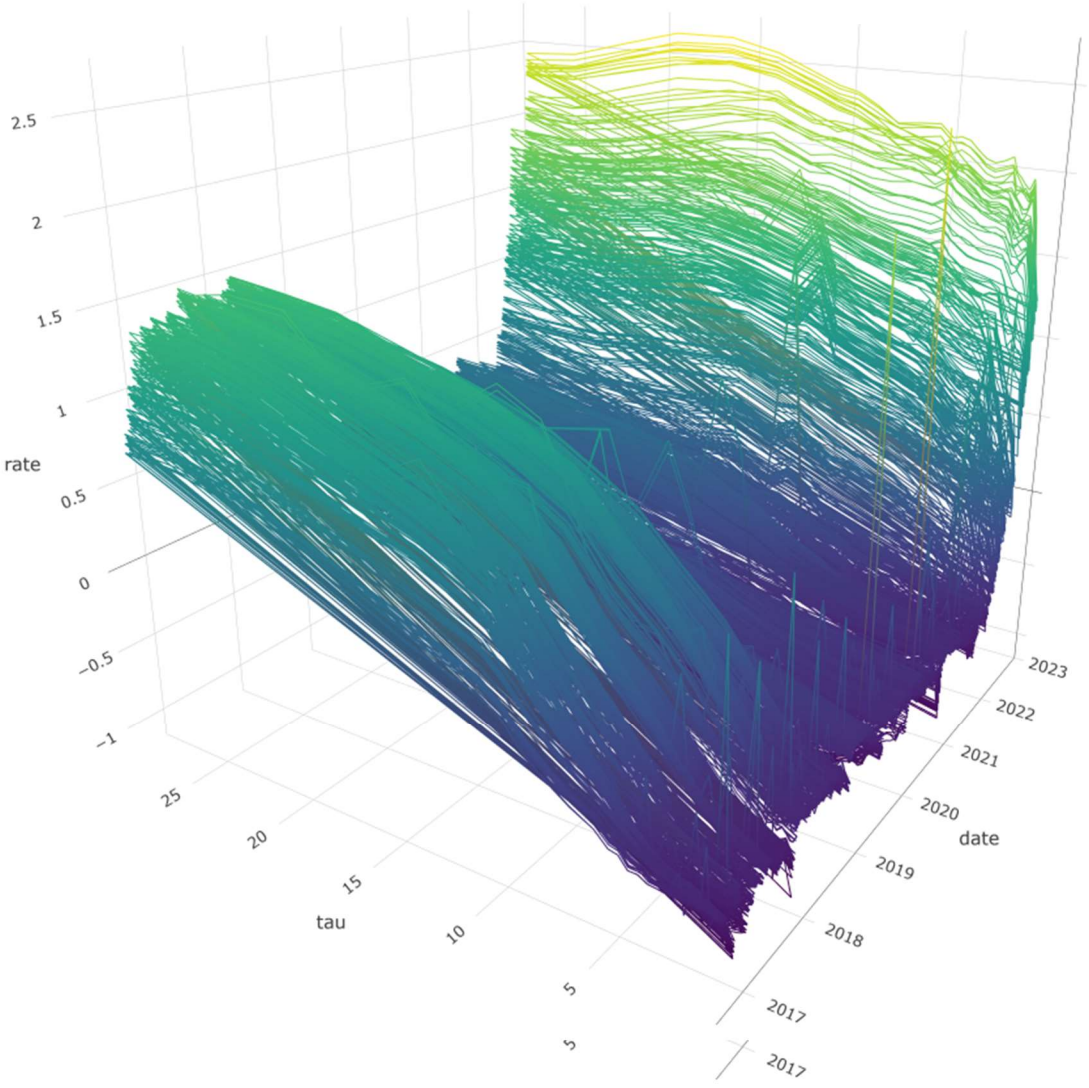
Prediction of the Swap Spread

A Novel Approach to Predict the Spread between Two Correlated Interest Rates.



J.V. Swanenburg

October 24, 2023

TU Delft

Predicting the Swap Spread with a Dynamic Nelson-Siegel Model

A Novel Approach to Predict the Spread between Two Correlated Interest Rates.

J.V. Swanenburg

to obtain the degree of Master of Science at the
Delft University of Technology

Student number:	4565614
Thesis committee	Dr. D. Kurowicka
	Dr. P. Chen
	Dr. R. van den Goorbergh
Date	August 24, 2023

Abstract

This thesis aims to develop a methodology for predicting the swap spread, which is defined as the difference between the German government bond interest rate and the Euribor swap rate. Thus far, the prediction of interest rates is limited to the prediction of a single interest rate. This thesis introduces the simultaneous prediction of the spread between two correlated interest rate curves.
The methodology developed in this thesis considers the dependence between the bond rate and the swap rate. The study utilizes a Dynamic Nelson-Siegel (DNS) model, which is extended to incorporate the correlation between these two rates. The simulation studies reveal that the variants simultaneously predicting both the swap and bond rates using a restricted VAR(1) model for factor dynamics outperform the other variants in predicting the swap spread. Another important aspect considered is the stationarity of the latent factors. The simulation studies demonstrate that the stationarity of the empirical DNS factors accurately represents the stationarity of the true DNS factors. This motivates the reformulation of the DNS model into a new variant where the first-order differences of both the swap and bond rate latent factors are modeled by a restricted VAR(1) model.
A case study validates the developed new variant of the DNS model, demonstrating predictions for the swap and bond curves that have an accuracy comparable to the accuracy of the benchmark model. The key advantage of the DNS model over this benchmark model is that the DNS model predicts the swap curve and bond curve over the whole maturity spectrum. The prediction over the whole maturity spectrum is crucial to compute the spread between the two rates, which emphasizes the relevance of the new model presented in this thesis.

Keywords: Interest rate prediction, Swap spread prediction, Dynamic Nelson-Siegel model, linear Gaussian state space model, Kalman filter

Acknowledgements

I would like to express my sincere gratitude to the following individuals who have played a significant role in the successful completion of my master's thesis.

First of all, I would like to thank Dorota for her supervision throughout the process. The insightful suggestions and discussions have been a major contribution to this thesis as well as to my personal development as a statistician. Her guidance has played a crucial role in the direction of this study.

I am also grateful to all people from APG that have been involved in my graduation project. Their expertise and help have been very important in reaching the goals of this thesis. A special word of thanks goes to Hans for his dedicated time, invaluable feedback, and continuous support.

I would like to express my appreciation to my housemates for their understanding, motivation, and support in being there for me in both good and difficult times. Another special thank you goes to Koen for his unlimited support during this challenging phase of my studies. Finally, I am thankful to my parents for their belief in my abilities. Their patience and constant motivation have been a source of strength throughout this academic journey.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Background

Interest rates are of great importance for pension funds investments. Pension funds promise future pension payments to their participants. The fund estimates the value of these payments using interest rates. To ensure there is enough money for these future payments, the fund invests its money. A prominent part of this money is invested in bonds and the return of a bond is determined by its interest rate. Thus, on one hand, the interest rate is used to estimate the present value of future pension payments and on the other hand, it is used to compute the return on the bond investments.

Interest rates are applied in these two calculations because they establish a connection between the present value of money and its future value. They represent the percentage of change in value over a specific period of time. The rate itself is influenced by the duration between the current moment and the future moment. For example, it can be that the interest rate for an investment of 1 year is 2% per year, whereas the interest rate for an investment of 10 years is 2.5% per year. The relation between the interest rate and the time that passes is represented by the interest rate curve. An example of an interest rate curve is given in Figure 1.

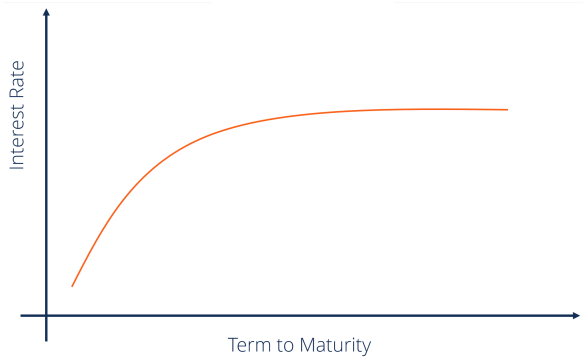


Figure 1: A visualization of an interest rate curve. The x-axis represents the term to maturity, which is the time between today and a future moment in time. The y-axis represents the corresponding interest rate per time unit.

In practice, the interest rate curve is not observed as a continuous function of time to maturity. The interest rate is only known for real market observations. The continuous interest rate curve is thus an approximation of the interest rate for each term. Moreover, the interest rates observed in the real market exhibit variability over time, which implies that the estimated interest rate curves also undergo continuous changes.

The interest rate that is used to compute the present value of the future pension payments must be estimated on a set of assets called 'swaps'. The choice for swaps as assets to compute today's value of the future pension is imposed by the Dutch Central Bank.
Although a pension fund invests part of its money in swaps, the majority of the money is invested in another asset called 'bonds'. That means, the return on the majority of these investments depends on another interest rate, namely the interest rate implied by the bonds. In other words, the interest rate that determines the return on the investments is another rate than the interest rate used to compute the value of future pension payments: the bond rate determines the return on the investments and the swap rate determines the value of the future pension payments.

A pension fund's role is to match the current investments with future pension payments. This could be a challenging task since the current investments depend on the bond rate, whereas the future pension payments depend on the swap rate. The difference between the bond and swap rates is thus a measure of the mismatch in the current investments and the value of the future pension payments. The difference between the bond and the swap rate with the same time to maturity is defined as the swap spread. The swap spread is visualized in Figure 2. When the two interest rates experience divergent changes, the swap spread changes and there will be a mismatch between the future pension payments and the return on investments. This mismatch implies the risk for the pension fund that there may not be enough funds available to pay future pensions.

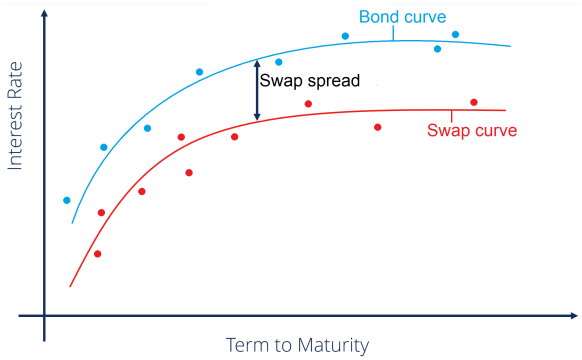


Figure 2: A visualization of the swap spread, which is defined as the difference between the bond rate and the swap rate with the same time to maturity. The points represent the bond and swap observations. The goal of this thesis is to develop a methodology to predict the swap spread as a continuous function of the time (or term) to maturity.

To ensure a match between future pension payments and investment returns, it is necessary to know the difference between the two rates. In order to address a potential future mismatch, it is necessary to predict the difference between the two interest rates for the whole spectrum of times to maturity. Therefore the aim of this thesis is to develop a model and a methodology that allows us to predict the swap spread as a continuous function of the time to maturity.

1.2 Related work

Modeling and predicting interest rates is a widely researched topic. The first models that have been proposed were published by Vasicek (1977), Cox, Ingersoll, and Ross (1985), and Hull and White (1990). It has been established that these models generally perform well in modeling interest rates but they are not suitable for forecasting purposes. Models that are used for both modeling and forecasting the interest rate stem from the work of Nelson and Siegel (1987), who proposed to model the interest rate in terms of three latent factors and their corresponding factor loadings. The extension of these models was presented on Svensson (1994). The model of Nelson and Siegel has been enriched by a fourth latent factor allowing for more flexibility. These models are referred to as Nelson-Siegel (NS) model and Nelson-Siegel-Svensson (NSS) model, respectively.

The first time the Nelson-Siegel model has been used for prediction purposes in Diebold and Li (2006). In this publication, three latent factors are modeled as an autoregressive process. Diebold and Li named their model 'Dynamic Nelson-Siegel (DNS) model'. This model is applied in this thesis to predict interest rates and it will be in detail explained in Section 3.

One of the earliest and most prominent contributions to the literature on DNS models was published by Diebold, Rudebusch, and Aruoba (2006), who extended the DNS model by introducing macroeconomic variables to the autoregressive process. Their model was therefore able to capture macroeconomic shifts. Additionally, Diebold et al. reformulated the DNS model to a state space representation. This representation allows for simultaneous estimation of the three latent factors and all other parameters, also known as the 'one-step' estimation. This is in contrast to the 'two-step' estimation for the original DNS model, where the latent factors are estimated by ordinary least squares. Then the other parameters are estimated for fixed values of latent factors. A disadvantage of this original two-step estimation is that the performance highly depends on the choice of factor loadings. Section 3.4.3 contains a detailed explanation of both the two-step and the one-step estimation procedure.
A third contribution of the publication of Diebold et al. (2006) is the interpretation of the three latent factors as level, slope, and curvature of the interest rate curve. This interpretation will be explained in Section 3.1.1.

