Factor Models for the Term Structure of STIBOR Rates

Jonatan Persson

Examensarbete för Kandidatexamen
2021:K28

Lunds universitet

Naturvetenskaplig fakultet
Matematikcentrum
Matematisk statistik

CENTRUM SCIENTIARUM MATHEMATICARUM

Abstract: The yield curve of a collection of debt contracts describes the yield of the debt contract as a function of the length-to-maturity of the contract. It turns out that these yield curves provide useful insight about the economy as a whole and can, for example, be used to predict short-term economic downturns. Therefore, it is of utmost importance that financial analysts and decision-makers are able to accurately estimate and model these yield curves.

This thesis will utilize a collection of parametric yield curve models, usually called the Nelson-Siegel family. More, specifically we will use a so called two-factor arbitrage-free dynamic Nelson-Siegel model. We will apply this model to data collected from the Swedish central bank, so called Stockholm Interbank Offering Rate (or STIBOR for short). We will also investigate whether including the repo rate, the interest rate set by the Swedish central bank, in the model improves modelling and forecasting capabilities.

Acknowledgments: I want to show my gratitude to my supervisor Magnus Wiktorsson for introducing me to the project, guiding me through the process of thesis writing and being able to answer any questions that I have posed.

Contents

1 Introduction

The term structure of a bond (sometimes called the term structure of the bond) describes yield of the bond as a function of the time-to-maturity. Being able to model the term structure of interest rates is something that is of interest to a variety of financial and economic decision makers, since it is used as an indicator of future economic growth. Because of this, there has been put considerable effort into developing flexible, parsimonious term structure models that are also simple for practitioners to use. One of the most commonly used family of models is the parametric Nelson-Siegel family. They have become popular because of their ability the capture the most common shapes of the yield curve, their relative simplicity and their ease of interpretation. The original Nelson-Siegel model was presented by Nelson and Siegel [16] in 1985 and has since then been further developed by a number of different people, e.g. Svensson [22] in 1994 and Diebold and Li [9] in 2006.

It is our intention to apply the arbitrage-free Nelson-Siegel model developed by Christensen et. al [8] to interbank reference rates used on the Stockholm market, called STIBOR. We will however make some modifications to this model. STIBOR rates are only available for maturities shorter than a year, and thus the arbitrage-free Nelson-Siegel model might be unnecessarily complex four our purposes. Therefore, we will propose a restriction of the model and use this for our term structure estimation. We will also look at the STIBOR from a macroeconomic perspective, investigating if the Nelson-Siegel model can be improved by including the repo rate as an exogenous variable in the model. The repo rate is the interest rate set by the Swedish central bank, in an effort to control inflation. The repo rate is set on regular, relatively long intervals and the financial decision makers usually have some relatively good idea of what repo rate is going to be in the short-term future. Thus we consider the repo rate to be a good predictive variable

There are also several approaches to estimating the parameters of the Nelson-Siegel model. We will make use of state-space models in this effort, essentially modelling the parameters of the Nelson-Siegel model as latent time-series entries. This allows us to use some classic time-series techniques, such as the Kalman filter, in an effort to find the estimates that best fit our data.

The time period we will be looking over is 1998 to 2020. This interval will be split into two subintervals, one lasting from 1998 to 2013 and the other lasting from 2013 to 2020. The reason for doing this is two fold. Firstly, the period from 1998 to 2013 is characterized by a series of economic shocks, first in 2000 during the dot-com bubble, then the great recession that occurred in 2008 and last the European debt crisis that started in 2009. The period from 2013-2020 is in contrast characterized by stable, low and sometimes negative interest rates. We want to investigate whether the Nelson-Siegel model can withstand such volatility as is observed in the first interval or if it will perform better during more stable economic periods. Secondly, data is available for a total 8 different time-to-maturities during 1998 to 2013, but only for 6 different time-to-maturities during 2013 to 2020. Thus, splitting of the interval into two separate parts is also of practical importance.

2 Interest Rate Theory

An integral part of modern financial markets is the credit market, where governments, banks, companies and investors issue new debt and trade debt contracts. Throughout this section we refer to Byström [6] whenever we are discussing the fundamental concepts related to the credit market.

A bond is a debt instrument traded on the credit market, where the issuer of the bond is indebted to holder and has an obligation to reimburse the holder at a pre-specified end date. In contrast to the issuer, the holder is not required to wait until the end date to get reimbursed, and has the option to sell the bond at an earlier date. The debtholders are compensated for issuing the loan, the exact amount of compensation is referred to as the interest rate.

This section will be devoted to establishing some of the mathematical theory behind interest rates that will be needed for the concepts discussed later in this thesis. We begin by discussing the simplest bond available on the credit market, the zero-coupon bond.

2.1 Zero-Coupon Bonds and Forward Rates

The zero-coupon bond consist of only a single cash flow, unlike other forms of bonds, where the owner might receive so called coupon payments at regular intervals during the lifetime of the bond. This single cash flow occurs when the bondholder is reimbursed at the end of the contract. Other more complicated types of bond contracts often consist of collections of zero-coupon bonds, which makes proper understanding of zero-coupons fundamental for further analysis of the credit market. Zero-coupon bonds are often referred to as discount bonds, because they are said to be traded at a discount. This means that the price of the bond is lower than the actual nominal amount that is to be paid at the end of the contract.

We will in this thesis use the same setup for the formula for the zero-coupon price (as well as the forward yield curves, which will be discussed later) as the one used in Diebold and Li [9]. Before stating the formal pricing formula for zero-coupon bonds we have to discuss different compounding rates. The rate of compounding describes how frequently the interest accumulates and creates so called "interest on interest". There are several different compounding rates, such as annual compounding or continuous compounding. In this text we will be assuming continuous compounding, where we let the compounding frequency go to infinity. This leads to the following pricing formula for a τ-maturity zero-coupon bond:

Definition 2.1. *Let $P(\tau)$ denote the price of a zero-coupon bond with time to maturity τ with a single future cash flow equal to one unit of currency (the nominal value). Then if $y(\tau)$ is the yield to maturity of the bond we have:*

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

for $\tau \geq 0$.

The forward rate of a bond is the yield of the bond, at some fixed time point in the future. We have following relationship between the price of the zero-coupon bond and the instantaneous forward rate:

Definition 2.2. *Let $f(\tau)$ denote the forward rate of zero-coupon bond at τ time units from today. If $P(\tau)$ denotes the price of a τ-maturity bond with nominal value of one currency, we have the following formula:*

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

for $\tau \geq 0$.

Using Definitions 2.1 and 2.2, we can establish the following relationship between the yield curve and the forward rate:

Corollary. *If $y(\tau)$ is the yield to maturity of a τ-maturity bond and $f(\tau)$ is the forward rate, the following holds:*

$$y(\tau) = \frac{1}{\tau} \int_0^\tau f(s) ds$$

for $\tau \geq 0$.

This means that there is a fundamental relation between the price of a zero-coupon bond, the forward rate and the yield curve. If we know one of them, it means we know the value of the other two. For the rest of this text, we will primarily be dealing with the yield curve.

2.2 The Yield Curve

The yield curve, often referred to as the term structure, describes the yield to maturity y of bonds in the market as a function of their time to maturity τ, at a specific point in time t. Note that we are thus able to analyze the yield curves in two different ways, first cross-sectionally at a fixed time t, for varying maturities, and secondly as they evolve dynamically over time, for a fixed maturity τ.

Yield curves are frequently used by finance professionals as a general performance metric for the debt market and as a useful indicator of future economic growth, which is discussed further in, for example, Choudhry [7]. Typically this is done by analyzing the shape of the yield curve, which is usually divided into three main categories.

When the yield curve is upwards sloping, i.e. higher maturities corresponds with higher yields, the yield curve is said to be normal. This is an indication that investors believe in positive economic growth in the future and that interest rates for shorter maturities will increase. Thus the investors demand higher yield to compensate for investing in longer maturities now.

When the yield curve is downwards sloping, i.e. higher maturities corresponds with lower yields, the yield curve is referred to as being inverted. This is the opposite case of the normal curve and indicates a belief in negative economic growth in the future. Thus investors are willing to pay more for lower risk bonds with shorter time to maturity.

When different maturities correspond to roughly the same yield we have a flat yield curve. This indicates that the investors are inconclusive about the future growth of the economy or that the economy is entering a transitional period.

The interpretations of the yield curve shapes are based on the expectation hypothesis, which states that risk neutral individuals are indifferent between borrowing money on longer time to maturity or borrowing a series of consecutive loans with shorter time to maturity during the same period. Thus, assuming the market is free of arbitrage, the long term interest rate of a τ-maturity loan should equal the average expected short term interest rate for a series of successive loans with shorter time to maturity over the same τ year long period. For more on the expectation hypothesis, see Fregert and Jonung [11].

Usually, we are unable to observe the yields directly. Instead, we have to rely on indirect measurements, that we can later transform into the proper yields. One such measurements is the LIBOR rate, an abbreviation for London Interbank Offered Rate.

2.3 The LIBOR Rate

The LIBOR rate refers to the estimated average interest rate at which the most prominent banks at the London market lend money to each other over a time period with length τ. It is used as a daily reference rate for the pricing of a variety of financial instruments. Corresponding rates also exist for other national markets, e.g. the EU market uses the EURIBOR, the Stockholm market uses the STIBOR, etc. The LIBOR can be connected to the yield curve (and thus the zero-coupon price and the forward rate) in the following way, as is done in Björk [5]:

Definition 2.3. *Let $LIBOR(\tau)$ denote the LIBOR rate over a time period $[T, S]$ with length $\tau = S - T$ and let $y(\tau)$ denote the yield to maturity of a τ-maturity bond. Then the following relation holds:*

$$1 + \tau LIBOR(\tau) = e^{\tau y(\tau)}$$

for $\tau \geq 0$.

In this study we will be looking at the Swedish market, and will thus be interested in the STIBOR rate. The STIBOR is published every workday at 11:00am CET and is calculated differently according to the number of banks that have chosen to submit their rates, as is stipulated by the Swedish Financial Benchmark Facility [23]. The calculation scheme is as follows:

Number of submissions	Calculation procedure
$9 \geq$	Remove the two highest and two lowest submissions and take the arithmetic mean of the remaining.
6-8	Remove the highest and lowest submission and take the arithmetic mean of the remaining.
4-5	The arithmetic mean of the submissions.
≤ 3	Use yesterday's rate instead.

The STIBOR rate is also affected by other macroeconomic factors. In a report published by the Swedish central bank, Sveriges riksbank, [21] called "Stibor synas på nytt - en uppföljning", they concluded that one of the factors that should have an effect on the STIBOR rate was the interest rate at which the other banks were able to loan and place money in the central bank.

2.4 The Repo Rate

Sveriges riksbank is an independent government body, responsible for formulating, implementing and administrating the official monetary policy of Sweden. It has a stated goal of maintaining low and stable levels of inflation, with yearly inflation goal of about 2% [18]. Riksbanken measures inflation using the KPIF-index (consumer price index with fixed interest rate), which is in turn calculated by Statistics Sweden (Statistiska centralbyrån) [13].

The primary way in which the Riksbank tries to control inflation is summarized in Fregert and Jonung [11]. This is done by setting the repo rate, which is the interest rate of week-long securities issued by the Riksbank to the other large banks, so called Riksbank certificates. This allows the other banks to get rid of their surplus liquidity, which in turn effects the overnight rate. This is the interest rate at which the large bank lend to each other from one day to the next, hence the name. The Riksbank also tries to influence the long term interest rates by regularly publishing their own 3-years-ahead prognosis of the repo, basing it on a series of underlying macroeconomic variables. This gives the public an idea of the Riksbanks intended future monetary policy and will thus effect the long term rates, according to the expectation hypothesis discussed in section 2.2.

The repo rate was first introduced in 1994, replacing the previously used marginalränta (marginal rate). It is revised on a regular basis, usually several times a year, from which the repo curve gets its piece wise constant shape. The repo rate curve is shown in figure 1, from June 1994 to July 2020.

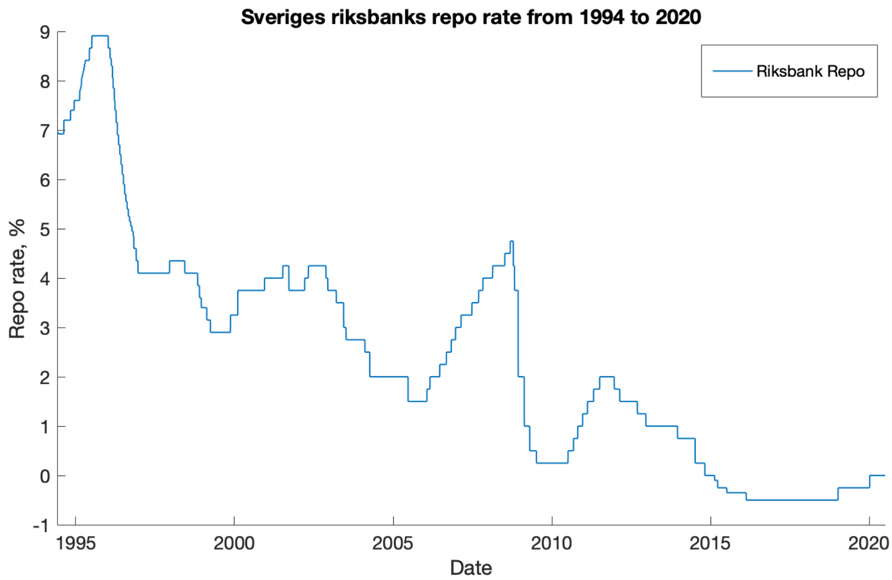


Figure 1: The repo rate set by Sveriges riksbank, from the start of use in June 1994 to July 2020.

3 Time Series Modelling

Before we are able to discuss the models used to estimate the STIBOR term structure, we have to familiarize ourselves with some of the statistical prerequisites of such models and estimation techniques.

3.1 Stochastic Processes and Stationarity

In this and the following subsection, we will be using the setup from Lindgren, Rootzén and Sandsten [14]. Random variables allow us to describe and analyze random phenomenon at a single point in time. However, we might also be interested in random phenomenon over time or in some higher dimensional space. To this end we use what is called stochastic processes, defined as follows:

Definition 3.1. *For a probability space* $(\Omega, \mathcal{F}, P)$ *and parameter space* T *we call the indexed family*

$$\{X(t); t \in T\}$$

of random variables a stochastic process. If $T \subset \mathbb{R}$ *we say the process is in continuous time. If* $T \subset \mathbb{Z}$ *we say the process is in discrete time, and it is sometimes called a time series instead.*

Next, we give definitions for two important categories of stochastic processes.

