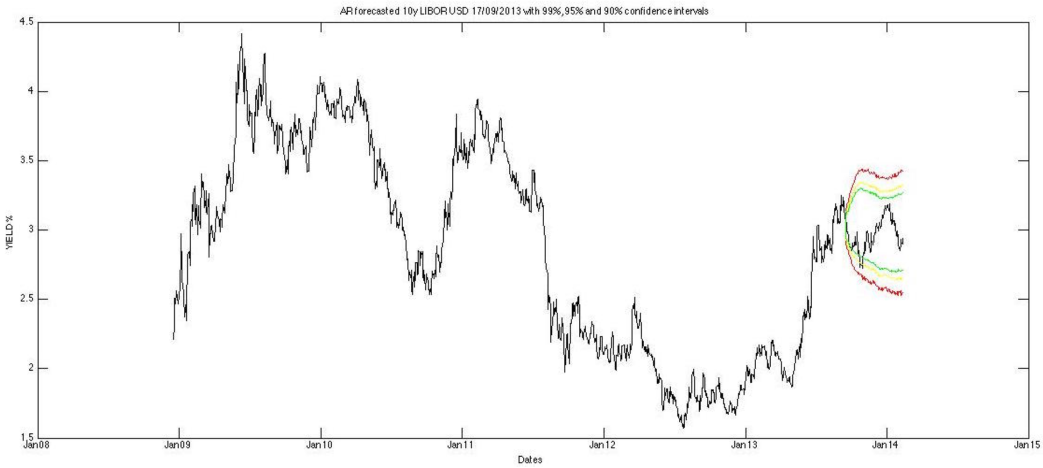


Figure 7.8: Forecasts confidence levels versus realized USD LIBOR 10 year, 17th September 2013



In conclusion we believe that the bad performance of the models in the backtest follows the same reasons as stated in chapter 6 i.e. bad model, incorrect coding or unpredictable yields using this model. Here we should also add that one could probably improve the backtesting performance by having a better model for the variance in the Monte Carlo and bootstrapping techniques. Also we could potentially improve our backtesting results by not keeping lambda fixed, this would however of course increase the complexity in the calibration and introduce a non-linear optimization problem which would increase the computing time substantially. As seen in the out-of-sample forecasting the DNS model with time dependent lambda provided slightly better results than that of the DNS model with fixed lambda.

Chapter 8

Concluding Remarks

In this thesis we have analyzed the forecasting performance of three versions of the Dynamic Nelson-Siegel (DNS) model applied to LIBOR rates during the time period January 2003 to March 2015. The first version represents the standard DNS AR(1) model with a fixed decay parameter (λ). This version is widely studied by researchers and also widely used by central banks. Our second DNS model also has the same AR(1) factor dynamics, but it has a time dependent decay parameter, i.e., (λ) varies over time. Lastly, for completeness we analysed a DNS model with VAR(1) factor dynamics, to fully test our DNS models. For comparison, we also employed a random walk model, simple AR(1) and VAR(1) models, AR(1) on three principal components, and a slope regression model.

First, we find that all our models demonstrate convincing in-sample performance. Starting with the DNS model with fixed lambda, the model shows good empirical fit to both the USD and DKK LIBOR yield curves, as shown in the graphs and tables in Chapter 5. We then provide results suggesting that the DNS model with time dependent lambda is superior, yielding a closer fit to actual rates than when lambda is time-invariant. Introducing the Dynamic Nelson-Siegel-Svensson model allows us to even further improve the closeness of fitted yields to the raw yield curve, by adding the second curvature factor to this model. We demonstrate these points graphically and in tables of results. The tables include the in-sample RMSE measure; our principal measure of goodness-of-fit. Thus, the Dynamic Nelson-Siegel-Svensson model is shown to be superior because it provides the smallest RMSE and hence the best model fit.