A subclass of DNS models was introduced by Christensen, Diebold, and Rudebusch (2011), who defined the DNS model under the absence of riskless arbitrage opportunities named 'Arbitrage Free DNS (AFDNS) models'. The idea for this model stems from the findings of Björk and Christensen (1999), who showed that the NS model is not arbitrage-free. Christensen, Diebold, and Rudebusch (2009) combined the NSS and AFDNS models into a more flexible arbitrage-free model. The no-arbitrage assumption is particularly useful for the arbitrage-free pricing of assets and is therefore often used for trading purposes.
Besides the arbitrage-free models, there is another class of extensions of the DNS model. These models allow for time variation in parameters that are not latent factors. Among others is the work of Koopman, Mallee, and Van

der Wel (2007), who allow their model to have time-varying factor loadings and time-varying volatility. This extension leads to an improvement in the in-sample fit and the prediction performance of the model. Another contribution of this type is the model from Van Dijk, Koopman, Van der Wel, and Wright (2014) that has a time-varying unconditional mean for the autoregressive process. Their model improves prediction performance but is not available in state space representation. Therefore, the model is limited to the two-step estimation for the parameters.
The last class of extensions on the DNS models that is worth mentioning is the regime-switching DNS models, published first by Xiang and Zhu (2013). This model is able to capture the influence of macroeconomic cycles, which can be interpreted as different regimes. Guidolin and Pedio (2019) show that the regime-switching DNS model is particularly useful when predicting interest rates in a changing macroeconomic environment such as the Great Financial Crisis.

All papers cited in this section so far use United States government bonds as data to estimate and predict the interest rate, with the exception of the paper written by Svensson (1994), which uses Swedish forward contracts. There are also examples where the DNS model is used to model and predict interest rates for other types of assets, such as other government bonds for Central Banks (Filipović, 2009, Chapter 3), and the application of the DNS on credit default swaps by Shaw, Murphy, and O'Brien (2014).

In this thesis, I use the DNS model to estimate and predict the spread between the bond rate and the swap rate. To the best of my knowledge, the existing literature on utilizing the DNS model for modeling and predicting interest rates is solely focused on the modeling of a single interest rate curve. In this thesis, the DNS model is used to model and predict the spread, or difference, between two interest rate curves.

1.3 Problem statement

The goal of this thesis is to develop a framework that uses a Dynamic Nelson-Siegel model to estimate and predict the spread between bond interest rates and swap interest rates. The prediction horizon is 30 days. The estimations and predictions must cover the whole range of times to maturity between 0 and 30 years. The base model for the framework is the DNS model in the form proposed by Diebold and Li (2006). The variants on this DNS model mentioned in Section 1.2 are not included in the framework.
The starting point for the framework is the interpretation of the three factors of the DNS model. By focusing on the dynamics of these factors, three new variants on the DNS model will be proposed. These three variants are, in addition to the framework, the scientific contribution of this thesis. The framework to estimate and predict the spread will take into account these three new variants additional to the DNS model as proposed by Diebold and Li (2006).

The aforementioned goal of this thesis is realized through the following structure. In Section 2, the real data for the problem is described. The relevant theory related to the DNS model of Diebold and Li (2006) and the theory related to the three new variants is explained in Section 3. The development of the framework consists of several simulation studies. The method for these simulation studies is described in Section 4. The final framework is constructed from the results of these simulation studies, which is described in Section 5. This framework is then applied to the real data in the case study described in Section 6. Finally, based on both the simulation studies and the case study, a conclusion is drawn in Section 7. Furthermore, a discussion is provided in Section 8 and potential future work will be discussed as well.

2 Data

In this chapter, I will present the data that is used to estimate and predict the spread curve. The spread is defined as the difference between the swap interest rate and the bond interest rate. The most important economic definitions are included in this chapter. For a complete overview of the economic terms and methodology, see Appendix A.
This chapter is structured as follows. In Section 2.1 and Section 2.2 is a description of the data for the bond and swap contracts respectively. Each subsection begins by describing the data collection process. Subsequently, the data is thoroughly examined and prepared for further analysis. The subsection concludes with an overview of the data characteristics.
In Section 2.3 is a description of the spread data. It will be explained that it is not possible to compute the spread directly from the bond and swap data. Furthermore, this subsection will include an analysis of the relationship between the bond and swap data.

2.1 Bond data

The specific bond curve that is used in this thesis to compute the spread, is the zero coupon German government bond curve. The zero coupon bond curve represents the relation between the time to maturity and the yield to maturity of a zero coupon bond contract, as explained in Section 1.1. A zero coupon bond contract is defined as follows.

Definition 2.1 (Zero-coupon bond contract)**.** A zero-coupon bond is a contract where an investor lends money to the issuer in exchange for a future payment.

A zero coupon German government bond is a zero coupon bond of which the issuer is the German government.

2.1.1 Bond data collection

The data for the construction of the set of German government bonds originates from bond contracts that are available for trading in the market. The specifications of such a bond contract are collected and published by a prominent financial information and media company called 'Bloomberg'. Bloomberg utilizes diverse data sources such as financial institutions, regulatory filings, and news agencies and is widely considered a reliable data source. They combine the information from all their data sources to publish, among other data, bond contracts with the following attributes:

- Date t - the date and time at which this bond is available for trading for price $P(t,T)$.
- Price $P(t,T)$ in euros - the price at which the bond can be traded at date t.
- Ticker of the bond contract which serves as a unique identifier in Bloomberg.
- Maturity date T, the date at which the issuer of the bond contract pays the face value of the contract to the holder of the contract.
- Face value of €100, the amount of money paid by the issuer to the holder at the maturity date.
- A coupon value in percentage of the face value. The coupon is annually paid by the issuer to the holder of the bond contract.

An example of two observations from Bloomberg is included in Table 1. Bloomberg provides bond data with a real-time and continuous update frequency. For this thesis, I have sampled the data on a daily basis from Bloomberg, where each observation is sampled at the same time. The utilization of market data indicates a data limitation whereby no observations are available during periods of market closure, such as weekends. The missing observations will be discussed later in Section 3.4.2.

The set of bond observations that are used in this thesis consists of 91.541 observations. The number of observations per day varies and is in the range of 25-30 observations per day, depending on the number of bond contracts available in the market.

Date	Ticker	Maturity date	Face value	Price	Coupon (%)
2020-12-23	EI650542 Corp	2021-04-04	€100	€103.62	0.00
2020-12-23	EI879238 Corp	2022-01-04	€100	€104.77	2.00

Table 1: Example of two out of all 91.541 bond observation with the attributes that are available after gathering the data from Bloomberg.

2.1.2 Bond data preprocessing

Preprocessing the raw data gathered from Bloomberg consists of data cleaning, data enrichment, and data modification.

The first step is data cleaning, where two types of bonds are removed from the set of observations: the 'green bonds' and the short-term bonds. A green bond is a financial instrument specifically issued to raise capital for sustainable projects. The price of a green bond is potentially biased due to different investor demands and market dynamics.
A short-term bond is a bond with a time to maturity of less than 3 months. Short-term bonds are more susceptible to short-term market fluctuations, which can introduce noise to the interest rate estimates (Fernandes & Vieira, 2019). The elimination of the green bonds and the short-term bonds results in the exclusion of 45.919 data points from the data set.

The second step is data enrichment, where two attributes are added to each observation: the time to maturity τ and the yield to maturity y. These two attributes will be used to estimate the interest rate curve and the spread curve, which will be explained later in Section 3. The time to maturity τ is the time difference between the date of the observation and the maturity date.[1] The yield to maturity y can be computed by

$$y = \left(\frac{FV}{P(t,T)}\right)^{\frac{1}{\tau}} - 1,$$

where FV is the face value, $P(t,T)$ is the price of the observation at time t with maturity date T, and τ is the corresponding time to maturity.

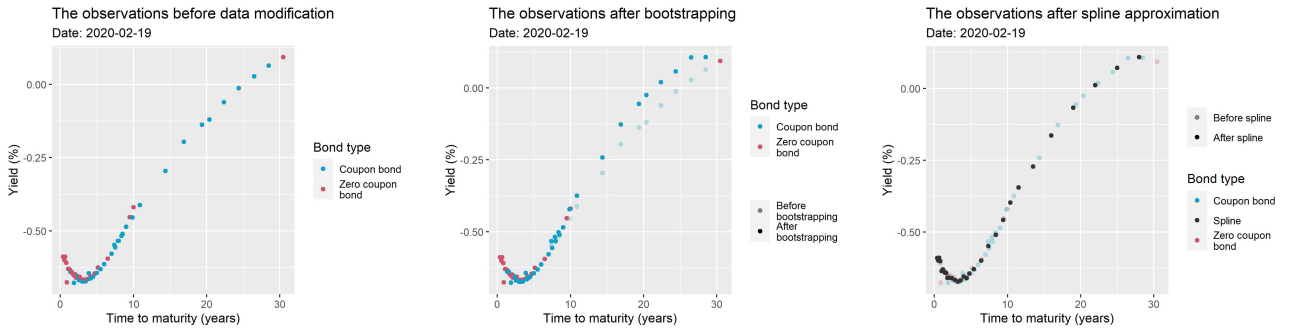
The third step is data modification. The first modification concerns all observations that have a coupon attribute that is not equal to zero. To estimate the zero coupon interest rate curve, we need a set of observations that have zero coupons. The yields of the observations with a nonzero coupon value are therefore modified such that they represent a zero coupon yield. This process is called 'bootstrapping' and is explained in detail in Section B.1.1.
For the estimation and prediction method that is used in this thesis, we need the same set of times to maturity τ for each day. But the observed data have different times to maturity for each day. To obtain a set of observations with a fixed set of times to maturity, we interpolate the observed yields by a cubic spline. This process leads to a data set of observations with times to maturity $\tau \in \{0.4, 0.525, 0.65, 0.8, 1, 1.2, 1.4, 1.6, 1.8, 2, 2.4, 2.8, 3.2, 3.6, 4, 4.4, 4.8, 5.4, 6.4, 7.4, 8.4, 9.4, 10.4, 11.5, 13.5, 16, 19, 22, 25, 28\}$, with τ in years. This modification is explained in detail in Section B.2. The effect of both modification steps is visualized for one day in the data set in Figure 3.

2.1.3 Bond data description and analysis

The set of bond observations that is used is used to estimate the spread curve consists of 46.860 observations. The first observations are from October 5, 2016, and the last observations are made on October 26, 2022. The time span is 2212 days. In Figure 4 is a three-dimensional plot of the zero coupon bond data. In this plot, you can see that the shape of the interest rate varies over time during our sample time interval. In the period 2016-2019 the curves were increasing and concave, during 2020-2021 the curves were flat and took values around zero, and in 2022 the curves remain flat but yields increase rapidly up to values around 2.5%.

The various yields will play a prominent role in the sequel. Hence I focus on them now in some detail. In Table 2, I present some descriptive statistics for the yields. From the mean column, it becomes clear that the average zero coupon interest rate curve is upward-sloping and from the standard deviation we see that the long rates are more volatile than the short rates. Also, from the autocorrelation, we know that the interest rates are persistent for each time to maturity.

[1] To compute the time difference, a day-count convention of 30/360 is used. More information about the day-count convention can be found in Section B.1.1.



(a) Observations without data modification

(b) Observations after bootstrapping the coupon bonds

(c) Observations after approximating the yields for a fixed set of maturities

Figure 3: Plots of the bond observations before and after the data modification steps. The modification consists of two steps: bootstrapping the coupon bonds into zero coupon bonds and approximating the yields for a fixed set of maturities. The spline points (black points in Figure 3c) will be used to estimate the zero coupon interest rate curve.

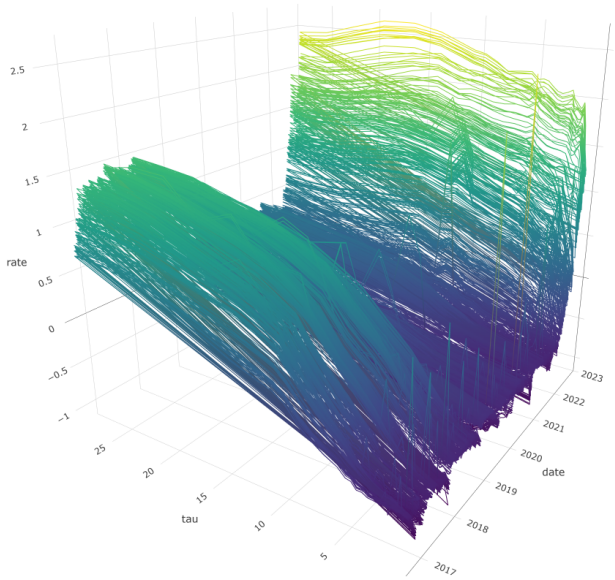


Figure 4: Three dimensional plot of the zero coupon German government bond data, 2016.08-2022.26. The sample consists of daily yield data at a fixed set of maturities for each day.