Definition 3.2. *Let* $\{X(t); t \in T\}$ *be stochastic process. Let* $\stackrel{d}{=}$ *refer to equality in distribution between two r.v.'s. Assume that the process satisfies*

$$(X(t_1), \dots, X(t_n)) \stackrel{d}{=} (X(t_1 + \tau), \dots, X(t_n + \tau))$$

for every $n \in \mathbb{Z}$, $t_1 \dots t_n \in T$ *and* τ *s.t.* $t_i + \tau \in T$. *Then* $\{X(t); t \in T\}$ *is referred to as a stationary process.*

The property of stationarity will ensure that the process is not affected by trends in deterministic values. Oftentimes, this definition of stationarity is very difficult for processes to fulfill, and we will instead use the following, less restrictive definition.

Definition 3.3. *Let* $\{X(t); t \in T\}$ *be stochastic process. If* $\mathrm{E}(X(t))$ *is constant for all* $t \in T$ *and* $\mathrm{Cov}(X(t), X(s))$ *is finite and only depends on* $\tau = s - t$ *then the process is referred to as being weakly stationary.*

Now we will move on to describe two frequently used time series models, the autoregressive (AR) and vector autoregressive (VAR) models.

3.2 AR and VAR Models

Before properly defining the AR and VAR models, we will give a definition for the white noise process.

Definition 3.4. *If* $\{\epsilon_t; t = 0, \pm 1, \ldots\}$ *is stochastic process with zero mean and autocovariance function defined by*

$$\mathrm{Cov}(\epsilon_s, \epsilon_t) = \begin{cases} \sigma^2, & \text{if } s = t \\ 0, & \text{otherwise} \end{cases}$$

we say that it is a discrete time white noise process, or innovation process.

Now we are ready to state the definition of the AR model.

Definition 3.5. *A discrete time stochastic process* $\{X_t; t = 0, 1, \ldots\}$ *is called an autoregressive process of order p, AR(p), if it satisfies*

$$X_t = \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + \epsilon_t$$

where $\{\epsilon_t\}$ *is a white noise process and the* ϕ_i*'s are scalars.*

Further we have that if the complex polynomial $\Phi(z) = z^p - \phi_1 z^{p-1} - \ldots - \phi_p$ has all of its roots inside the unit circle, then the corresponding AR(p) process satisfies the conditions of weak stationarity.

The AR model is a frequently used model in economic time series modelling. It also makes up one of the components of the commonly utilized ARMA model, together with the moving average (MA) process. The AR model can be further generalized to allow for modeling of multiple factors and how they relate to each other over time. This generalized AR model is called the vector autoregressive (VAR) model, presented in Shumway and Stoffer [20].

Definition 3.6. *A discrete time vector process* $\{X_t = (X_{1t}, \ldots, X_{nt})'; t = 0, 1, \ldots\}$ *is called a vector autoregressive process of order p, VAR(p), if it satisfies*

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

where $\{\epsilon_t\}$ *is a white noise vector process and the* Φ_i*'s are* $n \times n$ *matrices.*

Much like for the simple AR process, we can also formulate stationarity conditions for the VAR process. In particular, the VAR(1) model satisfies the conditions of weak stationarity if the matrix Φ has eigenvalues inside the unit circle, see Jakobsson [12]. Both the AR and VAR models can be further generalized by adding deterministic values as well as exogenous variables to the model. Sometimes we are not able to observe the realizations of our time series directly, and have to rely on indirect measurements. A way of getting around this is to use so called state-space model, sometimes referred to as dynamic linear models.

3.3 State-Space Representation

For this and the next subsection, we will be referring to Shumway and Stoffer [20]. The state space model relies on what is called a state equation to model the underlying dynamics of a particular system that we are interested in. This state equation takes the form of a VAR(1) model with an added exogenous variable:

$$X_t = A_t X_{t-1} + B_t u_t + w_t$$

where X_t is a $n \times 1$ state vector, A_t is a $n \times n$ state matrix, u_{t-1} is a $s \times 1$ vector of exogenous variables, B_t is a $n \times s$ matrix and $\{w_t\}$ is a Gaussian white noise vector process with a $n \times n$ covariance matrix Q.

However, we do not directly observe our state vectors and instead have to rely on indirect measurements. We relate these measurements to the state vector through the measurement equation:

$$Y_t = C_t X_{t-1} + D_t u_t + v_t$$

where Y_t is a $m \times 1$ vector of measurements made at time t, C_t is a $m \times n$ measurement matrix, D_t is a $m \times s$ matrix and $\{v_t\}$ is a Gaussian white noise vector process with a $m \times m$ covariance matrix R.

We will henceforth assume that the white noise processes $\{w_t\}$ and $\{v_t\}$ are uncorrelated. Given the indirect measurements, we want to be able to estimate our latent state vectors. This is most commonly accomplished using the so called Kalman filter.

3.4 Kalman Filter

The linear estimators of the state vectors that have the smallest average square of errors are obtained using the Kalman Filter, assuming we have constructed our state space model as above, with linear state and observation equations, as well as Gaussian noise processes. To compress the length of some of the equations below, we introduce the following notation:

$$X_t^s = \mathrm{E}(X_t | Y_1, \ldots, Y_s)$$

$$P_t^s = \mathrm{E}\left\{(X_t - X_t^s)(X_t - X_t^s)'\right\}$$

We will not go into detail about the derivation of the Kalman filter, but instead mention that the theoretical basis of the algorithm consists of viewing the conditional expectation X_t^s as a projection of X_t on to the function space spanned by the measurements $Y_1, \ldots, Y_s$. For more on this we refer to appendix B in Shumway and Stoffer [20].

The Kalman filter is a recursive algorithm that needs to be initialized by some X_0^0 and P_0^0. We assume that this initial state vector and covariance are normally distributed, and if this distribution is completely known through it's mean μ_0 and covariance Σ_0 we often let $X_0^0 = \mu_0$ and $P_0^0 = \Sigma_0$. Once the filter is initialized, we can get an a priori estimate of X_t^{t-1} using the previous state vector X_{t-1}^{t-1}, and similarly for P_t^{t-1} using P_{t-1}^{t-1}, by performing the prediction step:

$$
\begin{aligned}
X_t^{t-1} &= A_t X_{t-1}^{t-1} + B_t u_t \\
P_t^{t-1} &= A_t P_{t-1}^{t-1} A_t' + Q.
\end{aligned}
$$

Then, using our measurements, we can get a posterior estimates, by performing the update step:

$$X_t^t = X_t^{t-1} + K_t(Y_t - C_t X_t^{t-1} - D_t u_t)$$

$$P_t^t = (I - K_t C_t) P_t^{t-1}$$

where

$$K_t = P_t^{t-1} C_t' (C_t P_t^{t-1} C_t' + R)^{-1}$$

is called the Kalman gain. We reiterate that the Kalman filter only gives meaningful results if the state-space model in question is linear and Gaussian.

Often we do not have the parameters of our state-space model prespecified, and in such cases we have to estimate these parameters using, for example, the maximum likelihood method.

3.5 Maximum Likelihood

In this section, we will look at the likelihood function of the state-space model. First, the innovations of the Kalman filter are given by:

$$\epsilon_t = Y_t - \mathrm{E}(Y_t | Y_1, \ldots Y_{t-1}) = Y_t - C_t X_t^{t-1} - D_t u_t.$$

The innovations have zero mean and covariance matrix given by:

$$\Sigma_t = \mathrm{Var}(\epsilon_t) = C_t P_t^{t-1} C_t' + R.$$

The likelihood function of a Gaussian linear state-space system can be formulated through the innovations and their covariance matrices. If we let Θ denote the unknown parameters of the state-space model, and if we have a total of N observations, the likelihood function looks as follows:

$$L(\Theta | Y_1 \ldots Y_N) = \prod_{i=1}^{N} \left[(2\pi)^m \det(\Sigma_i) \right]^{-1/2} \exp\Big(-\frac{1}{2} \epsilon_i' \Sigma_i^{-1} \epsilon_i \Big)$$

This can be transformed from a product to a sum, using the natural logarithm. We arrive at the following formula for the loglikelihood:

$$\log L(\Theta|Y_1 \dots Y_N) = -\frac{1}{2}\Big[\sum_{i=1}^{N} \log\ \det(\Sigma_i) + \sum_{i=1}^{N} \epsilon_i' \Sigma_i^{-1} \epsilon_i\Big] + \text{constant} \qquad (1)$$

Then we can maximize the loglikelihood with respect to Θ using some numerical optimization algorithm, e.g. quasi-Newton or simplex methods.

4 Nelson-Siegel Models

Since we are only able to observe the yield curve from discrete time market data, we have to try and estimate the entire yield curve from these discrete observations. Several different approaches to this problem has been developed, such as polynomial interpolation, parametric models and regression methods. The challenge lies in trying to make sure that these models can capture the stylized facts of the yield curve (presented in section 2.2) for large amounts of market data, whilst still keeping the model relatively simple and easily interpretable.

Some of the most commonly used parametric models for term structure estimation belong to the Nelson-Siegel family of parametric models, first developed by Nelson and Siegel [16]. They have become popular because they are based on a relatively small number of latent factors, that can be easily interpreted, and are able to capture the stylized facts of the yield curve. In this section we will present some of the different variants of the Nelson-Siegel model, beginning with the three factor version, first developed by Nelson and Siegel.

4.1 Three Factor Nelson-Siegel Model

The original model developed by Nelson and Siegel attempts to capture all of the market data at time t with three latent factors, β_1, β_2 and β_3. The model is derived by first proposing a model for the forward rate curve, which is then transformed into the yield curve, using the relation presented in section 2.1, finally arriving at the following expression:

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

where $\lambda > 0$ is called the decay parameter. This parameter regulates how fast the factor loadings reach 0, with smaller λ resulting in faster movement and bigger λ in slower movement towards the limit.

It is possible to interpret the three different factors in terms of different parts of the curve by studying the asymptotic behavior of the factor loadings. The first factor, β_1, can be interpreted as the long-term factor, because it has a constant factor loading. It is thus unaffected by the value of τ and represents where the curve will level off as τ approaches infinity. We can interpret β_2 as the short-term factor, by first noting that

$$\lim_{\tau \to 0} \frac{1 - e^{-\lambda\tau}}{\lambda\tau} = 1$$

and further that

$$\lim_{\tau\to\infty} \frac{1-e^{-\lambda\tau}}{\lambda\tau} = 0.$$

Thus β_2 controls the shape of the curve at shorter maturities, whilst not affecting the curve for bigger τ. Finally, β_3 can be interpreted as the medium-term factor. We begin by taking the limit as $\tau \to 0$

$$\lim_{\tau\to 0} \frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} = 0$$

and then the limit as $\tau \to \infty$

$$\lim_{\tau\to\infty} \frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} = 0.$$

We notice that β_3 leaves the yield curve unaffected for short and long maturities τ, but instead controls the shape at medium-term maturities. The Nelson-Siegel model is thus able to capture the three standard shapes of the yield curve discussed in section 2.2, through just three different factors. However, it is also possible to extend the model by adding additional factors. We will now have a look at such an extension.

4.2 Four Factor Nelson-Siegel-Svensson Model

This extended version of the Nelson-Siegel model was first proposed by Svensson [22] and adds an additional medium-term factor, giving us the following expression:

$$y(\tau) = \beta_1 + \beta_2(\frac{1-e^{-\lambda_1\tau}}{\lambda_1\tau}) + \beta_3(\frac{1-e^{-\lambda_1\tau}}{\lambda_1\tau} - e^{-\lambda_1\tau}) + \beta_4(\frac{1-e^{-\lambda_2\tau}}{\lambda_2\tau} - e^{-\lambda_2\tau})$$

The purpose of this extension was to the give the model added tractability and allow for better fitting of yield curves with more complicated shapes. The downside of this extension is the addition of an extra decay parameter λ_2, which makes factor estimation substantially more involved.

In addition to extending the original Nelson-Siegel model, it is also possible to restrict it. We will now look at one such proposed restriction, a two factor model.

4.3 Two Factor Nelson-Siegel Model

The two factor model is based on the work done by Litterman and Scheinkman [15], in which they used a variance decomposition technique to show that the first two factors of the three factor Nelson-Siegel model make up almost all of the variance in the yields. The two factor model has the following form:

$$y(\tau) = \beta_1 + \beta_2(\frac{1 - e^{-\lambda\tau}}{\lambda\tau})$$

Using the the two factor model, we could potentially capture the properties of the yield curve just as well as the three factor model does, whilst at the same time simplifying the model greatly, making estimation easier. However, in contrast to the four factor model proposed by Svensson, this comes at the cost of reduced flexibility and ability to model more complex yield curves.

All the Nelson-Siegel models we have introduced so far describes the yield curve cross-sectionally at a fixed time point t. In the following section we will introduce a method that also allows for a time series interpretation of the standard models.

4.4 Dynamic Nelson-Siegel Model

In an effort to find suitable models for forecasting the yield curve, Diebold and Li [9] introduced the following modification of the standard Nelson-Siegel model:

$$y_t(\tau) = L_t + S_t(\frac{1 - e^{-\lambda_t\tau}}{\lambda_t\tau}) + C_t(\frac{1 - e^{-\lambda_t\tau}}{\lambda_t\tau} - e^{-\lambda_t\tau})$$

The principal difference between this dynamic model and the standard model, is that we let the factors L_t, S_t and C_t, as well as the decay parameter λ_t, vary over time. This allows us to forecast yield curves using state-space models and time series. If we assume that the factors vary independently of each other, we model the factors as a AR(1) model. If we instead want the factors to be dependent on each other over time, we will have to use a VAR(1) model.

Diebold and Li would also introduce an interpretation of the model factors, different from the one used in section 4.1. First they noted that the long-term factor L_t can be interpreted as the level of the curve, since

$$\lim_{\tau\to\infty} y_t(\tau) = L_t$$

and that shifting L_t also shifts all of the yields by the same amount across maturities. Then, if you define the slope of the yield curve as

$$\lim_{\tau\to\infty} y_t(\tau) - y_t(\frac{1}{\tau}) = -S_t$$

you can directly identify the short-term factor S_t as a slope factor. Since shifts in the medium-term factor C_t mainly affects the medium-term yields, we can interpret it as the factor controlling the curvature of the yield curve. We thus refer to C_t as the curve factor.