These results are consistent with the findings of e.g. Diebold and Rudebusch (2012) and Christensen et al. (2009). Relating this evidence back to the research questions in the introduction, the results obtained for the in-sample analysis allow us to provide positive answers to the first two questions:

1. Are we able to produce the good empirical fit that the Dynamic Nelson-Siegel model has become known for?

2. Are we able to improve the fitted yields by using the Svensson (1995) extension, the Dynamic Nelson-Siegel-Svensson model?

Second, our results indicate that it is not only possible to improve the in-sample performance, but also the out-of-sample forecasting performance by extending the DNS model to include a time-varying decay parameter, i.e., by letting lambda vary over time. Relative to Koopman et al. (2010) we find similar results in this context, that it might be possible to improve the forecasting results using this model setting. As argued by Diebold and Li (2006) it does enforce some further complexity in the estimation procedure, however our results indicate that it might be worth the while. In 14 out of 21 possible cases included in the out-of-sample forecasting analysis, the DNS model with time dependent lambda performs better than the DNS fixed-lambda counterpart. This enables us to provide a positive answer to the third question stated in the introduction:

3. Are we able to improve the forecasted yields by letting lambda vary over time?

However, when it comes to the fourth question:

4. How does the DNS model perform compared to other natural forecasting competitors?

we can conclude that the DNS models does not yield superior results relative to both the Random Walk model and the simple AR(1) on yields model. This indicates that the DNS type models are not good when it comes to forecasting the yield curve. For any maturity and horizon included in our analysis, the DNS models have poorer performance than the benchmark models. Thus, in contrast to Diebold and Li (2006), who find that the DNS model are able to better predict government bond yields than the Random Walk model for longer horizons, we argue that there is no convincing advantage in using the more complicated Dynamic Nelson-Siegel model over a simple AR(1) or random walk model, even for longer horizons in our context. As illustrated in Figure 6.1 it might be that for some periods it is equally likely that the rate will move up and down after the sample period, and in those cases the random walk model will perform well against a model like the DNS model.

Our fifth research question was meant to determine if we had a model that could be used to forecast yield curves in a way which could then be inserted as a risk factor into a CCR model:

5. Are we able to produce a forecasting model for the LIBOR rates that could be used in a CCR model?

The answer to this last question goes hand in hand with the results pertaining to the fourth question: our backtesting results show that the model does not forecast well and more development needs to be done before it can be inserted as a risk factor in a comprehensive CCR model. As it stands, we believe that it does not add any value to include the forecasted DNS model yields in a CCR model, since it would only create more uncertainties than it solves because of the poor performance.

Further studies need to be conducted to develop a forecasting model suitable to include in CCR measurement. It seems like the model either is not sophisticated enough or not simple enough for this matter, related to the rather famous "KISS principle" in forecasting, developed by Zellner (1992) – "Keep it sophisticatedly simple".

Despite the poor results in this section we would still like to enhance the fact that we add some value and insights about the predictive ability to forecast LIBOR rates using the DNS model, an area that is not yet as explored as investigating the DNS model performance applied to Government Bonds. Our recommendation is to conduct a similar study using the State-Space Model approach, in order to provide a better suited model for including in a CCR measurement.

8.1 Limitations and Extensions

In our thesis we used the two-step estimation procedure for estimating the latent factors and parameters of the Dynamic Nelson-Siegel specifications, as proposed by Diebold and Li (2006). Another approach is to use the State-Space Model (SSM) and Kalman filter to estimate and forecast the Dynamic Nelson-Siegel model parameters via maximum likelihood. Using this approach, one could possibly obtain a better suited model for forecasting future yields of LIBOR rates. We did not include this in our thesis due to space limitations, but this would be a natural extension for future research.

Furthermore, we believe that the backtesting section could be improved by introducing a dynamic decay factor (λ), as done in the out-of-sample forecasting analysis. We also envisage possible improvements from using a better model for computing the variance. Furthermore, if space had permitted, it would have been interesting to further extend the backtesting section by including multiple risk factors (e.g., foreign exchange rates, or multiple interest rates) and providing an application towards specific CCR measurement.