Time to maturity (years)	Mean	Std. dev.	Minimum	Maximum	$\hat{\rho}(1)$	$\hat{\rho}(14)$	$\hat{\rho}(30)$
0.4	-0.6253	0.2749	-1.062	1.25	0.9805	0.7517	0.5253
1	-0.5564	0.447	-0.9590	2.459	0.8965	0.6526	0.4914
2.4	-0.5310	0.4842	-1.027	2.052	0.9892	0.8408	0.6778
4	-0.4063	0.5131	-1.014	2.151	0.9872	0.8461	0.6987
5.4	-0.2783	0.5348	-1.000	2.230	0.9916	0.8647	0.7301
16	0.4376	0.6154	-0.7263	2.642	0.9945	0.9131	0.8314
22	0.594	0.6137	-0.5803	2.553	0.9952	0.9242	0.8577
28	0.6936	0.6044	-0.4875	2.438	0.9957	0.9324	0.8709
level	0.7479	0.6046	-0.4875	2.438	0.9957	0.9324	0.8709
slope	-1.237	0.5652	-2.342	-0.2932	0.9966	0.9602	0.9284
curvature	-0.1967	0.8108	-3.454	1.07	0.9812	0.874	0.7799
level diff	0.0009377	0.04175	-0.2078	0.2212	-0.3350	0.0213	0.0441
slope diff	0.0002706	0.04546	-0.263	0.2598	-0.4800	0.0460	-0.0239
curvature diff	-0.001829	0.1422	-2.461	2.594	-0.5020	-0.0131	0.000485

Table 2: A table with the descriptive statistics for the daily sampled zero coupon German government bond data set. The level, slope, and curvature represent the empirical counterparts for the level, slope, and curvature as defined by Diebold and Li (2006), and their diff values represent the series transformed by the linear difference operator. The last three columns contain sample autocorrelations at displacements of 1, 14, and 30 days.

2.1.4 Empirical counterparts of the bond data factors

In Section 1.2 it was explained that the interest rate curve can be described by three factors which can be interpreted as the level, slope, and curvature. Diebold and Li (2006) defined the empirical counterparts for the level, slope, and curvature being $y(\infty)$, $y(\infty) - y(0)$, and $2y(2) - y(0.25) - y(10)$ respectively. In Section 3.1 the definition for the empirical counterparts and the interpretation of the factors will be explained in more detail. Because these three factors also play a prominent role in the sequel, I include the analysis of their empirical counterparts in this section.
To analyze the empirical counterparts from the data, I use the following methodology:

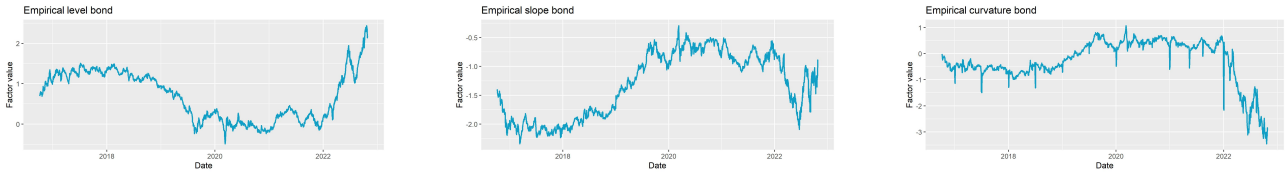
- The empirical counterpart for the level is extracted from the data by using the yield of the observation with the highest time to maturity, i.e. $y(\tau_{max})$.
- The empirical counterpart for the slope is extracted from the data by the difference between the yields of the observations with the highest and lowest times to maturity, i.e. $y(\tau_{max}) - y(\tau_{min})$.
- The empirical counterpart for the curvature is extracted from the data by the yields of the observations that have a time to maturity closest to 2, 0.25, and 10 years. The curvature counterpart is then computed by $2 \cdot y(\tau_2) - y(\tau_{0.25}) - y(\tau_{10})$, where τ_k is the time to maturity that is closest to k.

Table 2 includes the descriptive statistics for the three empirical counterparts of the factors. From the autocorrelation, we see that all three empirical factors are highly persistent.
In Section 3.2 it will be explained that the DNS model assumes a stationary autoregressive process for the factor dynamics. In Figure 5 are the time series of the values for the empirical counterparts. These plots suggest that the time series is not a stationary process. The nonstationarity is confirmed by the ADF test results in Table 3.[2]

In order to obtain a stationary process for the factor dynamics, I applied the linear difference filter

$$\nabla x_t = x_t - x_{t-1}$$

on each series separately. The ADF test results on the transformed series are presented in Table 4 and the descriptive statistics are included in Table 2. According to these statistics, the first-order differences of the factor dynamics are expected to be stationary and the transformed factors are less persistent than the time series without the transformation.



(a) level (b) slope (c) curvature

Figure 5: The time series of the empirical counterparts for the factors that describe the interest rate. The empirical factors are computed from the bond data according to to the methodology described in Section 2.1.4.

Factor	ADF test statistic	p value
Level	0.4220	> 0.99
Slope	-2.132	0.5223
Curvature	-0.2060	> 0.99

Table 3: The ADF test results for the empirical counterparts for the level, slope and curvature of the zero coupon German government bond data.

Factor	ADF test statistic	p value
Level	-28.71	< 0.01
Slope	-29.49	< 0.01
Curvature	-18.75	< 0.01

Table 4: The ADF test results for empirical counterparts transformed by the linear difference operator for the level, slope, and curvature of the zero coupon German government bond data.

[2]Not rejecting H_0: time series is stationary is based on a critical value of -2.87, corresponding to the 5% critical value of the ADF test for more than 500 observations. The number of lags used in the ADF test is determined by Akaikes Information Criterion.

2.2 Swap data

The specific swap curve that is used in this thesis to compute the spread, is the Euribor swap curve. A swap curve represents the fixed rate of a swap contract.[3] A swap contract is defined as follows.[4]

Definition 2.2 (Swap contract)**.** A swap contract is a contract between two parties that defines a scheme where a payment stream at a fixed interest rate is exchanged for a payment stream of a floating interest rate.

The Euribor swap rate is the fixed rate of a swap contract of which the floating rate is the Euribor 6m interest rate.

2.2.1 Swap data collection

The data that is gathered to construct the swap data set originates from swap contracts that are available for trading in the market. These swap contracts have a floating rate that is the 6-month Euribor. Just as the bond observations, the swap observations are published by Bloomberg with the following attributes:

- Date t - the date at which the swap contract is observed.
- Ticker of the swap contract which serves as a unique identifier in Bloomberg.
- Time to maturity τ in years.
- Face value of €100.
- Rate in the percentage of the face value that represents the fixed rate of a swap contract.

Just as for the bond observations, the swaps are on a daily basis. In Table 5 an example of two observations of the swap data set is presented.

Date	Ticker	Time to maturity (years)	Face value	Rate (%)
2020-12-23	EUSA1 Curncy	1	€100	-0.517
2020-12-23	EUSA2 Curncy	2	€100	-0.517

Table 5: Example of two swap observations with the attributes that are available after gathering the data from Bloomberg.

2.2.2 Swap data preprocessing

The preprocessing process of the swap data consists only of data modification because there is no cleaning or enrichment needed. The swap rate is already published by Bloomberg.

The swap observations gathered from Bloomberg have semi-annual payments of the fixed lag and annual payments of the floating lag. This means that a swap contract with a time to maturity of 2 years implies four payments of the fixed lag and two payments of the floating lag. Therefore, the rate attribute of a swap observation represents multiple cash flows at multiple days.
To compare the swap rate with the bond rate, it is necessary that both rates represent the same cash flows. Therefore, the swap data rates must be modified to get a fixed rate that represents the cash flow at the maturity date only. This modified rate is the rate that can be compared with the yield of a bond observation that has the same maturity date.
The method to modify the rate of a swap contract is the 'bootstrapping method', which is the same method that was used to modify coupon bonds into zero-coupon bonds. The procedure to bootstrap the swap rate is explained in Appendix B.1.2.

[3] To understand this thesis, it is not necessary to perfectly understand the relation between a swap contract and the swap rate. For clarity purposes, the explanation has been omitted from the current context.

[4] This is the definition for a plain Vanilla fixed-for-floating interest rate swap, which is a specific type of swap contract. Because there are no other swap contracts included in this thesis, I refer to it as a swap contract without any further elaboration.

2.2.3 Swap data description and analysis

The set of swap data points consists of 46.080 observations. Each observation contains the attributes as specified in Table 5. The first observations are on October 5, 2016, and the last observations are on October 26, 2022, which is exactly the same time span as the set of bond observations. Each date has exactly 30 observations, with a time to maturity $\tau \in \{1, 2, 3 \ldots, 30\}$.
In Figure 6 is a three-dimensional plot of the swap data. In this plot you can see that the swap curves take approximately the same shapes and values for the same time intervals as the bond curves in Figure 4.

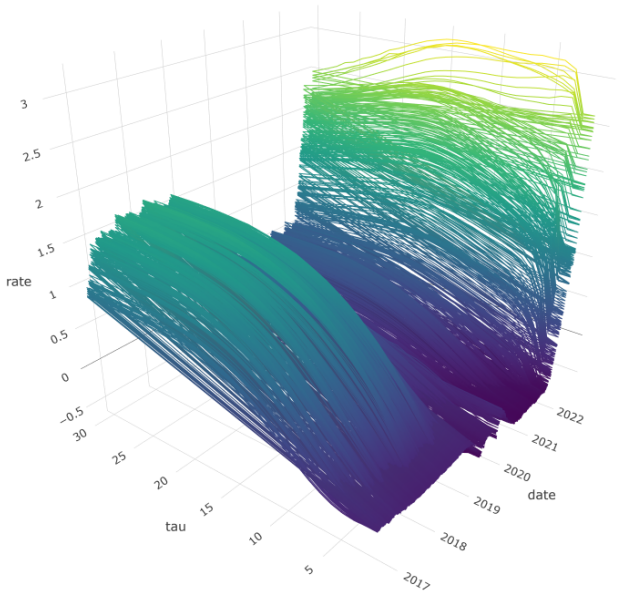


Figure 6: Three-dimensional plot of the swap data, 2016.08-2022.26. The sample consists of daily swap data at a fixed set of maturities for each day.

In Table 6 are the descriptive statistics for the swap rates, the empirical counterparts of the level, slope, and curvature, and the level, slope, and curvature transformed by the linear difference operator. The descriptive statistics indicate that the shape of the swap curves is comparable to the shape of the bond curve.

Time to maturity (years)	Mean	Std. dev.	Minimum	Maximum	$\hat{\rho}(1)$	$\hat{\rho}(14)$	$\hat{\rho}(30)$
1	-0.261	0.4404	-0.5445	2.62	0.9872	0.7977	0.6234
2	-0.261	0.4404	-0.5445	2.62	0.9872	0.7977	0.6234
3	-0.1913	0.5419	-0.5704	3.065	0.9892	0.8332	0.6952
5	-0.01	0.5882	-0.5705	3.141	0.9908	0.8642	0.7514
10	0.4862	0.6327	-0.3824	3.149	0.9931	0.9062	0.8189
15	0.8594	0.6525	-0.1948	3.156	0.9945	0.9271	0.8536
20	1	0.6405	-0.12	2.96	0.9954	0.9387	0.8743
30	1.039	0.6304	-0.2355	2.484	0.9965	0.953	0.9011
level	1.040	0.6306	-0.2355	2.484	0.9965	0.953	0.9011
slope	-1.112	0.5117	-1.939	0.381	0.9962	0.9364	0.8843
curvature	-0.4878	0.6329	-3.149	0.3824	0.9931	0.9062	0.819
level diff	0.001009	0.03652	-0.1745	0.2187	-0.4656	0.02924	-0.01849
slope diff	0.0007201	0.03349	-0.1744	0.1571	-0.4742	0.02814	-0.0005536
curvature diff	-0.001808	0.0345	-0.231	0.222	-0.4827	0.0029	0.02365

Table 6: A table with the descriptive statistics for the daily sampled swap data set. The level, slope, and curvature represent the empirical counterparts for the level, slope, and curvature as defined by Diebold and Li (2006). The level diff, slope diff, and curvature diff are the factors transformed by the two-times repeated linear difference operator. The last three columns contain sample autocorrelations at displacements of 1, 14, and 30 days.

2.2.4 Empirical counterparts of the swap data factors

The empirical counterparts are computed from the data according to the methodology that was described in Section 2.1.4. In Figure 7 are the plots of the empirical counterparts of the DNS factors computed from the

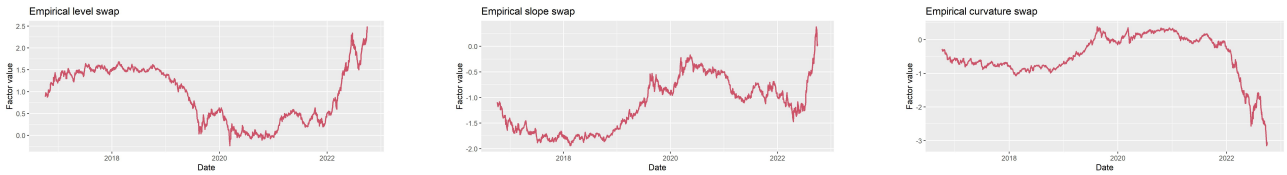
Factor	ADF test statistic	p value
Level	0.8871	> 0.99
Slope	-1.999	0.5789
Curvature	2.553	> 0.99

Table 7: The ADF test results for the empirical counterparts for the level, slope, and curvature of the swap data.