The previously mentioned Nelson-Siegel models perform well from a empirical perspective, but are lacking from a theoretical perspective. For example, Diebold, Piazzesi and Rudebusch [10] showed that the Nelson-Siegel models do not take lack of arbitrage into account, something that is usually assumed in economic theory. There are however ways of rectifying this problem and making the Nelon-Siegel models more theoretically rigorous.

4.5 Arbitrage-Free Nelson-Siegel Model

Trying to combine the empirically well-performing Nelson-Siegel models with more theoretically rigorous models, Christensen, Diebold and Rudebusch [8] would propose the following expression for the yield curve:

$$y_t(\tau) = L_t + S_t(\frac{1-e^{-\lambda_t \tau}}{\lambda_t \tau}) + C_t(\frac{1-e^{-\lambda_t \tau}}{\lambda_t \tau} - e^{-\lambda_t \tau}) - \frac{YA(\tau)}{\tau}$$

where $\frac{YA(\tau)}{\tau}$ is called the yield-adjustment term. This model has the nice property of assuming that the market is arbitrage-free. The model also assumes that the factors L_t, S_t and C_t are modelled in continuous time, by stochastic differential equations. We will not show the derivation of the model, and refer to Christensen, Diebold and Rudebusch for a more detailed discussion of the underlying assumptions. However, to allow for time series modelling, we have to discretize the model and it will thus be necessary to present the underlying stochastic differential equations.

The set-up we use in this thesis for the theory of stochastic calculus is based on the theory presented in Björk [5]. First, we mention the concept of risk-neutral measure. When buying an asset, assuming all investors are risk-neutral, the price an investor would be willing to pay for the asset is the discounted expected future value of the asset. However, in reality the average investor behaves in a risk-averse manner and in an effort to take this behavior into account, we have to use risk-neutral measure. Thus we make a distinction between the real-world probability measure $\mathbb{P}$, which does not adjust for risk-aversion, and the risk-neutral probability measure $\mathbb{Q}$ which does adjust for this.

Next we cursorily mention the concept of filtrations. For a probability triple $(\Omega, \mathcal{F}, \mathbb{P})$ we can think of the σ-algebra $\mathcal{F}$ as containing all the information about an experiment, with the possible outcomes of the experiment given by Ω. Now, if we only want to concern ourselves with the information available on the interval $[0,\ t]$, we can denote this as $\mathcal{F}_t$. We call the collection $\{\mathcal{F}_t\}_{t\geq 0}$ a filtration if for $s < t$ it holds that $\mathcal{F}_s \subseteq \mathcal{F}_t$. Given the probability triple $(\Omega, \mathcal{F}, \mathbb{P})$, we say that the stochastic process $\{X_t; t \geq 0\}$ is adapted to the filtration $\{\mathcal{F}_t\}_{t\geq 0}$, if X_t for all $t \geq 0$ depends only on the information contained in $\mathcal{F}_t$. This is a very informal way of defining filtrations and adapted processes, but it gives a good intuition and the formal definition requires more advanced measure theory beyond the realm of this thesis. We refer to Appendix B in Björk for a more stringent and measure theoretical approach to filtrations.

Next, we need to define the Wiener process (sometimes referred to as Brownian motion).

Definition 4.1. *The continuous time stochastic process $\{W_t; t \geq 0\}$ is called a Wiener process if it satisfies the following properties:*

(i) $W_0 = 0$.

(ii) If $0 \leq r < s \leq t < u$, then $W_u - W_t$ and $W_s - W_r$ are independent.

(iii) If $0 \leq s < t$, then $W_t - W_s \in N(0, t-s)$.

(iv) All realizations of the process are continuous.

To state the definition of a stochastic integral, we first have to determine which processes are considered integrable.

Definition 4.2. *A stochastic process $g = \{g_t\}$ is said to belong to the class $\mathcal{L}^2[a,\ b]$ if the following conditions are met:*

(i) $\int_a^b \mathrm{E}(g_s^2)ds \leq \infty$.

(ii) g is adapted to the filtration $\mathcal{F}_t^W$, where $\mathcal{F}^{\mathcal{W}}$ denotes the σ-algebra generated by the Wiener process $\{W_t; t \geq 0\}$.

Next, we define a stochastic integral for simple processes $g \in \mathcal{L}^2[a,\ b]$.

Definition 4.3. *A stochastic process $g \in \mathcal{L}^2[a,\ b]$ is called simple if there exists a partition $a = t_0 < t_1 \ldots < t_n = b$, such that $g_s = g_{t_k}$ for $s \in [t_k, t_{k+1})$, i.e. there exists a sub-interval where g is piecewise constant. We define the Itô integral of a simple stochastic process g as*

$$\int_a^b g_s dW_s = \sum_{k=0}^{n-1} g_{t_k}(W_{t_{k+1}} - W_{t_k})$$

where $\{W_t; t \geq 0\}$ is a Wiener process.

From this definition, we can give a definition for the Itô integral for general $g \in \mathcal{L}^2[a,\ b]$.

Definition 4.4. *If there exists a sequence of simple stochastic processes $\{g^n \in \mathcal{L}^2[a,\ b]\}$ and a non-simple stochastic process $g \in \mathcal{L}^2[a,\ b]$ such that*

$$\int_a^b \mathrm{E}\left[(g_s^n - g_s)^2\right] ds \to 0$$

as $n \to \infty$, then we define the Itô integral of g as

$$\int_a^b g_s dW_s = \lim_{n\to\infty} \int_a^b g_s^n dW_s.$$

where $\{W_t; t \geq 0\}$ is a Wiener process.

The Itô integral can be further extended in the following way, as per Oksendal [17]:

Definition 4.5. *Let $\{W_t = (W_{1t}, \ldots, W_{nt}); t \geq 0\}$ be an n-dimensional Wiener process and let $v = [v_{ij}]$ be an $m \times n$-matrix, where the $v_{ij} \in \mathcal{L}^2[a,\ b]$ are stochastic processes. We define*

$$\int_a^b v dW_s$$

as the $m \times 1$ column vector whose i'th entry is the sum of Itô integrals

$$\sum_{j=1}^{n} \int_a^b v_{ij} dW_{js}$$

Now we are ready to look at the underlying structures of the arbitrage-free Nelson-Siegel model. The factors $X_t = (L_t, S_t, C_t)'$ are assumed to be given by the following system of SDEs with risk-neutral measure $\mathbb{Q}$:

$$\mathrm{d}X_t = \Lambda(\theta - X_t)\mathrm{d}t + \Sigma \mathrm{d}W_t^Q \tag{2}$$

where

$$\Lambda = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -\lambda & \lambda \\ 0 & 0 & \lambda \end{pmatrix}$$

and θ is column vector of factor means. Σ is referred to as the volatility matrix. The SDE (2) should be interpreted as short-form for the following expression:

$$X_{t+\Delta t} - X_t = \int_t^{t+\Delta t} \Lambda(\theta - X_s)\mathrm{d}s + \int_t^{t+\Delta t} \Sigma \mathrm{d}W_s^Q$$

where the first integral is a standard Riemann (or Lebesgue) integral and the second integral is referred to as an Itô integral with respect to a Wiener process $\{W_t^Q; t \geq 0\}$. This SDE does in turn give rise to a set of ordinary differential equations, whose solutions give us the factor loadings as well as an expression for the yield-adjustment term.

Given the volatility matrix

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

we can express the yield-adjustment term in the following way:

$$\frac{YA(\tau)}{\tau} = \bar{A}\frac{\tau^2}{6} + \bar{B}\Big[\frac{1}{2\lambda^2} - \frac{1}{\lambda^3}\frac{1-e^{-\lambda\tau}}{\tau} + \frac{1}{4\lambda^3}\frac{1-e^{-2\lambda\tau}}{\tau}\Big]$$

$$+\bar{C}\Big[\frac{1}{2\lambda^2} + \frac{1}{\lambda^2}e^{-\lambda\tau} - \frac{1}{4\lambda}\tau e^{-2\lambda\tau} - \frac{3}{4\lambda^2}e^{-2\lambda\tau} - \frac{2}{\lambda^3}\frac{1-e^{-\lambda\tau}}{\tau} + \frac{5}{8\lambda^3}\frac{1-e^{-2\lambda\tau}}{\tau}\Big]$$

$$+\bar{D}\Big[\frac{1}{2\lambda}\tau + \frac{1}{\lambda^2}e^{-\lambda\tau} - \frac{1}{\lambda^3}\frac{1-e^{-\lambda\tau}}{\tau}\Big] + \bar{E}\Big[\frac{3}{\lambda^2}e^{-\lambda\tau} + \frac{1}{2\lambda}\tau + \frac{1}{\lambda}\tau e^{-\lambda\tau} - \frac{3}{\lambda^3}\frac{1-e^{-\lambda\tau}}{\tau}\Big]$$

$$+\bar{F}\Big[\frac{1}{\lambda^2} + \frac{1}{\lambda^2}e^{-\lambda\tau} - \frac{1}{2\lambda^2}e^{-2\lambda\tau} - \frac{3}{\lambda^3}\frac{1-e^{-\lambda\tau}}{\tau} + \frac{3}{4\lambda^3}\frac{1-e^{-2\lambda\tau}}{\tau}\Big]$$

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

5 Term Structure Estimation

We would like to apply the Nelson-Siegel models discussed in section 4 in an effort to estimate the historical term structure of the modern Swedish credit market. To do this in a way that would also allow for yield curve forecasting, we have to link the Nelson-Siegel models with the state-space modelling techniques presented in section 3. Before doing this however, we have to discuss what data would be suitable for such an investigation.

5.1 Data

We are able to estimate the short-term yield curve of the Swedish credit market using STIBOR rates. The STIBOR used to be published by the Swedish central bank [21] up until May 2020, when they instead began to be published by the Swedish Financial Benchmark Facility. The Swedish central bank completely seized publishing the STIBOR rate in July 2020. For more on how the STIBOR rate is calculated and published, we refer to section 2.3. From June 1998 to March 2013, the STIBOR was available for the following maturities: overnight, 1 week, 1 month, 2 months, 3 months, 6 months, 9 months and 12 months. From March 2013 and onwards, the 9 and 12 month maturities seized to be published.

Because of this we chose to apply the models over two intervals, from June 1997 to March 2013, when we have maturities up to 1 year available, and from March 2013 to July 2020, when we only have maturities up to 1/2 year available. We chose to retrieve data only from the Swedish central bank and thus cut-off the second interval at July 2020.

It is worth mentioning that the original data set contained some missing values. To fill these we used linear interpolation, which means that the values used to fill the missing ones lie on the straight line connecting the closest available values.

The STIBOR rates have to be converted to the yields through the relation given in definition 2.3. Once we have done this, we are ready to apply the Nelson-Siegel models to the data. The yields for the two intervals are plotted in figure 2.

Some descriptive statistics for the two intervals 1998 to 2013 and 2013 to 2020 are given in table 1 and table 2 respectively. We notice that the empirical mean yield curves for both intervals seem to have the shape of a normal yield curve, with yields increasing for longer maturities. For the interval 1998 to 2013, the shorter maturity yields are more volatile compared to the longer maturities. We do not see the same tendency for the second interval, where volatility seems to be highest for medium-term yields. It is also worth noting that the 1 week yields for the interval 2013 to 2020 have a minimum yield significantly smaller than for the other maturities. We also notice this graphically by looking at figure 2, where there is a large downward spike in december 2015.

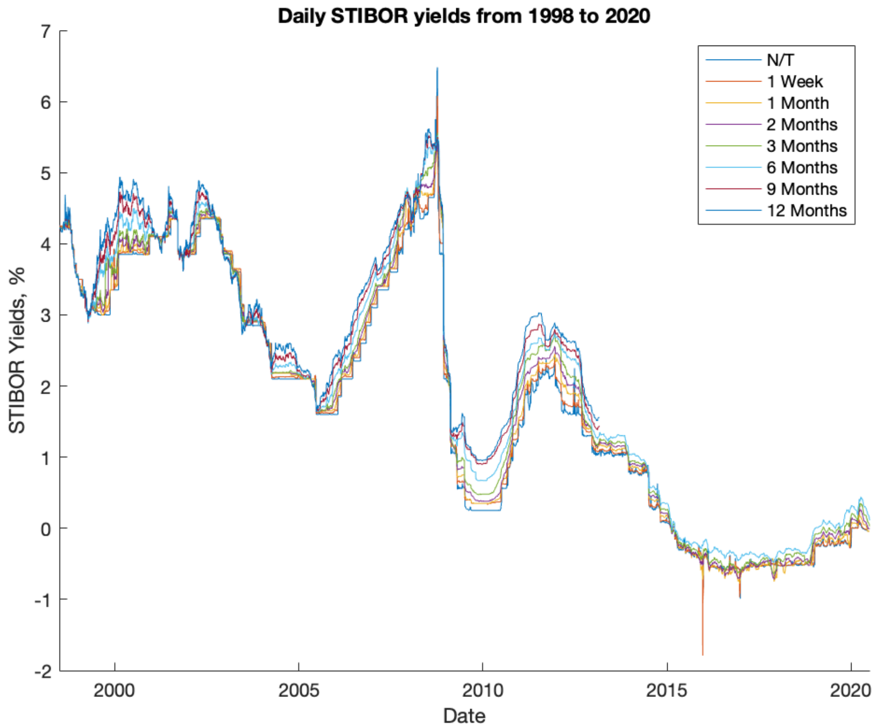


Figure 2: The daily yields for 8 different maturities over the time period June 1998 to July 2020. The 9 and 12 month maturities are not available after March 2013.

Overall, the period of 1998 to 2013 is characterized by a series of economic crises, such as the dot-com bubble around the beginning of the millennium, the subprime crisis of 2008 and the European sovereign debt crisis starting in 2009. In contrast, the period of 2013-2020 is distinguished by relative stability and very low and negative interest rates.

Table 1: Table containing the sample mean, sample standard deviation, maximum and minimum of the yields (%) from 1998 to 2013, for all the available maturities.

Maturity	Mean	Standard dev.	Max.	Min.
Overnight	2.7190	1.2817	6.48	0.25
1 week	2.7636	1.2612	6.08	0.33
1 month	2.8113	1.2606	5.43	0.35
2 months	2.8596	1.2526	5.43	0.38
3 months	2.9190	1.2406	5.56	0.47
6 months	3.0210	1.2119	5.52	0.67
9 months	3.1163	1.1824	5.52	0.90
12 months	3.1989	1.1773	5.62	0.95

Table 2: Table containing the sample mean, sample standard deviation, maximum and minimum of the yields (%) from 2013 to 2020, for all the available maturities.