Another possible extension would also be to follow Diebold and Rudebusch (2012) and make the DNS model consistent with the absence of arbitrage. However, as pointed out by Diebold and Li (2006), the use of no-arbitrage models does not necessarily increase the forecasting ability, although this is to be determined in further studies.

Appendix A

Proposition AFNS-adjustment.

PROPOSITION 2.1. Assume that the instantaneous risk-free rate is defined as

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

Also, assume that the state variables $X_t = (X_t^1, X_t^2, X_t^3)$ are described by the following system of stochastic differential equations (SDEs) under the risk-neutral Q-measure:

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

Then, zero-coupon bond prices are given by

$$P(t,T) = E_t^Q \left[exp\left(-\int_t^T r_u du \right) \right] = exp\left(B^1(t,T)X_t^1 + B^2(t,T)X_t^2 + B^3(t,T)X_t^3 + C(t,T) \right),$$

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

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

and

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

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

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

and

$$
\begin{aligned}
C(t,T) =& (K^Q\theta^Q)_2 \int_t^T B^2(s,T)ds + (K^Q\theta^Q)_3 \int_t^T B^3(s,T)ds\\
&+ \frac{1}{2}\sum_{j=1}^{3}\int_t^T (\Sigma' B(s,T)B(s,T)'\Sigma)_{j,j} ds
\end{aligned}
$$

Finally, zero-coupon bond yields are given by

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

The proof for Proposition 2.1 is given in Christensen et al. (2007). Proposition 2.1 defines the class of AFNS models. The factor loadings in the AFNS models exactly match the ones for Nelson-Siegel models, except for the last additional term in the yield function, namely, $-\frac{C(t,T)}{T-t}c$, which depends only on maturity. This additional term is the "yield-adjustment" term, and is the crucial difference between AFNS and DNS models, bringing the DNS models into the arbitrage free setting. The yield-adjustment term has the following form:

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

Given a general volatility matrix

$$
\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13}\\ \sigma_{21} & \sigma_{22} & \sigma_{23}\\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix},
$$

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

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

where

$$
\begin{aligned}
\bar{A} &= \sigma_{11}^2 + \sigma_{12}^2 + \sigma_{13}^2, \\
\bar{B} &= \sigma_{21}^2 + \sigma_{22}^2 + \sigma_{23}^2, \\
\bar{C} &= \sigma_{31}^2 + \sigma_{32}^2 + \sigma_{33}^2, \\
\bar{D} &= \sigma_{11}^2\sigma_{21}^2 + \sigma_{12}^2\sigma_{22}^2 + \sigma_{13}^2\sigma_{23}^2, \\
\bar{E} &= \sigma_{11}^2\sigma_{31}^2 + \sigma_{12}^2\sigma_{32}^2 + \sigma_{13}^2\sigma_{33}^2, \\
\bar{F} &= \sigma_{21}^2\sigma_{21}^2 + \sigma_{22}^2\sigma_{22}^2 + \sigma_{23}^2\sigma_{23}^2.
\end{aligned}
$$

According to Christensen et al. (2009) this result has two implications. First, empirical implementation of AFNS models are greatly facilitated since zero-coupon bond yields in the AFNS models are given by an analytical formula. Second, only six of the nine underlying volatility parameters can be identified, $\bar{A}, \bar{B}, \bar{C}, \bar{C}, \bar{D}, \bar{E}$ and $\bar{F}$. Hence, compared to a general volatility matrix we are missing three volatility parameters, resulting in the most flexible identifiable specification of the AFNS model is given by the following triangular matrix (upper or lower triangular being irrelevant for model fit):

$$
\Sigma = \begin{pmatrix} \sigma_{11} & 0 & 0 \\ \sigma_{21} & \sigma_{22} & 0 \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix}.
$$

Bibliography

Abramowitz, M. and Stegun, I. A. (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York, ninth dover printing, tenth gpo printing edition.

BIS (2005). *Zero-coupon yield curves: technical documentation*. Number 25 in BIS Papers. Bank for International Settlements.