Factor	ADF test statistic	p-value
Level	-16.27	< 0.01
Slope	-27.51	< 0.01
Curvature	-26.54	< 0.01

Table 8: The ADF test results for empirical counterparts transformed by the linear difference operator for the level, slope, and curvature of the swap data.

swap data. In Table 7 and Table 8 are the results of the ADF tests on the series. These plots and numbers show that the empirical counterparts of the level, slope, and curvature do not follow a stationary autoregressive process.
Following the same reasoning as for the bond data, I applied the first-order difference transformation to the series of empirical factors. According to the ADF test results in Table 8, these series are stationary.



(a) level (β_1) (b) slope (β_2) (c) curvature (β_3)

Figure 7: The time series of the empirical counterparts for the factors that describe the interest rate. The empirical factors are computed from the swap data according to to the methodology described in Section 2.1.4.

2.3 Spread data

The swap spread is defined as the difference between the swap rate and the bond rate with corresponding time to maturity τ. In this paragraph, it will be explained that it is not possible to compute the swap spread from our data set of bond and swap observations. Furthermore, the relation between the bond and swap observations is analyzed.

2.3.1 Empirical swap spread does not exist due to mismatch in times to maturity

The bond data has times to maturity $\tau \in \{0.4, 0.525, \ldots, 28\}$ and the swap data has times to maturity $\tau \in \{1, 2, \ldots, 30\}$. This makes it impossible to compute the spread directly from our data sets. This is visualized in Figure 8. For a certain swap observation, there is no matching bond observation, hence the swap spread does not exist empirically for that time to maturity.

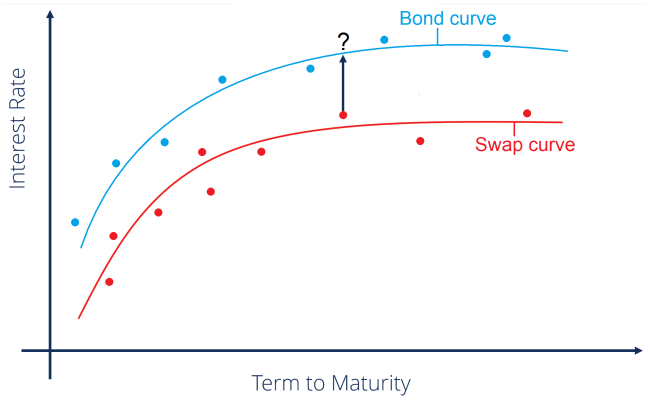


Figure 8: A plot that visualizes that the swap spread can not be computed from the data due to a mismatch in times to maturity.

The data is thus limited to two separate data sets. This motivates the need for a model that estimates and predicts the whole bond curve and the whole swap curve to be able to compute the prediction for the swap spread.

2.3.2 Analysis of the relation between the swap and bond observations

Although we have two separate data sets for both instruments, the analysis in the previous paragraph shows that there is a strong relationship between both data sets: Figure 4 and Figure 6 show that the shape of the curves is comparable for each time period, and the dynamics of the empirical counterparts for the level, slope and curvature in Figure 5 and Figure 7 take approximately the same values.
The relation between the bond data and the swap data is confirmed by the cross-correlation coefficients. The correlation coefficients of the empirical counterparts of both the swap and the bond data are presented in Table 9. The correlation coefficients of the series transformed by the linear difference operator are also presented in the table. All cross-correlation coefficients are significantly different than zero, which confirms the relationship between the series.

	level	slope	curvature	level diff	slope diff	curvature diff
Cross-correlation (1)	0.926	0.859	0.906	-0.413	0.374	-0.157
Cross-correlation (-1)	0.930	0.861	0.908	-0.427	-0.294	-0.184

Table 9: Cross-correlation coefficients for lag 1 for the empirical counterparts of the level, slope, and curvature of the bond data and the swap data. The 'diff' columns are the coefficients for the series transformed by the linear difference operator.

In summary, our data consist of a bond data set and a swap data set. The shape of the bond curve and the swap curve is analyzed through the empirical counterparts of the level, slope, and curvature, which will be explained in more detail in Section 3.1. The empirical counterparts are persistent but do not follow a stationary autoregressive process. The empirical counterparts transformed by the linear difference operator do follow a stationary autoregressive process but are less persistent than the not transformed series. This holds for both the bond and swap data and we will use this information later in Section 3.3.2.
It is not possible to compute the spread directly from the bond and swap data sets due to a mismatch in times to maturity. This motivates to use a model that estimates and predicts the swap curve and the bond curve as a continuous function for the time to maturity.

3 Dynamic Nelson-Siegel model

The Dynamic Nelson-Siegel model is a model that is widely used by, among others, central banks to predict the interest rate curve (Filipović, 2009). A main advantage of the DNS model is that it models the interest rate curve as a continuous function of the time to maturity τ. In Section 2.3.1 it was explained that to predict the swap spread, it is necessary to predict the swap curve and the bond curve as a continuous function of the time to maturity. This is the main motivation to use the DNS model to predict the swap spread.

In Section 1.2 it was explained that the DNS model models the interest rates by three latent factors. These three latent factors are then modeled by an autoregressive model in order to obtain predictions for the factors. These predictions are used to obtain predicted interest rate curves. This section contains the theoretical framework of the DNS model step-by-step. Section 3.1 elucidates the NS model with its interpretation, serving as the foundation for the DNS model, and in Section 3.2 is the original DNS model as proposed by Diebold and Li (2006), along with its matrix representation.

Diebold et al. (2006) remarked in 2006 that the factors of the DNS model were strongly correlated with macroeconomic variables. They improved the prediction accuracy of the DNS model by extending the autoregressive processes for the latent factors into a vector autoregressive process that included the macroeconomic variables. The same reasoning can be applied in the prediction of two correlated interest rate curves. If the latent factors of both curves are correlated, modeling the latent factors in one vector autoregressive process could improve the predictions of both rates. This leads to the first new variant of the DNS model, which will be presented in Section 3.3.1.

The original DNS assumes a stationary autoregressive process for the latent factors. In Section 2.1 and in Section 2.2, the interpretation of the latent factors is used to obtain a series of empirical factors. The series of empirical factors are not stationary, which could imply that modeling the latent factors by an autoregressive model is incorrect.
A linear first-order difference transformation of the empirical factors is stationary, as shown in Section 2.1 and in Section 2.2. This is the main motivation for the second new variant of the DNS model in this thesis. This variant models the first-order differences of the latent factors of the DNS model by an autoregressive process. This variant is presented in Section 3.3.2.

There is a third new variant of the DNS model that combines the reasoning of the first two variants. This model models the first-order differences of the latent factors of both the swap curve and the bond curves in one vector autoregressive model. This variant will be presented in Section 3.3.3.

To estimate the parameters and to construct predictions, the model is formulated as a state space model. The utilization of a state space representation offers numerous advantages compared to the original formulation of the DNS model. These advantages, among others, include the capability to effectively handle missing observations and to obtain theoretically correct inference. Both proposed models in Section 3.3 can be formulated in state space form as well, which is shown in Appendix E.
Using the state space representation, the parameters of the model can be estimated by maximum likelihood estimation. The likelihood of the parameters is evaluated with the Kalman filter. The Kalman filter and the parameter estimation are shown in Section 3.4.2 and Section 3.4.3 respectively. In Section 3.4.4 it is explained how to obtain MSE optimal predictions and the prediction intervals of the interest curves and the spread curves. Finally, the model performance assessment will be presented in Section 3.5.

3.1 The Nelson-Siegel model

Nelson and Siegel (1987) published a simple model that describes the interest rate y as a function of time to maturity τ, $y(\tau)$, with only four parameters: $\beta_1, \beta_2, \beta_3$ and λ. The model relates τ to the interest rate y by

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

where the interest rate $y(\tau)$ is a sum of three factors β_1, β_2 and β_3 weighted by their factor loadings 1, $\left(\frac{1-e^{-\tau\lambda}}{\tau\lambda} \right)$ and $\left(\frac{1-e^{-\tau\lambda}}{\tau\lambda} - e^{-\tau\lambda} \right)$. The NS model is in the literature referred to as the three-factor model for interest rates.[5]

[5]The factorization in Equation 1 is different than the original factorization in the article of Nelson and Siegel. The factorization in Equation 1 was published by Diebold and Li Diebold and Li (2006) in 2006. Their factorization matches the original NS model by reordering the equation but has the advantage over the original model that the factors β_1, β_2 and β_3 can be interpreted intuitively.

3.1.1 The interpretation of the NS model

We follow the interpretation of the model presented by Diebold and Li Diebold and Li (2006) which starts with the three-factor loadings. These loadings determine the weight on the factors per value of τ and depend on the value of τ and the parameter λ.

The relation between each factor loading and τ is visualized in Figure 9. The first loading is independent of τ, which implies that the weight on the first factor β_1 is constant over all values of τ. The second loading is a downward-sloping curve starting at a value of 1 and converging to zero for increasing values of τ. The third loading is a 'bumped' curve that starts at zero and converges to zero for higher values of τ. The shape of the third loading implies that the third factor has a smaller weight for rates with very short or very long maturities relative to rates with a middle maturity.
The relation between the factor loadings and τ forms the basis for the interpretation of the parameter λ and the factors β_1, β_2, and β_3.[6]

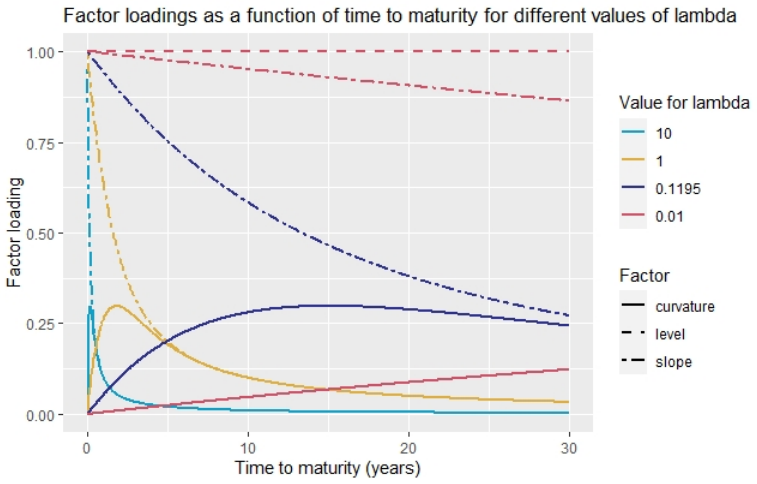


Figure 9: The factor loadings in the NS model in Equation 1 for different values of λ.

The interpretation of λ The parameter λ included in the second and third-factor loadings influences the exponential decay rate of the factor loadings. Small values of λ lead to slow decay and large values to fast decay. The value of λ also determines where the third loading achieves its maximum. In Figure 9 the loadings are plotted for four different values of λ. It can be observed that smaller values for λ imply a faster decay of the β_2-loading. Also, a smaller λ results in the β_3-loading attaining its maximum at a smaller value of τ.

Interpretation of the factors Diebold and Li (2006) provided interpretation for the factors β_1, β_2 and β_3 as being respectively the level, slope, and curvature of the interest rate curve. They defined the level of the interest rate curve as $y(\infty)$, the slope as $y(\infty) - y(0)$, and the curvature as $2 \cdot y(2) - y(0.25) - y(10)$.

For infinitely large values of τ, i.e. $\tau \to \infty$, the β_1-loading is a constant and the β_2- and β_3-loadings converge to zero. That implies that the rates for infinitely large values of τ are determined only by the first factor β_1. The factor β_1 thus can be interpreted as the level of the interest rate curve and is therefore referred to as the long-term factor or the level factor.
For small values of τ, i.e. $\tau \to 0$, the β_1 and β_2-loading are equal to 1 and the β_3-loading converges to zero. Hence the rates for small values of τ are determined by the first and second factors β_1 and β_2. By the definition of the slope as $y(\infty) - y(0)$, the slope of the interest rate curve is equal to the negative value of β_2. The factor β_2 is referred to as the short-term factor or the slope factor.
The factor β_3 starts at zero and converges to zero. An increase in β_3 will have little effect on very short or very long interest rates. An increase in β_3 will increase medium-term rates, which load more heavily on β_3. Thereby, β_3 is closely related to the curvature of the interest rate curve and is referred to as the medium-term factor or the curvature factor.

[6] The computations of the convergence of the factor loadings are shown in Appendix C.1.

The NS model in Equation 1 models the interest rate curve by defining the relation between the time to maturity and the interest rate $y(\tau)$ as a sum of three factors weighted by the factor loadings. These factors can be interpreted as the level, slope and curvature of the interest rate curve.
The NS model is limited to modeling the curve for one date and hence it is not directly useful for forecasting purposes. Therefore, the model is extended to the Dynamic Nelson-Siegel (DNS) model.