Maturity	Mean	Standard dev.	Max.	Min.
Overnight	-0.0894	0.5429	1.17	-0.99
1 week	-0.0869	0.5521	1.13	-1.79
1 month	-0.0656	0.5758	1.19	-0.78
2 months	-0.0319	0.5808	1.22	-0.72
3 months	0.0185	0.5777	1.26	-0.65
6 months	0.1190	0.5605	1.35	-0.48

5.2 Model Selection

In an effort to estimate the term structure of the two discussed periods as well as try to predict the future term structure, we will employ a dynamic two factor arbitrage-free Nelson-Siegel model, given by:

$$y_t(\tau) = L_t + S_t\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) - \frac{YA(\tau)}{\tau}$$

where the yield-adjustment term is given by the expression:

$$\frac{YA(\tau)}{\tau} = \bar{A}\frac{\tau^2}{6} + \bar{B}\Big[\frac{1}{2\lambda^2} - \frac{1}{\lambda^3}\frac{1-e^{-\lambda\tau}}{\tau} + \frac{1}{4\lambda^3}\frac{1-e^{-2\lambda\tau}}{\tau}\Big]$$
$$+\bar{D}\Big[\frac{1}{2\lambda}\tau + \frac{1}{\lambda^2}e^{-\lambda\tau} - \frac{1}{\lambda^3}\frac{1-e^{-\lambda\tau}}{\tau}\Big]$$

where $\bar{A} = \sigma_{11}^2 + \sigma_{12}^2$, $\bar{B} = \sigma_{21}^2 + \sigma_{22}^2$, and $\bar{D} = \sigma_{11}\sigma_{21} + \sigma_{12}\sigma_{22}$, for some volatility matrix Σ. Note that this expression for the yield-adjustment term is the same as for the three factor arbitrage-free model, but with $\sigma_{13} = \sigma_{23} = \sigma_{33} = 0$. Also note that we do not let the decay parameter λ vary over time, in contrast to the original formulation of the dynamic model by Diebold and Li [9]. We do this because it significantly simplifies the estimation procedure.

This choice of model is motivated by the fact that the STIBOR rates are only given for maturities less than or equal to a year. This means we are only going to be able to model the short-term yield curve and these curves are unlikely to be of a more complex shape. Thus a two factor model might perform just as well as the three factor model, whilst reducing the number of factors and model parameters needed to be estimated. The arbitrage-free model has been previously shown to perform well both in- and out-sample (for example by Christensen, Diebold and Rudebusch [8]) and also imposes a theoretical rigorousness that is absent from the other Nelson-Siegel models.

Using the dynamic version of the model allows for time series forecasting. However, since the factors $X_t = (L_t, S_t)'$ are modelled in continuous time in the arbitrage-free case, we have to discretize the model before being able to perform any state-space modelling and forecasting. This is accomplished by first transforming the model from risk-neutral measure $\mathbb{Q}$ to the real-world measure $\mathbb{P}$. Christensen, Diebold and Rudebusch showed that the factors are modeled by the following stochastic differential equation, under $\mathbb{P}$-measure:

$$\mathrm{d}X_t = K(\theta^P - X_t)\mathrm{d}t + \Sigma \mathrm{d}W_t^P$$

where K is a 2×2 matrix called the mean-reversion matrix, θ^P is a column vector of factor means under $\mathbb{P}$-measure, and $\{W_t^P; t \geq 0\}$ is the Wiener process under $\mathbb{P}$-measure. Before proceeding, we give a definition for the matrix exponential:

Definition 5.1. *Let X be a $n \times n$ matrix, then the series*

$$I + \sum_{k=1}^{\infty} \frac{1}{k!} X^i$$

is called the matrix exponential and is denoted by e^X.

Next, let

$$\tilde{X}_t = e^{Kt}(\theta^P - X_t)$$

where e^{Kt} refers to the matrix exponential.

Then we get the following:

$$\mathrm{d}\tilde{X}_t = Ke^{Kt}(\theta^P - X_t)\mathrm{d}t + e^{Kt}(-\mathrm{d}X_t)$$

$$= Ke^{Kt}(\theta^P - X_t)\mathrm{d}t - e^{Kt}K(\theta^P - X_t)\mathrm{d}t - e^{Kt}\Sigma \mathrm{d}W_t^P = -e^{Kt}\Sigma \mathrm{d}W_t^P$$

where the first equality follows from Itô's formula, see Proposition 4.2 in Björk [5]. The above equation is short form for:

$$\tilde{X}_{t+\Delta t} - \tilde{X}_t = -\int_t^{t+\Delta t} e^{Ks}\Sigma \mathrm{d}W_s^P,$$

which we can rearrange to get:

$$\tilde{X}_{t+\Delta t} = \tilde{X}_t - \int_t^{t+\Delta t} e^{Ks}\Sigma \mathrm{d}W_s^P$$

$$\Updownarrow$$

$$e^{K(t+\Delta t)}(\theta^P - X_{t+\Delta t}) = e^{Kt}(\theta^P - X_t) - \int_t^{t+\Delta t} e^{Ks}\Sigma \mathrm{d}W_s^P$$

$$\Updownarrow$$

$$\theta^P - X_{t+\Delta t} = e^{-\Delta t K}(\theta^P - X_t) - \int_t^{t+\Delta t} e^{-K(t+\Delta t - s)}\Sigma \mathrm{d}W_s^P$$

$$\Updownarrow$$

$$X_{t+\Delta t} = (I - e^{-\Delta t K})\theta^P + e^{-\Delta t K}X_t + \int_t^{t+\Delta t} e^{-K(t+\Delta t - s)}\Sigma \mathrm{d}W_s^P. \tag{3}$$

Let $w_t := \int_t^{t+\Delta t} e^{-K(t+\Delta t - s)}\Sigma \mathrm{d}W_s^P \in N(0, Q)$, where Q is a 2×2 covariance matrix. The reason why we are able to do this follows from definition 4.5 and the fact that for square integrable functions, the ordinary Itô integrals is normally distributed, see e.g. Baldi [4]. Q is given by

$$Q = \mathrm{Var}(w_t) = \int_t^{t+\Delta t} e^{-K(t+\Delta t - s)}\Sigma\Sigma' e^{-K'(t+\Delta t - s)}\mathrm{d}s$$

$$= \Big[\bar{s} = t + \Delta t - s\Big]$$

$$= \int_0^{\Delta t} e^{-K\bar{s}}\Sigma\Sigma' e^{-K'\bar{s}}\mathrm{d}\bar{s}$$

If we rearrange the 2×2 Q matrix into a 4×1 column vector:

$$\tilde{Q} := \begin{pmatrix} q_{11} \\ q_{12} \\ q_{21} \\ q_{22} \end{pmatrix}$$

and similarly for the matrix $\Sigma\Sigma'$

$$\tilde{\Sigma\Sigma'} := \begin{pmatrix} \bar{A} \\ \bar{D} \\ \bar{D} \\ \bar{B} \end{pmatrix}.$$

Before stating the expression for $\tilde{Q}$, we need to define the Kronecker product:

Definition 5.2. *If A is an $m \times n$-matrix and B is an $p \times q$-matrix, then we define the Kronecker product $\otimes$ as*

$$A \otimes B = \begin{pmatrix} a_{11}B & \cdots & a_{1n}B \\ \cdots & \ddots & \cdots \\ a_{m1}B & \cdots & a_{mn}B \end{pmatrix}$$

i.e. a $pm \times qn$ matrix.

Now, we get the following expression for $\tilde{Q}$:

$$\tilde{Q} = \int_0^{\Delta t} e^{-(I\otimes K + K \otimes I)\bar{s}} \tilde{\Sigma\Sigma'} \mathrm{d}\bar{s}$$

where $\otimes$ refers to the Kronecker product. For more on how this works, see section 10.2 on vector operators for matrices in Petersen and Pedersen [19].To compress some of the equations, we introduce the following notation:

$$\Gamma := I \otimes K + K \otimes I$$

Then, if Γ is not a singular matrix we have the following formula $\int_0^T e^{\Gamma t} dt = \Gamma^{-1}(e^{\Gamma t} - I)$, which can be shown by first noting that the definition of the matrix exponential looks very similar to the Taylor expansion of the ordinary exponential function and then doing a similar proof to the one for integral of an ordinary exponential function using Taylor series. We then arrive at the following:

$$\tilde{Q} = \int_0^{\Delta t} e^{-\Gamma \bar{s}} \tilde{\Sigma\Sigma'} \mathrm{d}\bar{s} = \Gamma^{-1}(I \otimes I - e^{-\Delta t \Gamma}) \tilde{\Sigma\Sigma'} \tag{4}$$

Thus, we have found an expression for Q in terms of the volatility matrix Σ, which will be needed later when calculating the yield-adjustment term. Now, letting

$$A := e^{-\Delta t K} \tag{5}$$

and $\mu := \theta^P$, we can write equation (3) as:

$$X_{t+1} = (I - A)\mu + AX_t + w_t.$$

This is the state equation of our state-space model. The measurement equation is given by:

$$Y_t = B + CX_t + v_t$$

where Y_t is a column vector containing the yields for different maturities, B is a column vector containing the yield-adjustment terms for the corresponding maturities, C is matrix containing the factor loadings and v_t is a zero-mean normal distributed error term with covariance matrix R.

5.3 Parameter Estimation

Before stating the estimation problem, we restate our state-space model

$$X_{t+1} = (I - A)\mu + AX_t + w_t, \qquad w_t \in N(0, Q)$$

$$Y_t = B + CX_t + v_t, \qquad v_t \in N(0, R)$$

which we can also write with the matrices expressed explicitly as

$$\begin{pmatrix} L_{t+1} \\ S_{t+1} \end{pmatrix} = \begin{pmatrix} 1 - a_{11} & -a_{12} \\ -a_{21} & 1 - a_{22} \end{pmatrix} \begin{pmatrix} \mu_L \\ \mu_S \end{pmatrix} + \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} L_t \\ S_t \end{pmatrix} + \begin{pmatrix} w_{t1} \\ w_{t2} \end{pmatrix}$$

$$\begin{pmatrix} y(\tau_1) \\ \vdots \\ y(\tau_n) \end{pmatrix} = \begin{pmatrix} -\frac{YA(\tau_1)}{\tau_1} \\ \vdots \\ -\frac{YA(\tau_n)}{\tau_n} \end{pmatrix} + \begin{pmatrix} 1 & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} \\ \vdots & \vdots \\ 1 & \frac{1-e^{-\lambda\tau_n}}{\lambda\tau_n} \end{pmatrix} \begin{pmatrix} L_t \\ S_t \end{pmatrix} + \begin{pmatrix} v_{t1} \\ \vdots \\ v_{tn} \end{pmatrix}$$

and the covariance matrices are given by

$$Q = \begin{pmatrix} q_{11} & q_{12} \\ q_{12} & q_{22} \end{pmatrix} \qquad R = \begin{pmatrix} r_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & r_n \end{pmatrix}.$$

For the interval 1998 to 2013, we have $n = 8$ and for the interval 2013-2020 we have $n = 6$. Note that we here make an assumption of R being diagonal, just as was done by Christensen, Diebold and Rudebusch [8].

There are several approaches to estimating the state vectors, and we will quickly mention some here. The method originally used by Nelson and Siegel [16] was based on fixing an appropriate value for λ, which in turn makes the measurement equation into a linear equation that allows for use of ordinary least squares. Another commonly used method, e.g. in Christensen, Diebold and Rudebusch, is the Kalman filter, presented in section 3.4. There are also methods based on ridge regression, developed by Annaert et al. [2].

In this thesis we will be using the Kalman filter, since it, from a theoretical perspective, gives the best linear approximations of the state vectors. The Kalman filter however requires that all the parameters of the state space representation be specified. This is not the case in our model and we will thus have to estimate them. To this end we will use the maximum likelihood method presented in section 3.5. The parameters we need to estimate are the state vector $X_0^0 = (L_0, S_0)'$, and covariance matrix $P_0^0 = \begin{pmatrix} p_{11} & p_{12} \\ p_{12} & p_{22} \end{pmatrix}$ used to initialize the Kalman filter, the state transition matrix A, the mean vector μ, the covariance matrix Q, the decay parameter λ and the covariance matrix R. All in all, we end up with $15+n$ parameters to be estimated, given by $\Theta = (L_0, S_0, p_{11}, p_{12}, p_{22}, a_{11}, a_{12}, a_{21}, a_{22}, \mu_L, \mu_S, q_{11}, q_{12}, q_{22}, \lambda, r_1, \cdots, r_n)$, where n depends on the number of different time-to-maturities available.

The maximum likelihood estimate of Θ is achieved by maximizing the expression in equation (1), or alternatively minimizing its additive inverse. (1) is a highly non-linear function and finding the maximum analytically would be very difficult, we thus have to rely on some numerical optimization algorithm. There are several different options we can go for, e.g. the Newton-Raphson method which was recommended by Shumway and Stoffer [20], the Marquardt and Berndt-Hall-Hall-Hausman algorithms used in Aruoba, Diebold and Rudebusch [3] or the Nelder-Mead simplex algorithm used in Christensen, Diebold and Rudebusch [8]. In this thesis we will opt for the Nelder-Mead method, since it does not require knowing or numerically computing the first and second derivatives, and that it is also the one used in Christensen et al., which used the same arbitrage-free Nelson-Siegel model we are using.

To initialize the optimization procedure we need to find a suitable initial parameter Θ_0. If we can fix λ and find an appropriate estimate for Q the measurement equation becomes a linear equation that can be estimated using ordinary least squares. We can thus get OLS estimates for the factors by using the normal equation

$$\hat{X}_t = (C'C)^{-1}C'(Y_t + B).$$

From this we can in turn get estimates of the A, μ, R and a new estimate of Q, which we can use for our Θ_0. If we have a total of N observations and let $\hat{X} := (\hat{X}_1, \ldots, \hat{X}_N)$ and $Y := (Y_1, \ldots, Y_N)$, we have

$$\hat{X} = (C'C)^{-1}C'(Y + B).$$

We can view the normal equation as the orthogonal projection of $Y + B$ onto the space spanned by C. The projection error is then

$$PE := (Y + B) - (C'C)^{-1}C'(Y + B) = (I - (C'C)^{-1}C')(Y + B).$$

We pick the λ that minimizes the trace of the squared projection error

$$\hat{\lambda}_0 = \min_{\lambda} \operatorname{tr}(PE'PE) = \min_{\lambda} \operatorname{tr}((Y+B)'(I-(C'C)^{-1}C')(Y+B))$$

where the second equation follows from the fact that if P is an orthogonal projection matrix, then $I-P$ is also an orthogonal projection matrix and that for any projection we have $P^2 = P$.