BIS (2010). Sound practices for backtesting counterparty credit risk models . *Bank for International Settlements*.

Björk, T. (2009). *Arbitrage Theory in Continuous Time*. Oxford University Press, third edition.

Björk, T. and Christensen, B. J. (1999). Interest rate dynamics and consistent forward rate curves. *Mathematical Finance*, 9(4):323–348.

Black, F. (1995). Interest rates as options. *The Journal of Finance*, 50(5):1371–1376.

Blaskowitz, O. and Herwartz, H. (2009). Adaptive forecasting of the EURIBOR swap term structure. *Journal of Forecasting*, 28(7):575–594.

Brigo, D. and Mercurio, F. (2006). *Interest Rate Models: Theory and Practice - with Smile, Inflation and Credit*. Springer Verlag.

Christensen, J. H. E., Diebold, F. X., and Rudebusch, G. D. (2007). The Affine Arbitrage-Free Class of: Nelson-Siegel Term Structure Models. NBER Working Papers 13611, National Bureau of Economic Research, Inc.

Christensen, J. H. E., Diebold, F. X., and Rudebusch, G. D. (2009). An arbitrage-free generalized Nelson–Siegel term structure model. *Econometrics Journal*, 12(3):C33–C64.

Coroneo, L., Nyholm, K., and Vidova-Koleva, R. (2011). How arbitrage-free is the Nelson-Siegel model? *Journal of Empirical Finance*, 18(3):393–407.

Cox, J. C., Ingersoll, Jonathan E, J., and Ross, S. A. (1985). A Theory of the Term Structure of Interest Rates. *Econometrica*, 53(2):385–407.

Dai, Q. and Singleton, K. J. (2000). Specification Analysis of Affine Term Structure Models. *Journal of Finance*, 55(5):1943–1978.

Dai, Q. and Singleton, K. J. (2002). Expectation puzzles, time-varying risk premia, and affine models of the term structure. *Journal of Financial Economics*, 63(3):415–441.

Diebold, F. X. (2008). *Elements of Forecasting*. Thomson, South-Western, 4th edition.

Diebold, F. X., Gunther, T. A., and Tay, A. S. (1998). Evaluating Density Forecasts with Applications to Financial Risk Management. *International Economic Review*, 39(4):863–83.

Diebold, F. X. and Li, C. (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics*, 130(2):337–364.

Diebold, F. X. and Lopez, J. A. (1996). Forecast Evaluation and Combination. NBER Technical Working Papers 0192, National Bureau of Economic Research, Inc.

Diebold, F. X. and Rudebusch, G. D. (2012). *Yield Curve Modeling and Forecasting: The Dynamic Nelson-Siegel Approach*, volume 1 of *Economics Books*. Princeton University Press.

Diebold, F. X., Rudebusch, G. D., and Borag[caron]an Aruoba, S. (2006). The macroeconomy and the yield curve: a dynamic latent factor approach. *Journal of Econometrics*, 131(1-2):309–338.

Duffee, G. R. (2002). Term Premia and Interest Rate Forecasts in Affine Models. *Journal of Finance*, 57(1):405–443.

Duffie, D. and Kan, R. (1996). A Yield-Factor Model Of Interest Rates. *Mathematical Finance*, 6(4):379–406.

Duffie, D. and Singleton, K. J. (1997). An Econometric Model of the Term Structure of Interest-Rate Swap Yields. *Journal of Finance*, 52(4):1287–1321.

Fama, E. F. and French, K. R. (1992). The Cross-Section of Expected Stock Returns. *Journal of Finance*, 47(2):427–65.

Gregory, J. (2010). *Counterparty credit risk : the new challenge for global financial markets*. Wiley finance series. Chichester : Wiley, 1. ed. edition. Literaturverz. S. [405] - 409.

Gürkaynak, R. S., Sack, B., and Wright, J. H. (2007). The U.S. Treasury yield curve: 1961 to the present. *Journal of Monetary Economics*, 54(8):2291–2304.