3.2 The Dynamic Nelson-Siegel model

The DNS model is an extension of the NS model by allowing the parameters β_1, β_2, β_3, and λ to take different values over time. The DNS model was proposed by Diebold and Li (2006) and has the form

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

Just as the NS model, the DNS model in Equation 2 is non-linear in the parameter λ_t. Simultaneous estimation of λ and the factors β_1, β_2, and β_3 results in a nonlinear optimization problem with certain computational issues. Discussion about computational issues of the DNS model, when λ is not constant, can be found in Cottle and Thapa (2017). To avoid the computational issues, Diebold and Li (2006) propose to choose a fixed value for λ in advance. A fixed value for λ makes the model less flexible to the data.

3.2.1 The DNS model as a linear model of the factors

If the value of λ is known, the DNS model is linear in the three factors β_1, β_2, and β_3. Consider a data set of n observations of $\{(y_{i,t}, \tau_{i,t})\}_{i=1}^{n}$, where y_i represents the interest rate, τ_i represents the time to maturity in years, and the index t represents the date of the observed value. Assume that all n observations have the same date index t. Then the DNS model in Equation 2 models the interest rate as

$$\begin{pmatrix} y_t(\tau_1) \\ y_t(\tau_2) \\ \vdots \\ y_t(\tau_n) \end{pmatrix} = \begin{pmatrix} 1 & \frac{1-e^{-\tau_1\lambda}}{\tau_1\lambda} & \frac{1-e^{-\tau_1\lambda}}{\tau_1\lambda} - e^{-\tau_1\lambda} \\ 1 & \frac{1-e^{-\tau_2\lambda}}{\tau_2\lambda} & \frac{1-e^{-\tau_2\lambda}}{\tau_2\lambda} - e^{-\tau_2\lambda} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\tau_n\lambda}}{\tau_n\lambda} & \frac{1-e^{-\tau_n\lambda}}{\tau_n\lambda} - e^{-\tau_n\lambda} \end{pmatrix} \begin{pmatrix} \beta_{1,t} \\ \beta_{2,t} \\ \beta_{3,t} \end{pmatrix} + \begin{pmatrix} \varepsilon_t(\tau_1) \\ \varepsilon_t(\tau_2) \\ \vdots \\ \varepsilon_t(\tau_n) \end{pmatrix}, \tag{3}$$

where $\begin{pmatrix} y_t(\tau_1) & \dots & y_t(\tau_n) \end{pmatrix}^T$ is the vector of the n interest rates for observations $y_t(\tau_i) = y_{i,t}$. The $n \times 3$-matrix contains the time-independent factor loadings for each observation, $\beta_\mathbf{t}$ is the vector of factors at time t, and $\varepsilon_\mathbf{t}$ is the vector of error terms which are considered as being pricing errors. The pricing errors $\varepsilon_t(\tau_i)$ are assumed to be mutually independent, normally distributed, $\varepsilon_t \overset{iid}{\sim} \mathcal{N}(0, H)$, for H a diagonal matrix.

The value of λ in practice is not known, but estimated separately from the estimation of the factors. The representation of the DNS model in Equation 3 allows for this separate estimation of λ and the factors, which will be explained later in Section 3.4. For now, it is only important to notice that the interest rate is a linear combination of the factors $\beta_{1,t}, \beta_{2,t}$ and $\beta_{3,t}$ because the following subsections describe the dynamics of these factors.

3.2.2 Dynamics of the DNS factors

Diebold and Li (2006) recognize that market observations have a natural temporal ordering, and therefore the factors $\beta_t = \begin{pmatrix} \beta_{1t} & \beta_{2t} & \beta_{3t} \end{pmatrix}^T$ in the DNS model are strongly correlated over time. They argue that the dynamic movements of the factors can be captured with a vector autoregressive (VAR) model.

The proposed VAR model is a VAR model of order 1 on the three dynamic factors. The VAR(1) model describes the factor dynamics as

$$\begin{pmatrix} \beta_{1,t} \\ \beta_{2,t} \\ \beta_{3,t} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} + \begin{pmatrix} \phi_{11} & & \emptyset \\ & \phi_{22} & \\ \emptyset & & \phi_{33} \end{pmatrix} \begin{pmatrix} \beta_{1,t-1} \\ \beta_{2,t-1} \\ \beta_{3,t-1} \end{pmatrix} + \begin{pmatrix} \eta_{1,t} \\ \eta_{3,t} \\ \eta_{2,t} \end{pmatrix}, \tag{4}$$

where $\beta_{i,t}$ denotes the value of factor $i = 1, 2, 3$ at time t, c_1, c_2, c_3 denote the time independent trend of the model, and ϕ_{ij} is the coefficient that determines the relation between factor $\beta_{i,t}$ and factor $\beta_{j,t-1}$. The vector η_t contains the error terms for the factors in the model, which are assumed to be mutually independent, normally distributed, $\eta_t \overset{iid}{\sim} \mathcal{N}(0, Q)$, for Q a diagonal matrix.

Notice that the off-diagonal elements ϕ_{ij} in the autoregressive coefficient matrix are zero, and therefore the VAR model is restricted and is actually a combination of three univariate AR(1) models. This assumption is also used in several prominent articles about the DNS model (Diebold & Li, 2006), (Shaw et al., 2014). The argument for a restricted VAR model in these articles is that due to a large number of included parameters, there is a risk of in-sample overfitting. Additionally, a full VAR matrix would take much more computational time and lower tractability. There are also articles that use a full matrix Φ for the VAR model in Equation 4 (Diebold et al., 2006), (Koopman et al., 2007). As stated in Section 1.3, the DNS models in this thesis are derived from the original DNS model. For that reason, and for the computational tractability, I assume an autoregressive coefficient matrix Φ where all off-diagonal elements are equal to zero.

3.3 Variants of the DNS model

In this section, I propose three variants of the original DNS model. The first variant, referred to as sim-statDNS, combines the factor dynamics of two different interest rates into one autoregressive model. The second variant, referred to as nonstatDNS, models the dynamics of the factors β_1, β_2 and β_3 by an autoregressive model on the first-order difference of the factors, instead of the factors themselves. The third variant, referred to as sim-nonstatDNS, also combines the factor dynamics of two rates but models the first-order differences as a stationary autoregressive process.

3.3.1 sim-statDNS: DNS to estimate two interest rates simultaneously

This thesis focuses on the prediction of not one, but two interest rate curves. In 2006, Diebold et al. (2006) improved the prediction performance of the DNS by including macroeconomic variables in the vector autoregressive process for the latent factors. In Section 2, it was shown that the level (β_1), slope (β_2), and curvature (β_3) estimated on the swap data, were strongly correlated with the level, slope, and curvature of the bond data. Using the same reasoning as Diebold et al. (2006), the first variant of the DNS model is to combine the factors of the swap data with the factors of the bond data into one VAR model, where each factor is related to its own lagged value and the lagged value of the other interest rate curve,

$$\begin{pmatrix} \beta_{1,t,swap} \\ \beta_{2,t,swap} \\ \beta_{3,t,swap} \\ \beta_{1,t,bond} \\ \beta_{2,t,bond} \\ \beta_{3,t,bond} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \\ c_6 \end{pmatrix} + \begin{pmatrix} \phi_{11} & & & \phi_{14} & & \\ & \phi_{22} & & & \phi_{25} & \\ & & \phi_{33} & & & \phi_{36} \\ \phi_{41} & & & \phi_{44} & & \\ & \phi_{52} & & & \phi_{55} & \\ & & \phi_{63} & & & \phi_{66} \end{pmatrix} \begin{pmatrix} \beta_{1,t-1,swap} \\ \beta_{2,t-1,swap} \\ \beta_{3,t-1,swap} \\ \beta_{1,t-1,bond} \\ \beta_{2,t-1,bond} \\ \beta_{3,t-1,bond} \end{pmatrix} + \begin{pmatrix} \eta_{1,t} \\ \eta_{2,t} \\ \eta_{3,t} \\ \eta_{4,t} \\ \eta_{5,t} \\ \eta_{6,t} \end{pmatrix}, \tag{5}$$

where the empty entries of the matrix Φ represent value zero, $\eta_t \overset{iid}{\sim} \mathcal{N}(0, Q)$ and Q a 6×6-diagonal matrix. This implies that the error terms of the factors are mutually independent.

Thus, the complete first variant of the DNS model is given by Equation 2 and Equation 5 and will be referred to as 'sim-statDNS'. This variant uses more information from the relationship between the bond and swap rate by combining the factors of two interest rates and is therefore expected to have a better fit and a better prediction performance.

3.3.2 nonstatDNS: DNS with nonstationary factor dynamics

The estimation of a VAR model requires all the time series to be stationary. The time series of the empirical counterparts for the level (β_1), slope (β_2), and curvature (β_3) computed from the data were nonstationary processes, as shown in Section 2. Modeling these factor dynamics by a VAR model would therefore be not correct.

To model the factors by an autoregressive model, the data is transformed by taking the first-order differences. The transformation that implies a stationary process for the factor dynamics is the first-order differences on each of the time series, as demonstrated in Section 2.[7] The series of first-order differences are modeled by a VAR(1) process,

$$\begin{pmatrix} \nabla\beta_{1,t} \\ \nabla\beta_{2,t} \\ \nabla\beta_{3,t} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} + \begin{pmatrix} \phi_{11} & & \emptyset \\ & \phi_{22} & \\ \emptyset & & \phi_{33} \end{pmatrix} \begin{pmatrix} \nabla\beta_{1,t-1} \\ \nabla\beta_{2,t-1} \\ \nabla\beta_{3,t-1} \end{pmatrix} + \begin{pmatrix} \eta_{1,t} \\ \eta_{2,t} \\ \eta_{3,t} \end{pmatrix}, \tag{6}$$

[7] The nonstationarity of the factors is concluded from the ADF test on the empirical factors. Later, in Section 5.1, it will be shown that the nonstationarity of the empirical factors implies nonstationarity of the DNS factors

where $\nabla\beta_{i,t} = \beta_{i,t} - \beta_{i,t-1}$, c_i is the trend, ϕ_{ij} is the coefficient that determines the relation between $\nabla\beta_{i,t}$ and $\nabla\beta_{i,t-1}$. The vector η_t contains the error terms for the factors in the model.

Thus, the second variant of the DNS model is given by Equation 2 and Equation 6, and in the sequel this model will be referred to as 'nonstatDNS'. This variant should be used when the first-order difference of the factors follows a stationary autoregressive process.

There are two reasons for choosing the first-order difference transformation for the time series. First, as stated previously, this transformation ensures stationarity of the factor dynamics in the data that is used in this thesis. Second, this transformation is linear, which is necessary for the formulation of the model as a linear state space model, which will be explained subsequently in Section 3.4.1. Furthermore, the model for the transformed factor dynamics only uses one lag and could thereby ignore relevant information from earlier observations. Incorporating multiple lags in this model is possible but is out of the scope of this thesis.

3.3.3 sim-nonstatDNS: DNS to estimate two rates simultaneously with nonstationary factor dynamics

The third variant of the DNS is a model where the first-order differences of the factor dynamics are assumed to be stationary. This is a combination of the two previous variants in Section 3.3.1 and Section 3.3.2. The first-order differences of the factor dynamics are modeled by a VAR(1) model where the factors are assumed to correlate with each other. The autoregressive model is given by

$$\begin{pmatrix} \nabla\beta_{1,t,swap} \\ \nabla\beta_{2,t,swap} \\ \nabla\beta_{3,t,swap} \\ \nabla\beta_{1,t,bond} \\ \nabla\beta_{2,t,bond} \\ \nabla\beta_{3,t,bond} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \\ c_6 \end{pmatrix} + \begin{pmatrix} \phi_{11} & & & \phi_{14} & & \\ & \phi_{22} & & & \phi_{25} & \\ & & \phi_{33} & & & \phi_{36} \\ \phi_{41} & & & \phi_{44} & & \\ & \phi_{52} & & & \phi_{55} & \\ & & \phi_{63} & & & \phi_{66} \end{pmatrix} \begin{pmatrix} \nabla\beta_{1,t-1,swap} \\ \nabla\beta_{2,t-1,swap} \\ \nabla\beta_{3,t-1,swap} \\ \nabla\beta_{1,t-1,bond} \\ \nabla\beta_{2,t-1,bond} \\ \nabla\beta_{3,t-1,bond} \end{pmatrix} + \begin{pmatrix} \eta_{1,t} \\ \eta_{2,t} \\ \eta_{3,t} \\ \eta_{4,t} \\ \eta_{5,t} \\ \eta_{6,t} \end{pmatrix}, \tag{7}$$

where the empty entries of matrix Φ are equal to zero, $\eta_t \overset{iid}{\sim} \mathcal{N}(0.Q)$, and Q a 6×6-diagonal matrix.

The third variant of the DNS model is composed of Equation 2 and Equation 7. In the sequel, this model will be referred to as the 'sim-nonstatDNS' model.