To be able to minimize $\operatorname{tr}(PE'PE)$, we need to estimate Q for every step of the optimization procedure. If $\hat{\lambda}_{0t}$ is the value at iteration t of the optimization, we can estimate the factors X by fixing $\lambda = \hat{\lambda}_{0t}$ and using the normal equation

$$\hat{X} = (C'C)^{-1}C'Y$$

ignoring the vector of yield-adjustment terms. Then, by viewing the state equation

$$(X_{t+1} - \mu) = A(X_t - \mu) + w_t$$

as a regression equation and letting $X^{N-1} = (X_1, \dots X_{N-1})$ and $X^N = (X_2 \dots, X_N)$, we can get the maximum likelihood estimate of A

$$\hat{A} = X^{N\prime} X^{N-1}(X^{N-1\prime} X^{N-1})^{-1}.$$

Then we can get the conditional maximum likelihood estimate of the covariance matrix Q by

$$\hat{Q} = (N-1)^{-1} \sum_{k=1}^{N-1} (X_{t+1} - \hat{A}X_t)(X_{t+1} - \hat{A}X_t)'.$$

For more detail on these estimation techniques for VAR(1) models, see Shumway and Stoffer [20]. Once we have $\hat{Q}$ we can evaluate B by using relation (4) and can thus continue the optimization procedure. We reiterate that this has to be done at every step of the optimization.

Once the optimization is finished, we get an estimate $\hat{\lambda}_0$ that we can use to get estimates for A, μ, Q and R and use these for our initial parameters Θ_0. X_0^0 is usually picked as its unconditional expectation, but this is only possible if we know the distribution of X_0^0. In this thesis we simply set X_0^0 as the $\hat{X}_1$ estimated using $\hat{\lambda}_0$. P_0^0 represents the uncertainty in our choice of X_0^0 and for our purposes, we let $P_0^0 = 10^{-1}I$. In turned out that the for the interval 2013-2020 without using the repo rate, setting $P_0^0 = 10^{-1}I$ made the optimization converge to quickly, and so in this case we set $P_0^0 = 5 \cdot 10^{-4}I$ instead. We have now fully specified our initial guess Θ_0 and can perform the actual maximum likelihood estimation for the state space model.

5.4 Repo as an Exogenous Variable

We are also interested in investigating the potential explanatory value that the repo rate has on the LIBOR yields. We would think that the repo rate would serve as some form of floor for the LIBOR yields. To try and illustrate this relation, we have plotted the repo rates together with the overnight LIBOR yields over the entire interval 1998 to 2020, visible in figure 3. We see that the overnight rate is higher than the repo rate with only a few exceptions, particularly during the period after 2015, which is characterized by negative interest rates.

The question is now in what way we would include the repo rate in our model. One possible approach is to directly include the repo rate as an additional exogenous variable in the measurement equation of our state-space model, i.e. modifying the measurement equation to become:

$$Y_t = B + CX_t + Dr_t + v_t.$$

where r_t is the repo rate at time t. We would then estimate D simultaneously with the other variables using the Kalman filter method mentioned in section 5.3. However, this runs the risk of imposing multicollinearity into our model, i.e. we might not be able to distinguish which variables has which effect on the yields.

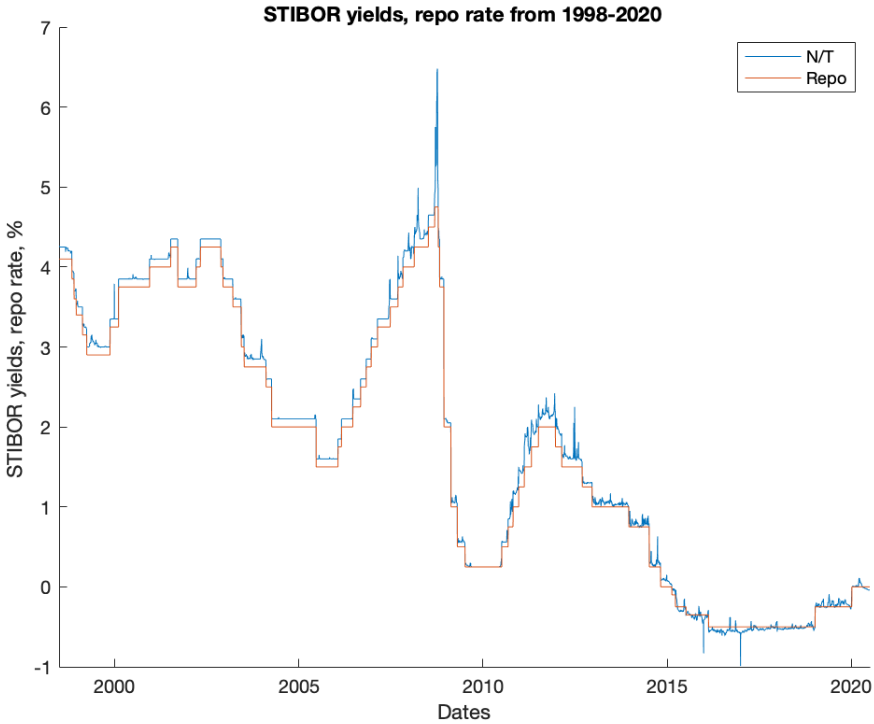


Figure 3: The daily yields for the overnight LIBOR and the repo rate set by the Riksbank over the time period June 1998 to July 2020.

We instead opt for a different tactic, where we postulate that there exists a static relation between the repo and the yields, given by:

$$Y_t = \delta r_t. \tag{6}$$

where δ is a column vector giving the relation between the yields and repo rate. Given the values of Y_t and r_t we can calculate the least square estimator of δ, i.e. the δ that minimizes

$$\sum_{t=1}^{N} (Y_t - \delta r_t)^2,$$

which is given by, see e.g. Anevski [1]

$$\hat{\delta} = \frac{\sum_{t=1}^{N} Y_t r_t}{\sum_{t=1}^{N} r_t^2}.$$

We modify our state-space model by changing the measurement equation to:

$$(Y_t - \hat{\delta} r_t) = B + CX_t + v_t.$$

This will constitute our model for including the repo rate as an exogenous variable.

5.5 Forecasting

For the forecasting of the yield curve, we use the same framework that was set up in Christensen et al. [8]. The one-step prediction for the state factors is given by the conditional expectation of X_{t+1} given the yields up to time t:

$$\mathrm{E}(X_{t+1}|\mathcal{F}_t^Y) = (I - A)\mu + AX_t^t$$

where $\mathcal{F}_t^Y$ is the filtration at t generated by $\{Y_t\}$, i.e. the information available about the yields up to time t. From this, we can get the one-step forecasts of the yields:

$$\mathrm{E}(Y_{t+1}|\mathcal{F}_t^Y) = B + C\,\mathrm{E}(X_{t+1}|\mathcal{F}_t^Y)$$

Since the state space model used here is Gaussian, we are also able to analytically calculate confidence intervals for the predicted yields. We can get the conditional variance of the yields:

$$\mathrm{Var}(Y_{t+1}|\mathcal{F}_t^Y) = C\,\mathrm{Var}(X_{t+1}|\mathcal{F}_t^Y)C' + R$$

where we can find an expression for the conditional variance of X_{t+1}

$$\mathrm{Var}(X_{t+1}|\mathcal{F}_t^Y) = AP_t^tA' + Q.$$

Since the multivariate normal distribution is completely specified by its mean vector and covariance matrix (which we have in an informal way referred to as the variance), and thus we can completely state the conditional distribution of Y_t:

$$\begin{aligned} Y_{t+1}|\mathcal{F}_t^Y &\in N(B + C\,\mathrm{E}(X_{t+1}|\mathcal{F}_t^Y),\, C\,\mathrm{Var}(X_{t+1}|\mathcal{F}_t^Y)C' + R) \\ &= N(B + C(I - A)\mu + CAX_t^t,\, CAP_t^tA'C' + CQC' + R). \end{aligned}$$

From this, we can construct a $1 - \alpha$ confidence interval for the different time-to-maturities τ_k:

$$I_{y_{t+1}(\tau_k),\,1-\alpha} = \left[(B+C(I-A)\mu+CAX_t^t)_k \pm z_{\alpha/2}\sqrt{(CAP_t^tA'C' + CQC' + R)_{kk}}\right]$$

where $z_{\alpha/2}$ is the upper $\frac{\alpha}{2}$ standard normal quantile and the subscripts k and kk refer to the k'th vector and diagonal matrix element, respectively.

These confidence intervals only take uncertainty in the factor estimation and actual prediction into account. There are other forms of uncertainty, such as in the parameter estimation that are not included. Including all possible different sources of uncertainty would make the construction of confidence intervals substantially more complicated, and so for our purposes we assume that they are negligible.

6 Results

This section will devoted to showing the results of the estimation procedure presented in section 5. We begin by looking at the estimated parameters.

6.1 Estimation Results

The initial estimate of the decay parameter $\hat{\lambda}_0$, the final ML estimate of the factor means μ and decay parameter $\hat{\lambda}$ as well as the value of the maximized log-likelihood function are available in table 3, for the two different intervals, with and without repo as an exogenous variable.

Table 3: Table containing initial λ estimate, the final estimated $\mu = \theta^P$ and λ, as well as the actual log-likelihood. μ is estimated using the yields in decimal form.

1998-2013 without repo

μ_L	μ_S	$\hat{\lambda}_0$	$\hat{\lambda}$	Log-likelihood
-0.0164	-0.0183	2.5620	2.0269	180763

1998-2013 with repo

μ_L	μ_S	$\hat{\lambda}_0$	$\hat{\lambda}$	Log-likelihood
0.0039	-0.0034	2.7396	2.1150	178893

2013-2020 without repo

μ_L	μ_S	$\hat{\lambda}_0$	$\hat{\lambda}$	Log-likelihood
0.0079	-0.0147	1.2444	0.5549	71230

2013-2020 with repo

μ_L	μ_S	$\hat{\lambda}_0$	$\hat{\lambda}$	Log-likelihood
0.0054	-0.0073	1.7038	0.8171	72312

We have also chosen to include some descriptive statistics of the estimated factors, available in table 4, as well as time series plots of the factors, visible in figure 4 and 5.

In the estimation procedure we have used, we do not estimate the mean reversion matrix K and volatility matrix Σ directly. Instead, we have to transform our estimates of the state matrix A and state noise covariance matrix Q through the relations (5) and (4) respectively. The directly estimated state matrices A, as well as the corresponding mean reversion matrices K are available in table 5. The directly estimated covariance matrices Q and corresponding volatility matrix Σ are available in table 6.

Table 4: Table containing descriptive statistics for the estimated level and slope factors for the different intervals, with and without repo. The factors are estimated using the yields in decimal form.

1998-2013 without repo

Factor	Mean	Standard dev.	Max.	Min.
L	0.0350	0.0117	0.0657	0.0141
S	-0.0075	0.0067	0.0091	-0.0255

1998-2013 with repo

Factor	Mean	Standard dev.	Max.	Min.
L	0.0027	0.0083	0.0189	-0.0176
S	-0.0025	0.0079	0.0148	-0.0188

2013-2020 without repo

Factor	Mean	Standard dev.	Max.	Min.
L	0.0163	0.0081	0.0385	-0.0055
S	-0.0173	0.0058	-0.0001	-0.0408

2013-2020 with repo

Factor	Mean	Standard dev.	Max.	Min.
L	0.0119	0.0045	0.0273	0.0013
S	-0.0119	0.0045	-0.0011	-0.0269

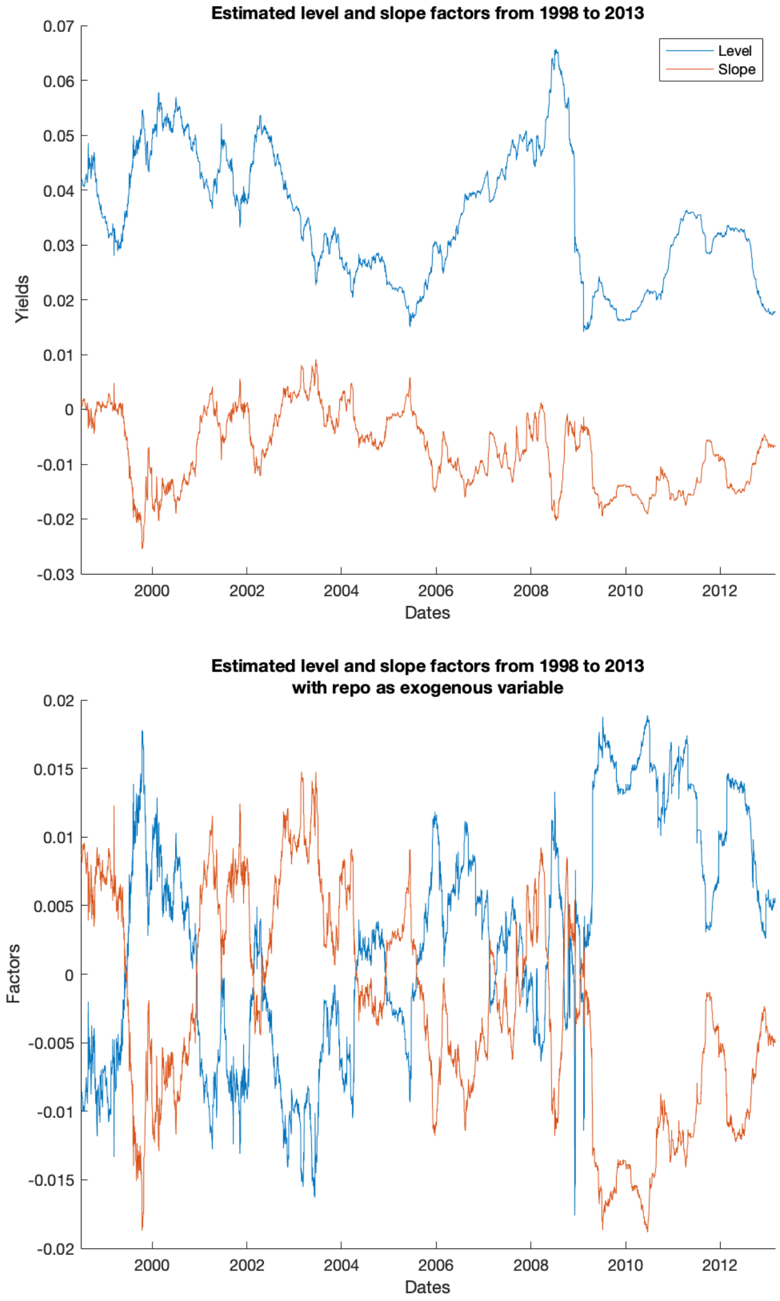
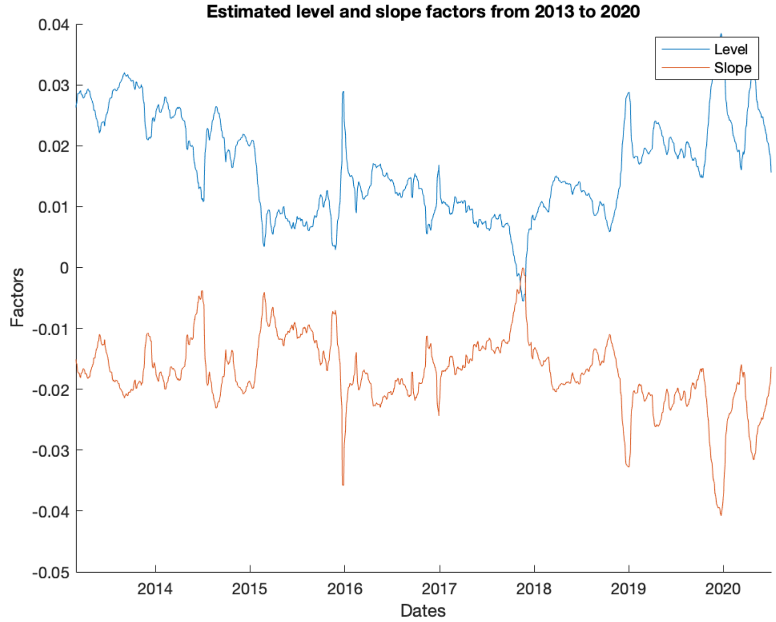