Heath, D., Jarrow, R., and Morton, A. (1992a). Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claims Valuation. *Econometrica*, 60(1):77–105.

Heath, D., Jarrow, R., and Morton, A. (1992b). Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claims Valuation. *Econometrica*, 60(1):77–105.

Hull, J. and White, A. (1990). Pricing Interest-Rate-Derivative Securities. *Review of Financial Studies*, 3(4):573–92.

Hull, J. and White, A. (2013). LIBOR vs. OIS: The derivatives discounting dilemma. *Journal of Investment Management*, 11(3):14–27.

Jolliffe, I. (2002). *Principal Component Analysis*. Springer Series in Statistics. Springer.

Joslin, S., Priebsch, M., and Singleton, K. J. (2014). Risk premiums in dynamic term structure models with unspanned macro risks. *Journal of Finance*, 69(3):1197–1233.

Kani, I. (2007). Interest Rate Fundamentals.

Kenyon, C. and Stamm, R. (2012). Discounting, libor, cva and funding : interest rate and credit pricing. http://www.gbv.de/dms/zbw/723683980.pdf.

Knez, P. J., Litterman, R., and Scheinkman, J. A. (1994). Explorations into Factors Explaining Money Market Returns. *Journal of Finance*, 49(5):1861–82.

Koopman, S. J., Mallee, M. I. P., and Van der Wel, M. (2010). Analyzing the Term Structure of Interest Rates Using the Dynamic Nelson-Siegel Model With Time-Varying Parameters. *Journal of Business & Economic Statistics*, 28(3):329–343.

Lando, D. and Poulsen, R. (2006). Lecture Notes for Finance 1 (and More). In *Lecture notes*, volume 1.

Litterman, R. B. and Scheinkman, J. (1991). Common Factors Affecting Bond Returns. *The Journal of Fixed Income*, 1(1):54–61.

Lord, R. and Pelsser, A. (2007). Level-Slope-Curvature - Fact or Artefact? *Applied Mathematical Finance*, 14(2):105–130.

Macaulay, F. R. (1938). Some Theoretical and Practical Difficulties of Comparing Long Term Interest Rates at Different and Especially at Widely Separated Dates. In *Some Theoretical Problems Suggested by the Movements of Interest Rates, Bond Yields and*

Stock Prices in the United States since 1856, NBER Chapters, pages 54–84. National Bureau of Economic Research, Inc.

Nelson, C. R. and Siegel, A. F. (1987). Parsimonious Modeling of Yield Curves. *The Journal of Business*, 60(4):473.

Nyholm, K. (2008). *Strategic Asset Allocation in Fixed Income Markets: A Matlab Based User's Guide*. The Wiley Finance Series. Wiley.

Ruiz, I. (2012). Backtesting counterparty risk: How good is your model? Technical report, Working Paper, iRuiz Consulting.

Senghore, M. (2013). *Modeling Swedish government yields with the Dynamic Nelson Siegel and the Dynamic Nelson Siegel Svensson Model*. PhD thesis, Lund University.

Sengupta, R. and Tam, Y. M. (2008). The LIBOR-OIS spread as a summary indicator. *Monetary Trends*, (Nov).

Svensson, L. E. O. (1995). Estimating and Interpreting Forward Interest Rates; Sweden 1992-1994. IMF Working Papers 94/114, International Monetary Fund.

Tay, A. S. and Wallis, K. F. (2000). Density forecasting: a survey. *Journal of Forecasting*, 19(4):235–254.

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5(2):177–188.

Yu, W.-C. and Zivot, E. (2011). Forecasting the term structures of Treasury and corporate yields using dynamic Nelson-Siegel models. *International Journal of Forecasting*, 27(2):579–591.

Zellner, A. (1992). Statistics, Science and Public Policy. Papers 92-21, California Irvine - School of Social Sciences.

Zhu, S. H. and Pykhtin, M. (2008). A Guide to Modeling Counterparty Credit Risk. *Social Science Research Network Working Paper Series*.