3.4 State space modeling

Diebold and Li (2006), who published the original DNS model, estimated the parameters in two steps: first, the factors, and then the parameters in the autoregressive model were estimated. In this two-step procedure, λ is not estimated Equation 2 and it is not possible to produce correct confidence and prediction intervals. Also, this approach assumes a complete set of observations and can not handle missing observations (which is the case for the data we will model).
To overcome these limitations, Diebold et al. (2006) reformulated the original DNS model in state space representation. In a state space model, the factors, λ, and the parameters can be estimated simultaneously. Furthermore, this one-step approach is able to handle missing observations in the data set.

This section will describe the process of estimating the parameters and generating predictions using a DNS model represented in the state space form. In Section 3.4.1, the DNS models will be formulated in state space representation. Although the original DNS model was already formulated in this form, the formulation of the three variants in state space representation is introduced in this thesis for the first time.
To estimate the parameters and obtain predictions, the Kalman filter will be used. The Kalman filter is an algorithm that estimates and predicts latent variables and simultaneously computes the loglikelihood of the parameters in a model. Details about the Kalman filter are included in Section 3.4.2.
In Section 3.4.3, it will be explained how to estimate all the parameters in the DNS models by maximum likelihood estimation. A crucial part of this estimation procedure is the choice for the initial parameters. The initial parameters will be estimated by the 2-step approach of Diebold and Li (2006), which will be explained in the same section.
Once the parameters are estimated, the model can be used to obtain predictions for the interest rate curves. Section 3.4.4 contains the details of how to produce MSE optimal predictions and corresponding confidence intervals.

3.4.1 DNS model in state space representation

The state space system for a DNS model consists of two equations: a measurement equation and a transition equation. The measurement equation captures the relation between the observed interest rates and the unobserved factors, and the transition equation captures the dynamics of the unobserved factors over time. For the original DNS model, the measurement and transition equation correspond to Equation 3 and Equation 4 respectively. The new DNS models defined in Section 3.3 are formulated as state space models in Appendix E.

The general form for a DNS model in state space representation is

$$b_t = c + \Phi b_{t-1} + \eta_t \tag{8a}$$
$$y_t = \Lambda_t b_t + \varepsilon_t, \tag{8b}$$

where y_t are the observations, b_t the unobserved unobserved factors, in the DNS model, ε_t the measurement errors and η_t the factor disturbances. Equation 8a is the transition equation and Equation 8b is the measurement equation.
The general transition equation uses the notation of b_t instead of β_t for the unobserved state vectors. That is because, for the variants of the DNS model, this vector includes the lagged factor values as well. By the usage of b_t the state space formulation is sufficient for all DNS models and there will be no confusion between β_t and b_t.

In this thesis, I assume that η_t, ε_t and b_0 are multivariate Gaussian random variables. With this assumption, the system in Equation 8a-8b is extended with the following conditions,

$$\eta_t \sim \mathcal{N}(0, Q) \tag{8c}$$
$$\varepsilon_t \sim \mathcal{N}(0, H_t) \tag{8d}$$
$$b_0 \sim \mathcal{N}\left(\bar{b}_0, P_0\right), \tag{8e}$$

such that the full state space system for the original DNS model is specified by Equation 8a-8e. The parameters in the state space model are collected in

$$\Theta := \{\Theta_t, ;\ t = 1, \ldots, T\} := \{\bar{b}_0, P_0, c, \Phi, Q, \lambda, H_t, ;\ t = 1, \ldots, T\}.$$

The formulation of the DNS models as state space models serve as a framework in which we can estimate the parameters of the model and make predictions. Both the estimation and prediction make use of the Kalman filter algorithm. Therefore, we now first introduce the Kalman filter algorithm.

3.4.2 Kalman filter

Consider the state space model given by Equation 8a-8e. Suppose all parameters Θ_t are known. For a data set of observations $\{y_{i,t}, \tau_{i,t}\}_{t=1}^{T}$, the Kalman filter estimates and predicts the unobserved state vectors b_t, predicts the future interest rates y_t for $t > T$ and computes the likelihood of the parameters in Θ_t on the data set of observations. How the Kalman filter is used to obtain each of these estimates and predictions will be explained subsequently in Section 3.4.3 and Section 3.4.4.

What will follow now is a short explanation of the Kalman filter and its application in the DNS models. The full algorithm, provided in Algorithm 3.1, consists of two steps: a prediction step and an update step. The prediction step predicts the latent factors $\hat{b}_t$ and the observations $\hat{y}_t$ one step ahead. The update step revises the prediction for the latent factor once the next observation has been observed. By iterating over these two steps, the algorithm obtains estimates and predictions for each moment t.

Firstly, the Kalman filter will be applied to obtain the log-likelihood of parameters Θ for a given set of observations. Hence, the Kalman filter will be used in the maximum likelihood estimation. This will be explained subsequently in Section 3.4.3.
Secondly, the Kalman filter is applied to obtain predictions for the latent factors and the interest rates. This will be explained later in Section 3.4.4.
Finally, an in-depth explanation of theory related to the Kalman filter is provided in Appendix F.

The Kalman filter is able to produce estimates and predictions even when one or more observations $y_{i,t}$ are missing. For missing observations, the algorithm skips the update step and uses the results of the prediction step as its estimates. This is particularly useful for our data set, as we do not have observations in the weekends.

Algorithm 3.1 (The Kalman filter)**.** Set the initial state vector equal to its mean $b_0 = \bar{b}_0$ and the initial covariance matrix of the state vectors $P_{0|0}$ equal to the covariance matrix of the initial state P_0. The Kalman filter algorithm obtains a series of estimated state vectors $b_1, \ldots, b_T$ by iteratively performing a prediction step and an update step. For $t = 1, \ldots, T$:

1. Prediction step from time $t-1$ to time t.
   Assume that we have observed $\mathcal{Y}_{t-1} = \{y_1, \ldots, y_{t-1}\}$, but not yet y_t.
   (a) Predict the mean $\bar{b}_t$ of factor b_t by the point estimate:
   $\bar{b}_{t|t-1} = \Phi \bar{b}_{t-1|t-1}$
   (b) Predict the covariance matrix P_t of the factor b_t:
   $P_{t|t-1} = \Phi P_{t-1|t-1} \Phi' + Q_t$
   (c) Predict the observation vector y_t by the point estimate:
   $\hat{y}_{t|t-1} = \Lambda_t \bar{b}_{t|t-1}$
   (d) Predict the covariance matrix F_t of the observation estimate:
   $F_t = \Lambda_t P_{t|t-1} \Lambda_t' + H_t$
2. Update step from at time t.
   Assume that we now have observed y_t as well.
   (a) Compute the Kalman gain K_t:
   $K_t = P_{t|t-1} \Lambda_t' \left(F_t\right)^{-1}$
   (b) Compute the prediction error of the observation prediction $\hat{y}_t^{t-1}$ from step 1(c):
   $v_t = y_t - \hat{y}_{t|t-1}$
   (c) Update the estimate for the mean $\bar{b}_t$ of factor b_t:
   $\bar{b}_{t|t} = \bar{b}_{t|t-1} + K_t v_t$
   (d) Update the estimate for covariance matrix on the states:
   $P_{t|t} = P_{t|t-1} - K_t \Lambda_t P_{t|t-1}$

3.4.3 Parameter estimation

The first application of the Kalman filter is to estimate the parameters $\Theta = \{\bar{b}_0, P_0, c, \Phi, Q, \lambda, H_t;\ \ t = 1, \ldots, T\}$ by maximum likelihood estimation. For a set of observations $Y_T = \{y_{i,t}, \tau_{i,t}\}_{t=1}^{T}$, the log-likelihood of Θ is given by[8]

$$\ln \ell\left(\Theta\right) = -\frac{nT}{2} \ln\left(2\pi\right) - \frac{1}{2} \sum_{t=1}^{T} \ln|F_t\left(\Theta\right)| - \frac{1}{2} \left(v_t\left(\Theta\right)\right)' F_t\left(\Theta\right)^{-1} \left(v_t\left(\Theta\right)\right), \tag{9}$$

where n is the number of observations per time step, T is the number of time steps, and $v_t(\cdot)$ and $F_t\left(\cdot\right)$ are the one step ahead forecast error and its covariance matrix given Y_{t-1}. Notice the dependence of v_t and F_t on the parameters in Θ.
The values for $v_t\left(\Theta_i\right)$ and $F_t\left(\Theta\right)$ directly follow from the Kalman filter algorithm. The value for v_t is obtained in step 2(b) and the value for F_t is obtained in step 1(d). Thus the log-likelihood for parameter Θ can be computed by running the Kalman filter algorithm once for Θ on the set of observations.

The computation of the log-likelihood by the procedure described in this section assumes that each day has an equal amount of observations. In Section 2.1 it was explained that the number of observed bonds was different for each day. In this thesis, it was chosen to preprocess the bond observations with splines, such that each day has the same amount of observations with the same times to maturity.
Another option could be to use all observations in the data in one large matrix, where the entries for bonds that were not observed at that specific date are empty. This, however, leads to a very large, sparse observation matrix Y_t, resulting in computational complexities.
There is a method that avoids the assumption that each date must have the same number of observations. This method is called the 'univariate treatment of a multivariate series' and is explained in Durbin and Koopman (2012a). The idea of this method is to introduce each single observation separately. This procedure is used in the context of a DNS model by Koopman et al. (2007), but is left out of the scope of this thesis.

[8] The derivation of this log-likelihood function is provided in Section F.2

Optimization algorithm To find $\hat{\Theta}_{MLE}$, we start with an initial Θ_0, for which we can evaluate the log-likelihood with the Kalman filter. Then the 'BFGS'-optimization algorithm is used to produce a new estimate Θ_1, for which we again evaluate the log-likelihood with the Kalman filter. This iterative process proceeds until it is converged for a relative convergence criterion of $1 \times e^{-8}$.
The 'BFGS' algorithm is a quasi-Newton method that was also used by Koopman et al. Koopman et al. (2007). Through iterative adjustment of each of the parameters, the BFGS algorithm seeks the optimal solution by maximizing the objective function, which is the log-likelihood function. The algorithm approximates the inverse Hessian matrix of the objective function based on gradient information.
The gradient of the log-likelihood function is known for some of the parameters in Θ, but not for all of them (Durbin & Koopman, 2012a). The only implementation of state space models that uses these gradient functions, is the implementation in the programming language `Ox` Koopman et al. (2007). Because the implementation for this thesis will be in the programming language `R` and because there is no implementation of the gradients of the log-likelihood available for `R`, I choose to approximate the gradient of the log-likelihood in each of the parameters by a finite-difference approximation.

Initial parameters To start the MLE optimization, we need the initial parameter Θ_0. This parameter is obtained by the so-called 'two-step' approach (Nelson & Siegel, 1987). For this method, we choose a certain value for λ. With this λ, the state vectors b_t are estimated by ordinary least squares, which is the 'first step'. Then in the 'second step', a VAR model is estimated by OLS on the time series of estimated factors. From these two estimated models, we can derive all initial parameters for the MLE estimation. See Appendix F.5 for a detailed explanation. Thus, for one initial choice of λ, the two-step approach provides all parameters in Θ.

Initial choice for λ The only initial parameter that must be chosen in advance, is the value of λ. In literature, there are various methods on how to choose the value of λ. Nelson and Siegel (1987) choose their value of λ as the value that provided the lowest in-sample error on the data among a grid of λ values. In contrast, Diebold and Li (2006) argued that λ must be chosen such that the $\beta_{3,t}$ factor loading attains its maximum for τ being the average time to maturity over all data points, see Figure 9.
Both these methods did not include the estimation of λ in the MLE procedure but rather kept the λ fixed to its initial value. Diebold et al. (2006) were the first that estimated λ by MLE, using the initial value from the method of Diebold and Li (2006). Later research on the DNS model adopts the MLE value of λ from Diebold et al. (2006) as the initial value, leading to fast convergence. This is possible because all articles apply the DNS model to the same type of data: US government bonds.

In this thesis, I do not use US government bonds for my data. Therefore, I choose the initial value of λ using the method of Diebold and Li (2006). Using this λ, the initial parameters are estimated by the two-step approach, and $\hat{\Theta}_{MLE}$ is estimated by maximum likelihood estimation.

3.4.4 Predictions

The predictions for the interest rate consist of point estimates, which are the predicted interest rates, and their uncertainty, which are the confidence intervals.

Predicting the interest rate The Kalman filter produces MSE optimal predictions for both the state vectors b_t and the observation vectors y_t. The MSE optimal prediction for b_t is the conditional expectation $\mathbb{E}[b_t|Y_{t-1}]$ and the MSE optimal prediction for y_t is the conditional expectation $\mathbb{E}[y_t|Y_{t-1}]$. By the assumption in Equation 8c-8e, we know that both b_t and y_t are normally distributed, conditional on the observations Y_{t-1}. That implies that the MSE optimal predictions for b_t and y_t are the conditional means predicted in step 1(a) and step 1(c) of the Kalman filter respectively.

For multiple steps ahead predictions, we skip the update step and use each predicted value at time t to predict the values at time $t+1$. The specific Kalman filter algorithm for making multiple steps ahead predictions are given in Section F.4.