Figure 4: The estimated level and slope factors for the interval 1998 to 2013, with and without repo. The factors are given using the yields in percentage.



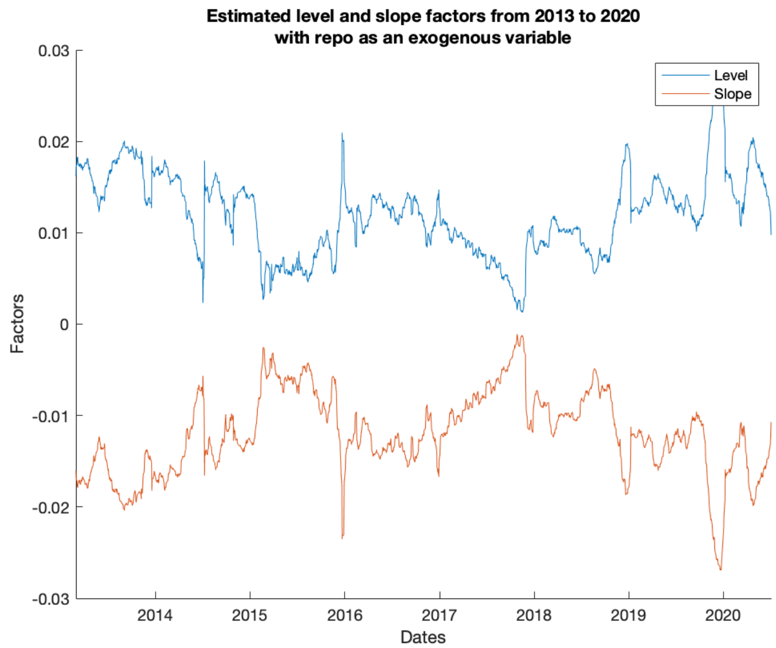


Figure 5: The estimated level and slope factors for the interval 2013 to 2020, with and without repo. The factors are given using the yields in percentage.

Table 5: Table containing the estimated state matrix A and the corresponding mean-reversion matrix K for the intervals, with and without repo.

1998-2013 without repo

A	$a_{.1}$	$a_{.2}$	K	$k_{.1}$	$k_{.2}$
$a_{1.}$	1.0003	-0.0022	$k_{1.}$	-0.0831	0.5574
$a_{2.}$	0.0006	0.9971	$k_{2.}$	-0.1406	0.7238

1998-2013 with repo

A	$a_{.1}$	$a_{.2}$	K	$k_{.1}$	$k_{.2}$
$a_{1.}$	0.8391	-0.1618	$k_{1.}$	42.6932	42.9747
$a_{2.}$	0.0621	1.0602	$k_{2.}$	-16.4818	-16.0435

2013-2020 without repo

A	$a_{.1}$	$a_{.2}$	K	$k_{.1}$	$k_{.2}$
$a_{1.}$	1.0026	0.0172	$k_{1.}$	-0.6752	-4.3719
$a_{2.}$	-0.0048	0.9765	$k_{2.}$	1.2117	5.9837

2013-2020 with repo

A	$a_{.1}$	$a_{.2}$	K	$k_{.1}$	$k_{.2}$
$a_{1.}$	1.0197	0.0287	$k_{1.}$	-5.0389	-7.3145
$a_{2.}$	-0.0338	0.9558	$k_{2.}$	8.6323	11.2606

Table 6: Table containing the estimated noise matrix Q and the corresponding volatility matrix Σ for the different intervals, with and without repo.

1998-2013 without repo

Q	$q_{.1}$	$q_{.2}$	Σ	$\sigma_{.1}$	$\sigma_{.2}$
$q_{1.}$	0.33×10^{-6}	-0.26×10^{-6}	$\sigma_{1.}$	0.0091	-0.0073
$q_{2.}$	-0.26×10^{-6}	0.29×10^{-6}	$\sigma_{2.}$	-0.0073	0.0046

1998-2013 with repo

Q	$q_{.1}$	$q_{.2}$	Σ	$\sigma_{.1}$	$\sigma_{.2}$
$q_{1.}$	0.93×10^{-6}	-0.50×10^{-6}	$\sigma_{1.}$	0.0159	-0.0083
$q_{2.}$	-0.50×10^{-6}	0.40×10^{-6}	$\sigma_{2.}$	-0.0083	0.0057

2013-2020 without repo

Q	$q_{.1}$	$q_{.2}$	Σ	$\sigma_{.1}$	$\sigma_{.2}$
$q_{1.}$	0.71×10^{-6}	-0.76×10^{-6}	$\sigma_{1.}$	0.0135	-0.0146
$q_{2.}$	-0.76×10^{-6}	0.84×10^{-6}	$\sigma_{2.}$	-0.0146	0.0017

2013-2020 with repo

Q	$q_{.1}$	$q_{.2}$	Σ	$\sigma_{.1}$	$\sigma_{.2}$
$q_{1.}$	0.46×10^{-6}	-0.37×10^{-6}	$\sigma_{1.}$	0.0108	-0.0087
$q_{2.}$	-0.37×10^{-6}	0.45×10^{-6}	$\sigma_{2.}$	-0.0087	0.0035

6.2 In-Sample Fit

In addition to the parameter estimates, we are also interested in how well the estimated model fits the actual yields. To do this, we first look at the residuals, given by

$$e_t^k = y_t(\tau_k) - \hat{y}_t(\tau_k)$$

where $\hat{y}$ is the estimated yield. Since our data contains some quite extreme outliers, we want to use an error measurement that is not affected by these. To this end, we use the following definition:

Definition 6.1. *For a data set $x_1, \ldots, x_n$, let $\bar{x}$ denote the median of the data set. Then we define the median absolute deviation as*

$$MAD = median(|x_i - \bar{x}|).$$

Some descriptive statistics for the residuals for the interval 1998 to 2013 and 2013 to 2020 are given in table 7 and 8 respectively. In addition, to illustrate how the residuals vary over time, we have included time series plots of the 3-month residuals over the different intervals in figure 6 and 7.

Table 7: Table containing mean, standard deviation and median absolute deviation of the residuals, for the the interval 1998 to 2013, with and without repo. The residuals are given in percentage points.

1998-2013 without repo

Maturities	Mean	Standard dev.	MAD
Overnight	-0.0345	0.1270	0.0379
1 week	-0.0013	0.0853	0.0306
1 month	0.0007	0.0062	0.0026
2 months	-0.0047	0.0446	0.0175
3 months	0.0064	0.0659	0.0306
6 months	-0.0089	0.0455	0.0234
9 months	0.0004	0.0043	0.0020
12 months	0.0188	0.0524	0.0281

1998-2013 with repo

Maturities	Mean	Standard dev.	MAD
Overnight	-0.0075	0.1240	0.0402
1 week	0.0052	0.0833	0.0311
1 month	0.0003	0.0186	0.0086
2 months	0.0011	0.0435	0.0196
3 months	-0.0015	0.0643	0.0306
6 months	-0.0060	0.0490	0.0268
9 months	-0.0002	0.0168	0.0081
12 months	0.0016	0.0579	0.0320

Table 8: Table containing mean, standard deviation and median absolute deviation of the residuals, for the interval 2013 to 2020, with and without repo. The residuals are given in percentage points.

2013-2020 without repo

Maturities	Mean	Standard dev.	MAD
Overnight	0.0087	0.0537	0.0301
1 week	0.0038	0.0492	0.0224
1 month	-0.0051	0.0267	0.0124
2 months	-0.0096	0.0275	0.0152
3 months	0.0038	0.0231	0.0135
6 months	-0.0003	0.0250	0.0152

2013-2020 with repo

Maturities	Mean	Standard dev.	MAD
Overnight	0.0080	0.0506	0.0271
1 week	0.0044	0.0494	0.0237
1 month	-0.0015	0.0264	0.0119
2 months	-0.0063	0.0324	0.0181
3 months	0.0055	0.0341	0.0207
6 months	-0.0001	0.0078	0.0033

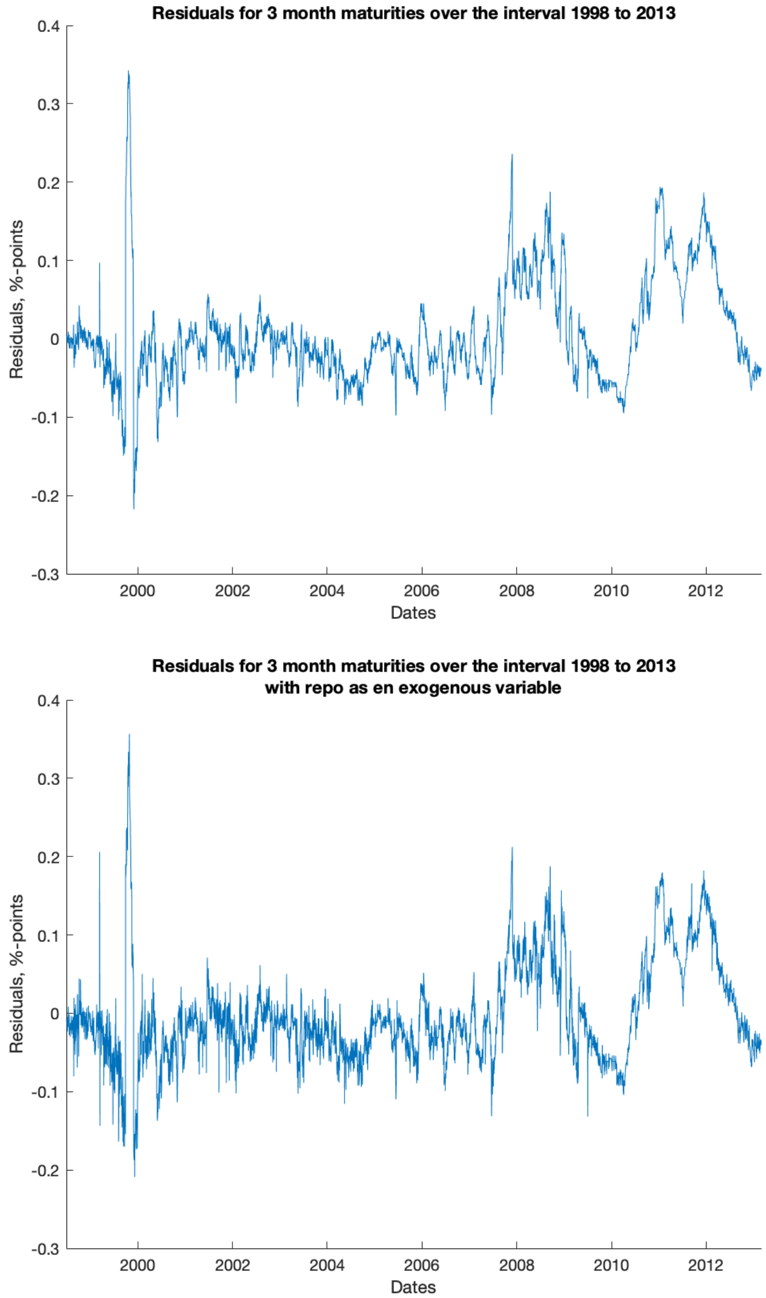


Figure 6: The residuals for the 3 month maturities over the interval 1998 to 2013, with and without repo. The factors are given using the yields in percentage.