Prediction intervals With the Kalman filter, we obtain not only the point estimates for b_t and y_t but also predictions for their covariance matrices P_t and F_t. Since b_t and y_t are normally distributed conditional on Y_t, by knowing the mean and covariance matrix, we know the full distribution of b_t and y_t. This property is used to construct the prediction intervals.

The prediction intervals are constructed using the standard formula for a prediction interval of a normal random variable,

$$[\mu - z \cdot \sigma, \mu + z \cdot \sigma],$$

where μ and σ are the mean and standard deviation of the normal random variable, and z is the standard score. For the 9% prediction intervals, a standard score of 1.96 will be used.
For the prediction intervals of b_t, μ is set to $\hat{b}_t$ and the standard deviation σ is set to the square root of the diagonal elements of P_t. For the prediction intervals of $\hat{y}_t$, the mean μ is set to $\hat{y}_t$ and the standard deviation σ is set to the square root of the diagonal elements of F_t.

3.5 Assessment of the model performance

The model performance will be assessed by its accuracy and uncertainty. The accuracy is measured by the root mean squared error (rMSE) and the uncertainty is determined by the confidence and prediction intervals of the modeled interest rate curves.

In Section 2.3.1 it was explained that due to a mismatch in times to maturity for the real data, it is not possible to compute the true swap spread. This implies that in the case study, the estimated or predicted swap spread curve can not be compared to the true swap spread curve. For that reason, the performance of the estimated and predicted swap spread curves will be assessed by analyzing the performance of the swap and bond curves separately. The swap and bond curve accuracy is measured by their rMSE on the data.

In the simulation studies, however, it is possible to simulate data such that the swap and bond observations have matching times to maturity. This makes it possible to determine the rMSE of the estimated and predicted swap spread curve in the simulation studies. The accuracy assessment of the estimated and predicted curves in the simulation studies consists thus of the rMSE of the swap, bond, and swap spread curves.

3.5.1 The random walk model as benchmark model

In statistics, it is common to compare the model performance to a benchmark model. The benchmark model in this thesis is the random walk model.
The random walk model is a simple mathematical concept used to describe the unpredictable movement of a variable over time. In this model, the variable's future position is determined solely by its current position and random changes or "steps" at each time step. The random walk model assumes that the interest rate at moment $t = t_i$ is equal to the interest rate at moment $t = t_{i+1}$. Therefore, the predicted interest rates are equal to the last observed interest rate observations.

A major disadvantage of the random walk model in the prediction of the swap spread is that the random walk model does not model the interest rates as a continuous function of the time to maturity. The mismatch in times to maturity of the real data implies that it is not possible to compute the swap spread from this benchmark model. That means that the random walk model only serves as a benchmark model to compare the prediction accuracy of each rate separately and can not be used to compare the prediction accuracy of the spread in the case study.

In the simulation studies, the data will be simulated in a way that the swap and bond data have matching times to maturity. That makes it possible to compute the swap spread for the benchmark model as well. In the simulation studies, the accuracy of the swap spread curve will thus be compared to the benchmark model as well.

4 Procedure for the simulation studies

The goal of this thesis is to develop a methodology to predict the swap spread using a DNS model. All theory related to the original DNS model of Diebold and Li (2006) is described in the previous section. Additionally, three new variants based on this DNS model are introduced. To define a methodology based on this theory, 13 different simulation studies are performed. This methodology is empirically tested on real data. In this section, it will be explained what simulation studies will be performed, how they will be performed, and how a methodology is derived from the simulation studies.

This section is structured as follows. Section 4.1 contains a short summary of all four DNS models that can be used in the methodology. Section 4.2 contains a description of the goal of each of the 13 different simulation studies that will be performed in this thesis. In Section 4.3, the general simulation procedure is described by four steps: data simulation, data analysis, model estimation and prediction, and the assessment of the model behavior and performance. This is followed by the description of the implementation in Section 4.4, where the names and usage of all software packages that will be used for the simulation study are included.

4.1 Four competitive DNS models

A main part of the simulation studies is to investigate which DNS model best fits the data. In the selection process, there are four potential DNS models to consider:

- The original DNS model as proposed by Diebold and Li (2006). This model is defined in Equation 3 and Equation 4. This model models a single interest rate of which the factor dynamics are assumed to follow a stationary autoregressive process of order one. Henceforth I will refer to this model as 'statDNS'.
- The DNS model of which the first-order differences of the factor dynamics are assumed to follow a stationary autoregressive process of order one. This model is described by Equation 3 in combination with Equation 6. I will refer to this model as 'nonstatDNS'.
- The DNS model that models two interest rates simultaneously. All factor dynamics are assumed to follow a stationary autoregressive process of order one. This model is described by Equation 3 in combination with Equation 5. I will refer to this model as 'sim-statDNS'.
- The DNS model that models two interest rates simultaneously of which the dynamics of the first-order differences of the factors are assumed to be stationary. This model is described by Equation 3 in combination with Equation 7. I will refer to this model as 'sim-nonstatDNS'.

These four models will be used throughout all simulation studies.

4.2 Goal of the simulation studies

A total of 13 distinct studies will be conducted to derive the methodology to estimate the DNS models. Each study has its objective and contribution to this methodology, which I will specify in this chapter. The simulation studies that will be performed, can be divided into five groups.

1. The empirical counterparts for the level, slope, and curvature factors.
   The simulations within this group focus on the relation between the empirical factors that can be computed from the data and the true underlying factors that can not be observed. The hypothesis is that the dynamics and systematic relation of the empirical factors are a good representation of the dynamics of the DNS model factors. In that case, the structure of the true DNS factors can be decided based on the empirical factors.

2. The model behavior.
   The simulations within this group focus on the estimation process for the parameters and the factors in the model, as well as their mutual relationships. By doing these simulations, we learn how certain parameters in the model behave relative to other parameters. In real-world scenarios, models can be difficult to understand due to the presence of numerous interconnected variables. These simulations allow us to explore the behavior under known conditions, which is useful in understanding the model behavior in more complex or real-world situations.

3. The model choices.
   The simulations within this group focus on the effect of the initial values and data preprocessing process on the model behavior and performance. This simulation study helps to investigate the impact of specific modeling choices on the output and therefore directly contributes to the methodology.

4. The validation of the new DNS variants.
   The simulations within this group aim to validate that in theory the new DNS variants are valid models to predict the interest rate curves. The validation process includes the ability to estimate the factors, the shape and accuracy of the estimated and predicted interest rate curves, and the validation of the model assumptions. The validation of the DNS models will be used to determine whether or not the DNS model should be used to predict the swap spread.

5. The model performance. These simulations aim to compare the competing DNS variants by their prediction accuracy and uncertainty to obtain an expectation of which model would give the best predictions.

The simulation studies within each of these five groups are summarized in Table 10.

Number	Simulation group	Simulation	Objective
1	Empirical counterparts	Empirical counterparts for a single interest rate	Find the relation between the value and dynamics of the empirical counterparts and the value and dynamics of the true DNS factors of a single interest rate.
2	Empirical counterparts	Empirical counterparts for two interest rates	Find the relation between the mutual correlation of the empirical counterparts and the correlation of the true DNS factors of two interest rates.
3	Model behavior	The estimation of $\hat{\lambda}_{MLE}$	Investigate the process and result of the estimation of $\hat{\lambda}_{MLE}$.
4	Model behavior	Interaction between λ, the factors and the factor loadings	Find the relation between the estimated λ and factors b_t to the model behavior and performance.
5	Model choices	Initial value for λ	Examine the effect of the initial value for λ on the estimated $\hat{\lambda}_{MLE}$.
6	Model choices	Missing observations	Investigate the effect of the missing weekend observations on the model performance.
7	Model choices	Distribution of the observations along the time to maturity axis	Investigate the effect of the choice to spline the bond data into a fixed set of maturities.
8	Model validation	Validation of the sim-statDNS model	Validating the sim-statDNS model
9	Model validation	Misleading results of the statDNS model	Explore the potential implications of estimating the statDNS model on nonstationary data.
10	Model validation	Validation of the nonstatDNS model	Validating the nonstatDNS model
11	Model validation	Validation of the sim-nonstatDNS model	Validating the sim-nonstatDNS model
12	Model performance	Stationary factor dynamics	Compare the estimation and prediction performance of the DNS models that will be applied when the factor dynamics are stationary autoregressive processes.
13	Model performance	Nonstationary factor dynamics	Compare the estimation and prediction performance of the DNS models that will be applied when the dynamics of the first order difference of the factors follow stationary autoregressive processes.

Table 10: Summary of the 13 simulation studies that are performed in this thesis.

4.3 Procedure for the simulation studies

In this section, I will describe the procedure that is used for all simulations in this thesis. A simulation study involves several steps: data simulation, analysis of data structure, model fitting, and assessment of the model performance. These steps will be explained subsequently.
For each simulation study, the whole simulation process will be repeated multiple times to check the consistency and correctness of the outcomes. The number of simulations varies from 12 to 36, and depends on the computational time to estimate the model. In Section 4.2, the exact number of simulations for each study is reported.

4.3.1 Data simulation

The data is simulated from one of the four DNS models stated in the introduction of this chapter. To simulate data, I assume a set of true parameters Θ_{true} for the DNS model. All true parameters are included in the appendix of this thesis, see Section G.1. These parameters are chosen such that the simulated data represents reasonable interest rate curves.
The simulated data consist of a set of simulated state vectors $\{b_t\}_{t=1}^T$ and a set of simulated interest rate observations $\{Y_t\}_{t=1}^T$. Initially, I simulate a series of state vectors $b_t = \{\beta_{1,t}, \beta_{2,t}, \beta_{3,t}\}$, for $t = 1, 2, \ldots, T$, using the process for the factor dynamics of the chosen DNS model. The series of state vectors incorporate randomness in the error terms ε_t. Subsequently, I simulate a set of observations $Y_t = \{y_{1,t}, y_{2,t} \ldots, y_{n,t}\}$, for $t = 1, 2, \ldots, T$, using the relation between the factors and the interest rates, see Equation 2. The observations incorporate randomness in the error terms η_t.

Both simulated data sets $\{b_t\}_{t=1}^T$ and $\{Y_t\}_{t=1}^T$ and are split into train and test data. The scope of this thesis is to predict the interest rate with a time horizon of 30 days. Therefore, the factors b_t and the observations Y_t for $t = 1, 2, \ldots, T - 30$ are used as train data. The last 30 factors and observations will be used as test data to assess the prediction performance. To assess the estimation performance of the DNS models, I will simulate a new set of observations Y_t for $t = 1, 2, \ldots, T - 30$ using the initial set of simulated factors b_t.

The steps for the simulation procedure are summarized in Table 11.

4.3.2 Analysis of the data structure

Once the data is simulated, I will analyze the structure of this data. This step is important to establish a methodology for selecting one of the DNS models because the model choice will mainly depend on the data structure. The four DNS models stated in the introduction of this chapter differ in their assumptions on the factor dynamics. In order to select the model that best fits the data, we will examine these factor dynamics.

Because the true factors are not known for real data, the empirical counterparts are the best representation for the factors of the DNS models and I will use these counterparts to examine the factor dynamics. Thus the initial step in analyzing the factors involves calculating the empirical counterparts for each factor. The computation for the empirical counterparts is based on the interpretation of the factors (Diebold & Li, 2006) and the exact computations are given in Section 2.1.4.

The primary examination of the factors involves assessing their stationarity. By looking at a plot of the factors, I get an initial guess on whether or not the process is stationary. Also, I will test the stationarity of the factors using the ADF test, where the number of lags in the autoregressive model is based on the AIC.
The subsequent step in the analysis of the factors is to identify the presence of any systematic relationship between the current observations and past observations at different time lags. To find such a relation, I will look at the autocorrelation plot and partial autocorrelation plot of the factor series.
The final step is to identify the presence of any systematic relationship between the factors of two different interest rates. To find such a relation, I will look at the cross correlation of two series of factors. This analysis only applies to the last simulation study on two interest rates.

The results of this analysis will be combined with the results of the fitted DNS model in order to obtain a methodology on which model suits the data with a certain structure.

4.3.3 Model fitting and predicting

One of the four DNS models stated in the introduction of this chapter will be used to obtain estimates and predictions for the interest rates. The process of estimating the parameters and predicting the interest rates is as described in Section 3.4.3. For convenience, the procedure is summarized in this paragraph.

The parameters of the model are estimated with maximum likelihood estimation. The likelihood of the parameters in Θ_i is evaluated with the Kalman filter. The likelihood is optimized with respect to the parameters in Θ with the BFGS-optimization algorithm with a convergence criterion of $1 \times e^{-8}$. Due to the absence of a theoretical gradient function, a finite-difference approximation will be used in the BFGS algorithm.