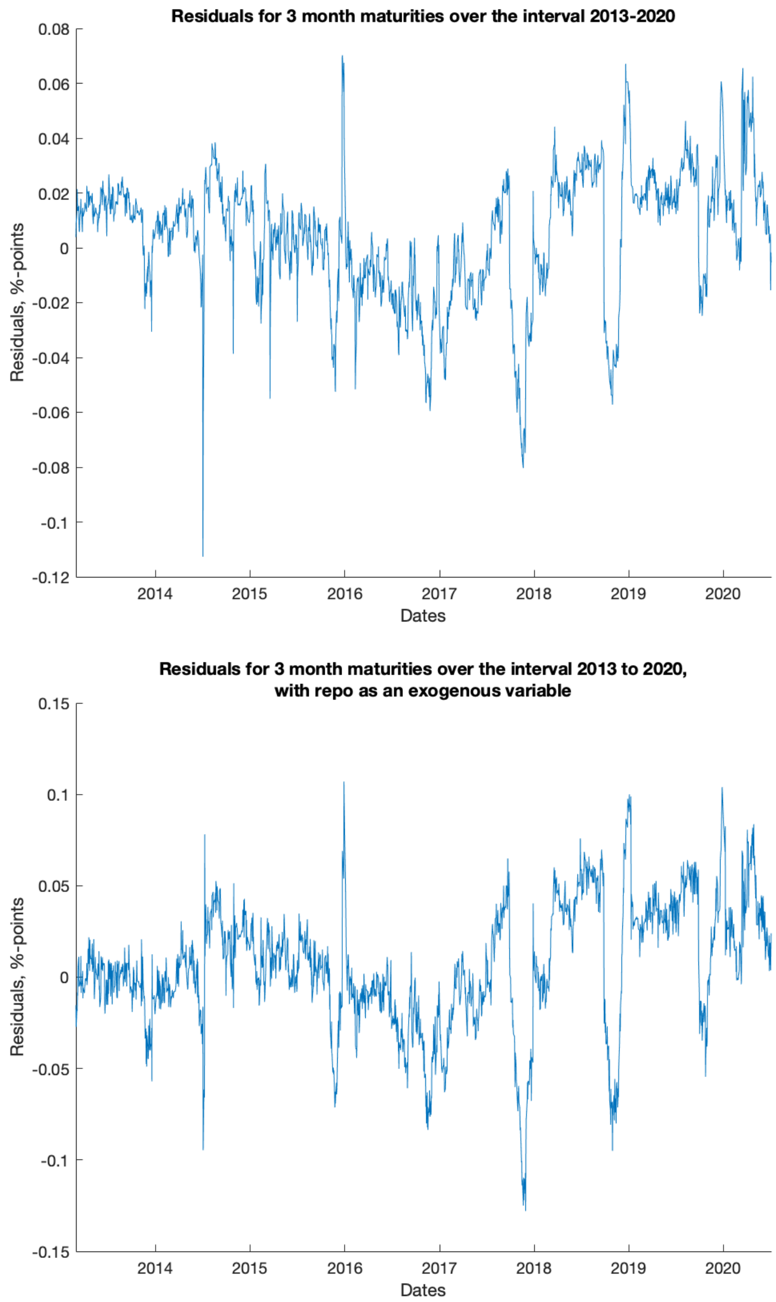


Figure 7: The residuals for the 3 month maturities over the interval 2013 to 2020, with and without repo. The factors are given using the yields in percentage.

6.3 In-Sample Forecasting

We would also like to test the forecasting capabilities of the model. We do this by using the one-step prediction setup described in 5.5, over the two different intervals. Some descriptive statistics for the prediction residuals for interval 1998 to 2013 and 2013 to 2020 are available in table 9 and 10 respectively.

In addition, we also want to make us of the confidence intervals for the predictions, also described in 5.5. First, we look at the ratio of actual yields contained in the 95%-confidence intervals. These ratios are available in table 11, for both intervals, with and without repo as an exogenous variable.

We have also included some plot of the confidence intervals, as they vary over time, available in figures 8, 9, 10 and 11. The plots, do not only contain the confidence intervals, but also the actual yields. These yields are color-coded in the following way: green yields are contained within the confidence interval, red yields are outside the confidence interval.

Table 9: Table containing mean, standard deviation and median absolute deviation of the prediction residuals, for the the interval 1998 to 2013, with and without repo. The residuals are given in percentage points.

1998-2013 without repo

Maturities	Mean	Standard dev.	MAD
Overnight	-0.0344	0.1254	0.0366
1 week	-0.0013	0.0895	0.0299
1 month	0.0007	0.0325	0.0061
2 months	-0.0047	0.0542	0.0196
3 months	0.0064	0.0719	0.0327
6 months	-0.0089	0.0539	0.0257
9 months	-0.0004	0.0340	0.0088
12 months	0.0189	0.0637	0.0331

1998-2013 with repo

Maturities	Mean	Standard dev.	MAD
Overnight	-0.0103	0.1178	0.0351
1 week	0.0025	0.0891	0.290
1 month	-0.0026	0.0611	0.0129
2 months	-0.0019	0.0757	0.0259
3 months	-0.0045	0.0929	0.0362
6 months	-0.0093	0.0812	0.0322
9 months	-0.0037	0.0648	0.0210
12 months	-0.0020	0.0848	0.0412

Table 10: Table containing mean, standard deviation and median absolute deviation of the prediction residuals, for the interval 2013 to 2020, with and without repo. The residuals are given in percentage points.

2013-2020 without repo			
Maturities	Mean	Standard dev.	MAD
Overnight	0.0082	0.0538	0.0303
1 week	0.0033	0.0526	0.0229
1 month	-0.0056	0.0314	0.0140
2 months	-0.0100	0.0309	0.0167
3 months	0.0034	0.0261	0.0147
6 months	-0.0006	0.0278	0.0171

2013-2020 with repo			
Maturities	Mean	Standard dev.	MAD
Overnight	0.0101	0.0520	0.0267
1 week	0.0065	0.0534	0.0234
1 month	0.0005	0.0350	0.0135
2 months	-0.0043	0.0394	0.0189
3 months	0.0074	0.0399	0.0221
6 months	0.0017	0.0210	0.0054

Table 11: Table containing the portion of actual yields that were contained inside the confidence intervals of the predicted values, for overnight and 12 month maturities for the interval 1998 to 2013, 6 month maturities for the interval 2013 to 2020.

Maturity	1998-2013, without repo	1998-2013, with repo	2013-2020, without repo	2013-2020, with repo
Overnight	0.9528	0.9683	0.9330	0.9488
12 month (6 month)	0.9753	0.9767	0.9744	0.9869

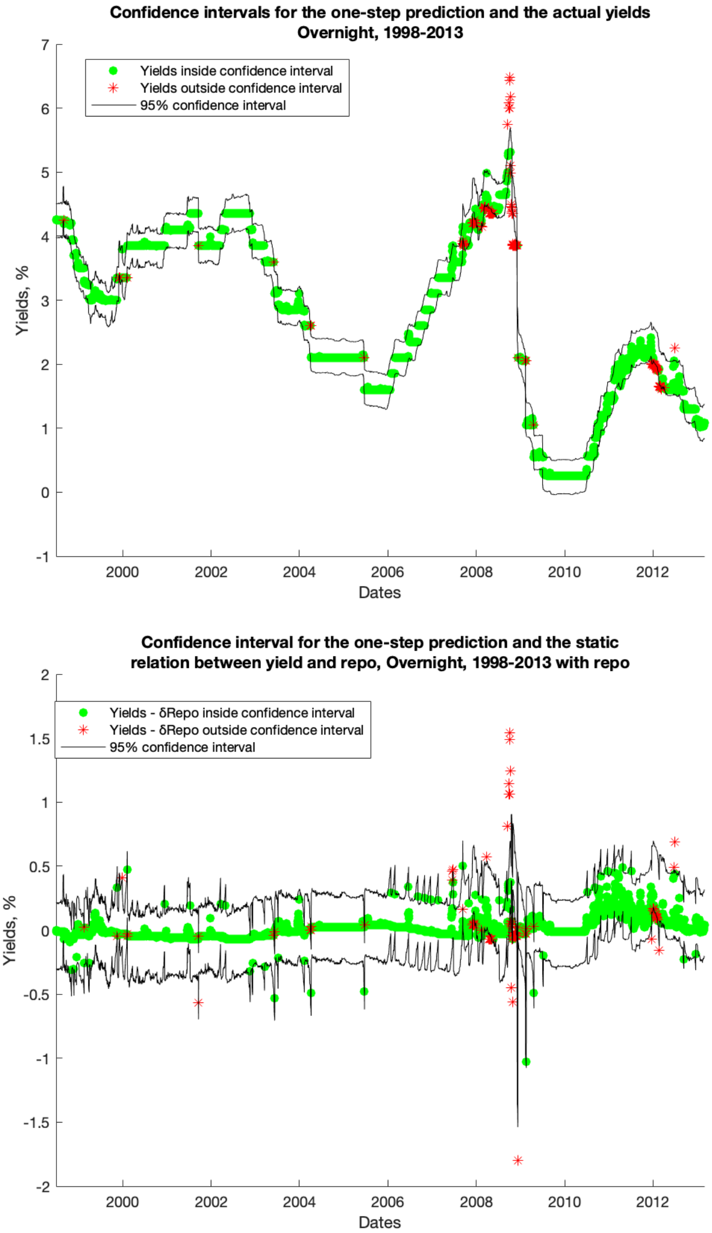


Figure 8: Plots containing the 95%-confidence intervals for the one-step predictions for the overnight yields as well as the actual overnight yields, y_t for the upper plot, and the static difference between the overnight yields and repo, $y_t - \delta r_t$, over the interval 1998 to 2013. The black lines are the confidence interval over time. The green dots signify yields inside the confidence interval. The red asterisks signify yields outside the confidence interval. Top: without repo. Bottom: with repo yields.

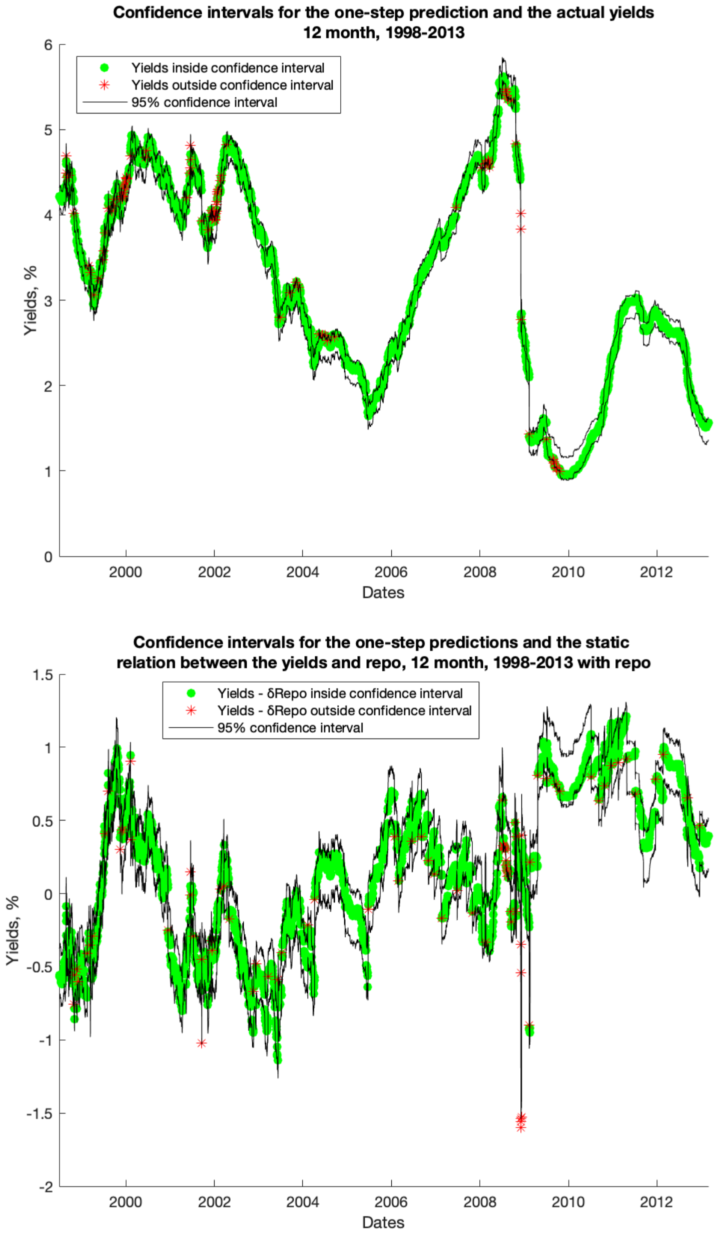


Figure 9: Plots containing the 95%-confidence intervals for the one-step predictions for the overnight yields as well as the actual 12 month yields, y_t for the upper plot, and the static difference between the 12 month yields and repo, $y_t - \delta r_t$, over the interval 1998 to 2013. The black lines are the confidence interval over time. The green dots signify yields inside the confidence interval. The red asterisks signify yields outside the confidence interval. Top: without repo. Bottom: with repo yields.

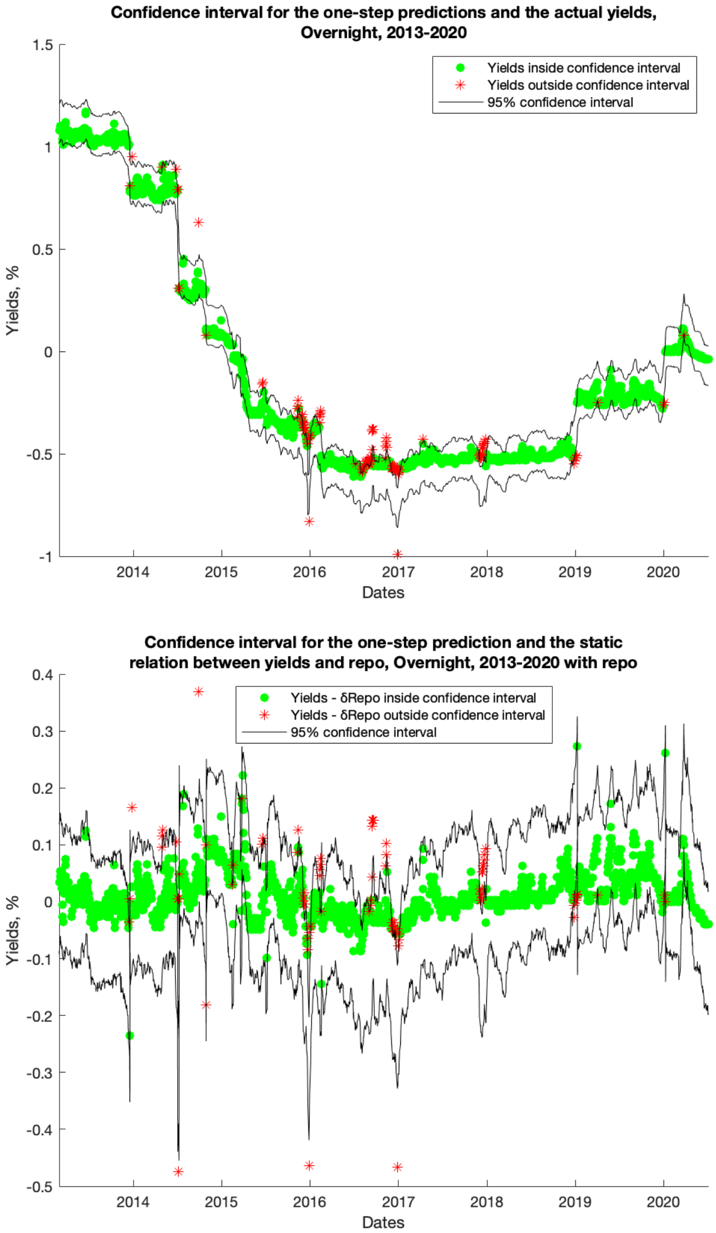
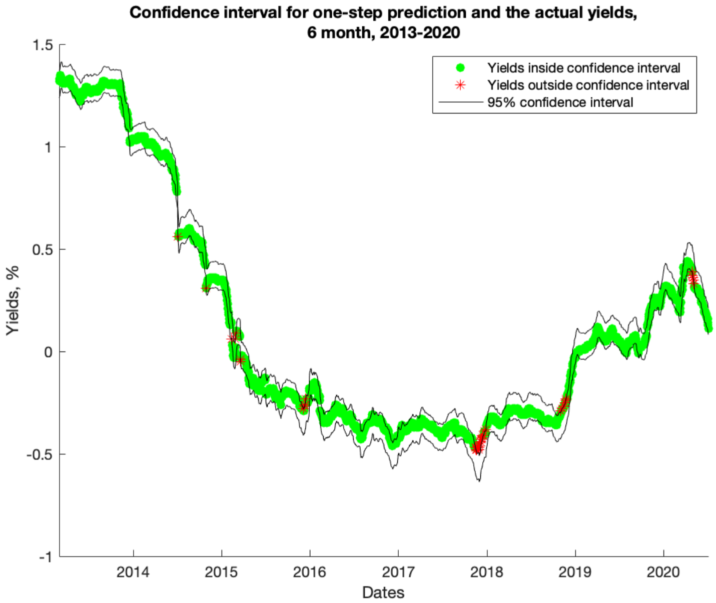


Figure 10: Plots containing the 95%-confidence intervals for the one-step predictions for the overnight yields as well as the actual overnight yields, y_t for the upper plot, and the static difference between the overnight yields and repo, $y_t - \delta r_t$, over the interval 2013 to 2020. The black lines are the confidence interval over time. The green dots signify yields inside the confidence interval. The red asterisks signify yields outside the confidence interval. Top: without repo. Bottom: with repo yields.



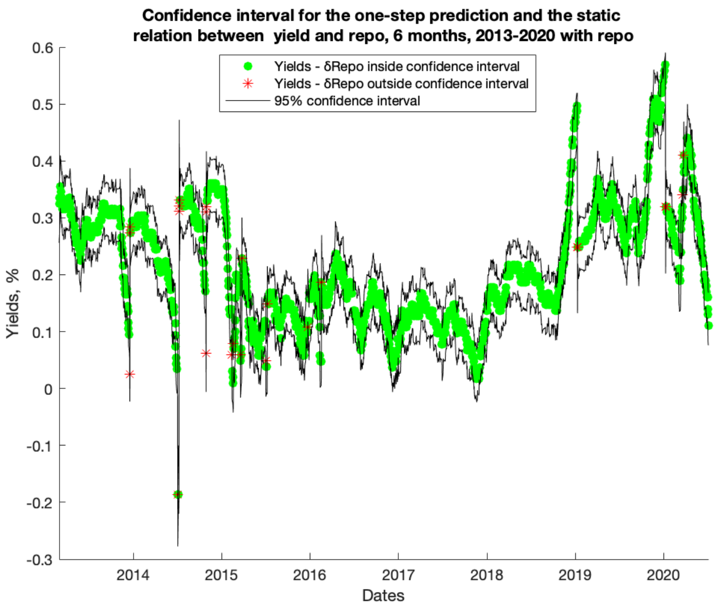


Figure 11: Plots containing the 95%-confidence intervals for the one-step predictions for the overnight yields as well as the actual 6 month yields, y_t for the upper plot, and the static difference between the 6 month yields and repo, $y_t - \delta r_t$, over the interval 2013 to 2020. The black lines are the confidence interval over time. The green dots signify yields inside the confidence interval. The red asterisks signify yields outside the confidence interval. Top: without repo. Bottom: with repo yields.

7 Discussion

In this section we intend to analyze, interpret and discuss the results presented in section 6, as well as suggesting some further areas of research that relates to the topic at hand.

Firstly, when looking at the state estimates for the different intervals, we notice that the inclusion of the repo rate as an exogenous variable seems to have a much greater effect on the state estimates from 1998 to 2013 than for 2013 to 2020. As we see in figures 4 and 5, the difference in the shape of the time series when we include the repo compared to when we do not include it is stark for the interval 1998 to 2013, whilst the difference is much smaller for the second interval. This is probably because the time series of yields seems much more prone to non-stationarity during the first interval compared to the second. We would thus expect that transforming the data using relation (6) would have a much bigger effect on the factor estimates from 1998 to 2013. The second interval from 2013 to 2020 only shows non-stationary behavior during the first part of the interval, from about 2013 to 2015, when the yields are decreasing rapidly. This is also the period where we observe the biggest difference between not including and including the repo rate in figures 4 and 5.

Another observation that we make when looking at the factor estimates is that for the interval 2013 to 2020 without repo, as well as both intervals when we include repo, the slope and level factors seem to be strongly negatively correlated. The Nelson-Siegel model we are using can be expressed as:

$$L_t + S_t \frac{1 - e^{-\lambda\tau}}{\lambda\tau} + \frac{YA(\tau)}{\tau} = y_t(\tau).$$

We also have:

$$\lim_{\tau\to\infty} \frac{1 - e^{-\lambda\tau}}{\lambda\tau} = 1.$$

In addition, it turns out that the yield-adjustment term is quite small. Thus, at least for for small τ, we can approximate the model by:

$$L_t + S_t \frac{1 - e^{-\lambda\tau}}{\lambda\tau} + \frac{YA(\tau)}{\tau} \approx L_t + S_t.$$

Then, we also have that from 2015 onwards the yields are roughly constant and close to zero. This is also true when we subtract the repo from the yields using the static relation, especially for the shorter maturities. Thus we get:

$$L_t + S_t \frac{1 - e^{-\lambda\tau}}{\lambda\tau} + \frac{YA(\tau)}{\tau} \approx L_t + S_t \approx 0$$

$$\Updownarrow$$

$$L_t \approx -S_t.$$

From this, it makes sense that the estimated factors show this strong negative correlation during the mentioned intervals.

When it comes to the state and mean reversion matrices presented in table 5, we observe that the A matrices are all quite close to the identity matrix, which would indicate that the the state equation has a long memory. This means that even if we take two points on the time series that lie far apart, they will still show a high degree of dependence on one another. We also notice, that the inclusion of the repo rate in the model has a substantial effect on the state matrices for the first interval from 1998 to 2013, whilst the two state matrices for the second interval are much closer to each other. Once again, this points to that the inclusion of the repo rate would have a much bigger effect on the first, more volatile interval compared to the second, more stable interval.

When it comes to the in-sample fit, visible in tables 7 and 10, the models seem to perform well overall. We notice that including the repo rate improves the model performance for the shorter maturities, in particular the overnight rates and the 1 month rates. The answer is more ambiguous for the 1 week rates, as well as 2 months and higher. In these cases, the mean residuals seems to decrease in general, whilst the variability increases in some cases, and decreases in others. Also the mean absolute deviation seems to increase in general when using the repo as an exogenous variable, with a few exceptions. It seems like the inclusion of the repo is mainly beneficial for the overnight rates, since these are the only maturities that see a consistent decrease in both mean and variability for the residuals. This lines up with the theory presented in section 2.4.

From figures 6 and 7, we gather that the models have a hard time capturing the term structure during periods of economic shocks. There are clear and noticable spikes in the residuals around 2000, 2008-2009 and 2011-2012. All of these are period characterised by economic crises, such as the dot-com bubble, the great recession and the beginning of the Eurocrisis, respectively.

Next, we look at forecasting. Overall the in-sample one-step forecasts perform quite well, and the prediction residuals, available in table 9 and 10, are close to the actual model residuals discussed above. The main difference seems to be in the variability, with the prediction residuals generally showing a higher standard and median absolute deviation than the model residuals. When it comes to the difference between excluding and including repo, we again notice that the results are inconclusive, with mean residuals decreasing for some maturities, whilst decreasing for others. In general, the variability of the prediction residuals seem to increase, the main exception being for the overnight rates. We see a significant decrease for the overnight mean residuals when we include repo for the first interval, and an increase over the second interval.

When looking at table 11, i.e. the proportion of actual yields contained inside the confidence intervals, we notice that the percentage is around 95% of the yields being inside. However, the proportion is consistently higher for the longer maturities, which could mean that we are overestimating the uncertainty for these maturities. Also, including the repo seems to lead to a higher proportion of yields being contained inside the confidence bounds.

Investigating the plots in figures 8, 9, 10 and 11, the yields not contained in the confidence intervals seem to be clustered during specific time periods, in particular during the great recession in 2008-2009, as well as the period after 2015, which is characterised by very low or interest rates. This again could point to the model performing worse during periods of economic uncertainty, as well as the predictions performing worse for very low rates.

To summarize, the two-factor arbitrage-free Nelson-Siegel model seems to perform quite well in-sample and generally is able to capture the term structure for most of the data set. The main exception is during periods of unstable economic times, when the model performs significantly worse. The results of including the repo rate are inconclusive, but generally seems to improve the model performance for the overnight rates.

7.1 Further Studies

Since the inclusion of the repo rates showed to be inconclusive, it could point to our choice of modelling a static relation between the yields and the repo to be flawed. One could potentially try to estimate the relation between the yields and the repo simultaneously with the other parameters during the maximum likelihood optimization. Doing this could however lead to problems with multicollinearity, where we are unable to determine which variables are having what effect.

As we observed earlier, inclusion of the repo rate seems to have a positive effect on the predictive power of the model, but only for shorter maturities. This lines up with the fact that the Riksbank mainly uses the repo to influence the short term rates, whilst it relies on publishing its own economic forecast to influence the longer term rates. In an effort to take this into account, one might want to include these economic prognostication that the Riksbank makes on a regularly basis in the model. We would then run into problems of both how to quantify the contents of these forecasts and how to collect them into a useable data set.

8 References

Books and Articles

[1] D. Anevski. *A Concise Introduction to Mathematical Statistics*. 1st ed. Studentlitteratur, 2017.

[2] J. Annaert et al. "Estimating the Yield Curve Using the Nelson-Siegel Model: A Ridge Regression Approach". In: *International Review of Economics & Finance, Forthcoming* (2012).

[3] B. Aruoba, F. Diebold, and G. Rudebusch. "The macroeconomy and the yield curve: a dynamic latent factor approach". In: *Journal of Econometrics* 131.1-2 (2006), pp. 309–338.

[4] P. Baldi. *Stochastic Calculus*. 1st ed. Springer, 2017.

[5] T. Björk. *Arbitrage Theory in Continuous Time*. 4th ed. Oxford Finance Series. Oxford University Press, 2019.

[6] H. Byström. *Finance*. 4th ed. Studentlitteratur, 2020.

[7] M. Choudhry. *Bonds: A Concise Guide for Investors*. 1st ed. Palgrave Macmillan, 2006.

[8] J. Christensen, F. Diebold, and G. Rudebusch. "The affine arbitrage-free class of Nelson-Siegel term structure models". In: *Journal of Econometrics* 164.1 (2011), pp. 4–20.

[9] F. Diebold and C. Li. "Forecasting the term structure of government bond yields". In: *Journal of Econometrics* 30.2 (2006), pp. 337–364.

[10] F. Diebold, M. Piazzesi, and G. Rudebusch. "Modeling Bond Yields in Finance and Macroeconomics". In: *American Economic Review* 95.2 (2005), pp. 415–420.

[11] K. Fregert and L. Jonung. *Makroekonomi: Teori, politik och institutioner*. 5th ed. Studentlitteratur, 2018.

[12] A. Jakobsson. *An Introduction to Time Series Modeling*. 3rd ed. Studentlitteratur, 2019.

[14] G. Lindgren, H Rootzén, and M. Sandsten. *Stationary Stochastic Processes for Scientists and Engineers*. 1st ed. CRC Press, 2013.

[15] R. Litterman and J. Scheinkman. "Common Factors Affecting Bond Returns". In: *The Journal of Fixed Income* 1 (1991), pp. 54–61.

[16] C. Nelson and A. Siegel. "Parsimoneous Modeling of Yield Curves for U.S. Treasury Bills". In: *Journal of Business* 60.3 (1987), pp. 473–489.

[17] B. Oksendal. *Stochastic Differential Equations*. 6th ed. Springer, 2003.

[20] R. Shumway and D. Stoffer. *Time Series Analysis and Its Applications*. 2nd ed. Springer Texts in Statistics. Springer, 2006.

[22] L. Svensson. "Estimating and Interpreting Forward Interest Rates: Sweden 1992-1994". In: *Sveriges Riksbankds Quarterly Review* 3 (1995), pp. 13–26.

Websites

[13] *Konsumentprisindex (KPI)*. May 6, 2021. URL: `https://www.scb.se/hitta-statistik/statistik-efter-amne/priser-och-konsumtion/konsumentprisindex/konsumentprisindex-kpi/`.

[18] *Penningpolitik*. May 6, 2021. URL: `https://www.riksbank.se/sv/penningpolitik/`.

[19] K. Petersen and M. Pedersen. *The Matrix Cookbook*. Version 2012-11-15. June 7, 2021. URL: `https://www.math.uwaterloo.ca/~hwolkowi/matrixcookbook.pd`.

[21] *Svenska marknadsräntor*. May 6, 2021. URL: `https://www.riksbank.se/sv/statistik/sok-rantor--valutakurser/forklaring-till-serierna/svenska-marknadsrantor/`.

[23] *Swedish Financial Benchmark Facility*. May 6, 2021. URL: `https://swfbf.se/`.

Bachelor's Theses in Mathematical Sciences 2021:K28
ISSN 1654-6229

LUNFMS-4057-2021

Matematisk statistik
Matematikcentrum
Lunds universitet
Box 118, 221 00 Lund

http://www.maths.lth.se/