To start the maximum likelihood optimization, I need to choose the initial parameters for the DNS model. For the initial value of λ, I will follow the approach of Diebold and Li (2006) by choosing λ such that the curvature loading attains its maximum on the middle maturity, as explained in Section 3.4.3. This applies to all simulations, except for the third simulation, where I choose various initial values for λ to perform a sensitivity analysis. Subsequently, all other initial parameters in $\Theta_{initial}$ are estimated using the 2-step approach as explained in Section 3.4.3.
Following the completion of estimating the initial parameters $\Theta_{initial}$ and estimating the maximum likelihood parameters $\hat{\Theta}_{MLE}$, the model with the maximum likelihood parameters are employed to estimate and predict the interest rates. The MSE optimal estimates and predictions for the factors and the interest rates are constructed by applying the Kalman filter on the train data for the DNS model with parameters $\hat{\Theta}_{MLE}$.

The steps for the model fitting and interest rate estimation and prediction are summarized in Table 11. Notice that I only use the train data for this part of the process.

Simulation	Estimation and prediction
1. Choose the true parameters for the model 2. Simulate a time series of state vectors b_t, for $t = 1, \ldots, T$ 3. Simulate a set of observations y_t, for $t = 1, \ldots, T$ for a fixed set of maturities 4. Split the simulated data in train data and test data.	1. Choose the initial value for λ 2. Estimate the initial parameters by the 2-step approach 3. Estimate the MLE parameters by the 1-step approach 4. Predict the factors and the observations with the Kalman filter

Table 11: The steps taken for simulating the data, estimating the parameters and predicting the observations for the simulation studies.

4.3.4 Assessment of the model behavior and performance

The models will be compared by their robustness of the parameter estimation, by their ability to estimate de factors and by their accuracy and uncertainty of the estimated and predicted interest rate curves.

The assessment of the model's behavior involves examining the estimated parameters $\hat{\Theta}_{MLE}$. A special focus will be on the estimated values of $\hat{\lambda}_{MLE}$ and $\hat{\Phi}_{MLE}$. The estimated $\hat{\lambda}_{MLE}$ determines the shape of the factor loadings, see Figure 9, and directly influence the values of the estimated factors β_t. The estimated $\hat{\Phi}_{MLE}$ determines the autoregressive process of the factor dynamics and thus directly influences the relationship between the factors and their lagged values.

An additional aspect of evaluating the model's behavior involves the estimated and predicted series of factors β_t. The factor values in combination with the factor loadings directly determine the estimated and predicted interest rates. Furthermore, the stationarity of the factor dynamics gives insight into the model behavior as well.

The goodness of fit and the prediction accuracy of the model is examined by the root mean squared error (RMSE) between the estimated and predicted interest rate and the simulated test observations. The RMSE serves as a measure of accuracy and makes it possible to compare different models. The performance of the model will consistently be evaluated in comparison to the benchmark model. I will use the random walk model as a benchmark model. The random walk model models the interest rate at moment $t + 1$ by setting it equal to the interest rate at moment t.

Finally, I will investigate the distribution of the error terms by looking at the Q-Q plot in combination with the Shapiro-Wilk test, which tests whether or not the residuals are normally distributed. By looking at the error distribution, I can determine if the model assumption of normally distributed errors holds. The error term distribution also shows potential outliers.
Potential heteroskedasticity, i.e. consistency of error variability across different levels of the independent variables, is tested with the Breusch-Pagan test.
The definitions of all these tests and their interpretation are attached in Appendix D of this thesis.

4.4 Implementation

All simulations are implemented in `R` version 4.1.0. For the estimation of the initial factors $\beta_{t,OLS}$ in the 2-step procedure, I use the function `lm` from the `stats`-package. To estimate the initial parameters $\Theta_{initial}$, I use the function `VAR` from the `vars`-package, with $p = 1$.

To evaluate the likelihood with the Kalman filter, I use the function `kalman_filter` from the `kalmanfilter`-package. After coding my own implementation for the Kalman filter, I found that the `kalman_filter` function returns the loglikelihood in Equation 9 without the constant first term. Because the likelihood of the parameters will only be used to mutually compare the models, this first term will not influence the conclusion. Therefore, the log-likelihood value computed by the `kalman_filter` function is sufficient.
The construction of the prediction intervals from the covariance matrix F_t is done manually following the steps in Algorithm 3.1 and using $P_{t|t}$ predicted by the function `kalman_filter`. The `kalman_filter`-function can not produce the prediction for the covariance matrix F_t directly.
The optimization of the log-likelihood function is implemented using the `optim` function from the `stats`-package, where the method is set to 'BFGS' and the fnscale is set to -1, corresponding to a maximization of the log-likelihood with the BFGS optimization algorithm. The gradient function is not specifically specified, which implies that the `optim` function uses a finite-difference approximation.

For the implementation see the code on GitHub via the following link or the next url:
`https://github.com/jswanenburg/DNS_model_for_swap_spread_prediction/tree/3e880e3847061f01f9b607db6e2270abfdba9345`

5 Simulation studies

The framework to predict the swap spread will be derived from the simulation studies in this section. An overview of all simulation studies and their objectives was presented in the previous section in Table 10. In this section, the results for these simulation studies will be presented.
Section 5.1 contains the first group of simulations that focuses on the empirical factors. This is followed by Section 5.2, in which the model behavior is investigated. Then, in Section 5.3, the simulations for the model choices are discussed. In Section 5.4, the three new DNS models are validated. Finally, in Section 5.5, the DNS models are compared by their prediction performance.
Each of these sections contains an intermediate conclusion for the simulation studies within that specific group. These conclusions are combined into a final framework in item 5.6. This framework is the framework that will be used to predict the swap spread.

5.1 Analysis of the empirical factors

The first group of simulation studies aims to find a relation between the empirical counterparts of the factors and the estimated DNS factors. If there is such a relation, the empirical factors can be used to choose a time series model for the factors (AR(p), VAR(p)) that fits the data best. There are two simulation studies within this group. The first simulation focuses on the value, stationarity, ACF and PACF of a single interest rate. The second simulation focuses on the CCF of two simulated interest rates.

5.1.1 Value and dynamics of the empirical factors

The objective of this study is to investigate the relation between the value and dynamics of the empirical counterparts and the true DNS factors for a single interest rate curve. Using this relationship, the empirical factors can be used to choose the correct DNS model, which is for example the statDNS or the nonstatDNS model. This simulation study is the main motivation to develop the nonstatDNS model.
This study only consists of data simulation and analysis of the data structure. The model fitting and prediction are not included. For the simulations, I simulated 36 data sets from the statDNS model and 36 data sets from the nonstatDNS model. The empirical counterparts for the factors are computed for each simulated set of observations. Following the procedure in Section 4.3.2, the empirical factors are analyzed by the ADF test, and the ACF and PACF plots. Finally, these results are compared to the true underlying factors.
The results for this simulation study are split in two parts. The first set of results is based on the 36 simulated data sets from the statDNS model. These are the results obtained under the assumption that the factors follow a stationary autoregressive process of order one.
The second set of results is based on the 36 simulated data sets from the nonstatDNS model. These are the results obtained under the assumption that the first-order differences of the factors follow a stationary autoregressive process of order one.

Stationary factor dynamics (statDNS) The initial step for the analysis of the factors is to look at the plots of the empirical and true level, slope, and curvature. In Figure 10 are three time series plots for the empirical (blue line) and true (black line) level, slope, and curvature for one simulation. In these plots, it can be seen that the values of the empirical factors are not equal to the true factor values. In this plot, it is also visualized that the processes of the empirical factors have the dynamics of mean-reverting processes. The latter suggests that the stationarity of the true factors is captured by the empirical factors.

The stationarity of the empirical factors is further investigated by performing the ADF test on each series of factors. In Table 12 are the results presented as the percentage of series that is stationary or nonstationary according to the ADF test. The results imply that all series of empirical counterparts follow a stationary process. This confirms the suggestion that the stationarity of the true factors is captured by the empirical factors.

	Level (%)	Slope (%)	Curvature (%)
Stationary	100	100	100
Nonstationary	0	0	0

Table 12: The percentage of series of empirical factors that follow a stationary or nonstationary process given that the true factors follow a stationary autoregressive process of order one. Stationarity is determined by the ADF test with a critical value of -3.44 and a p-value lower than 0.05. All data is simulated from the statDNS model, 100% corresponds to all 36 simulations.

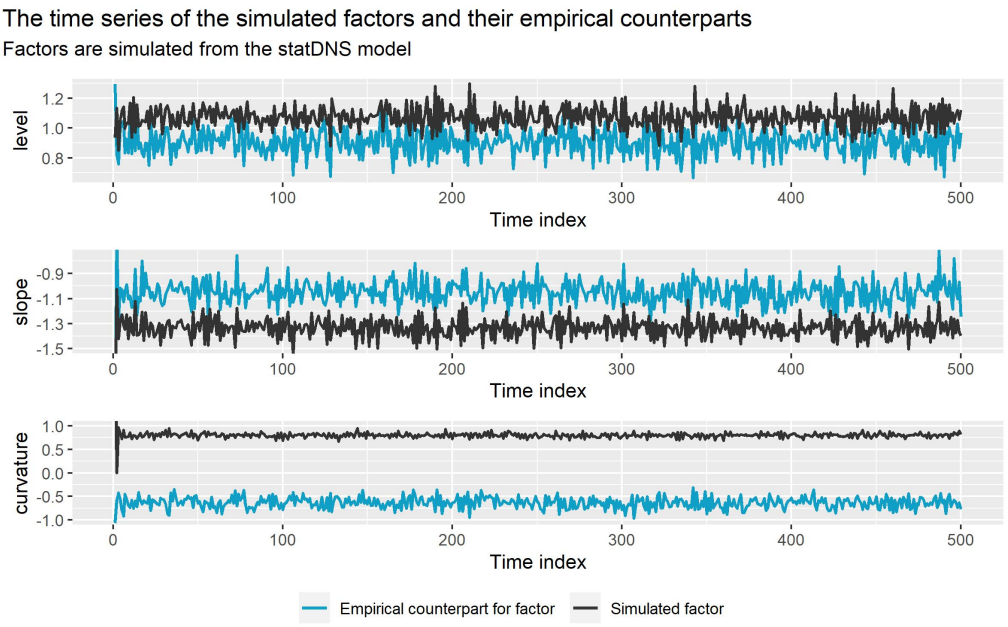


Figure 10: Plot of the series of empirical and simulated factors for one of the 36 simulations of the statDNS model.

The next step in analyzing the empirical factors is the presence of any systematic relationship in the series of factors. In this simulation, it is assumed that the factors follow an AR(1) process. The ACF and PACF plots of an AR(1) process are presented in Figure 11. The ACF of an AR(1) process tails off. The PACF of this process cuts off after one lag. The ACF and PACF of the empirical factors will be analyzed in comparison to the ACF and PACF of an AR(1) process.

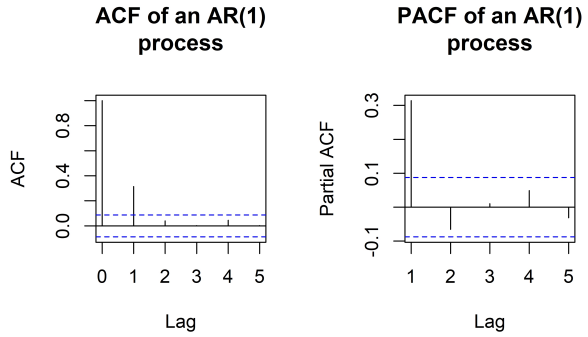


Figure 11: The theoretical ACF and PACF plot of an AR(1) process, The ACF and PACF of the empirical factors will be compared to this figure.

In Figure 12 are the ACF and PACF plots of the series of empirical level factors for one of the simulations. The ACF plot in this figure tails off and the PACF plot cuts off after one lag. These plots correspond to the ACF and PACF of an AR(1) model. Since the true factors are simulated from an AR(1) model, this result implies that the systematic relationship of the empirical factors corresponds to the systematic relationship of the true factors.

Nonstationary factor dynamics (nonstatDNS) Following the structure of the previous results, the plot of the time series is included in Figure 13, the results for the ADF tests are presented in Table 13, and the ACF and PACF plots of the level factor after a linear first order difference transformation are contained in Figure 14.

In Figure 13 it can be seen that in the case of nonstationary factor dynamics, the values of the empirical factors are not equal to the values of the true factors. However, the dynamics of the empirical level and slope show some similarity, but from these plots it can not be concluded whether the empirical factors follow a stationary autoregressive process.

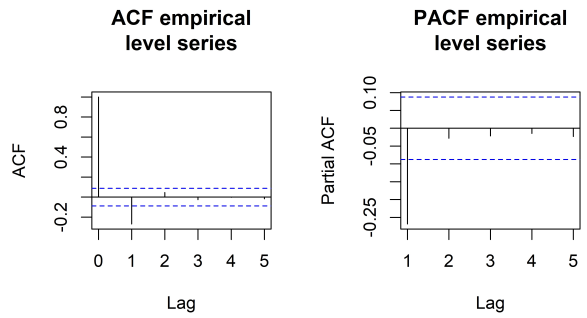


Figure 12: The ACF and PACF plot of the series of empirical counterparts for the level factor. Results for data simulated from the statDNS model with the assumption that the factors follow a stationary AR(1) model.

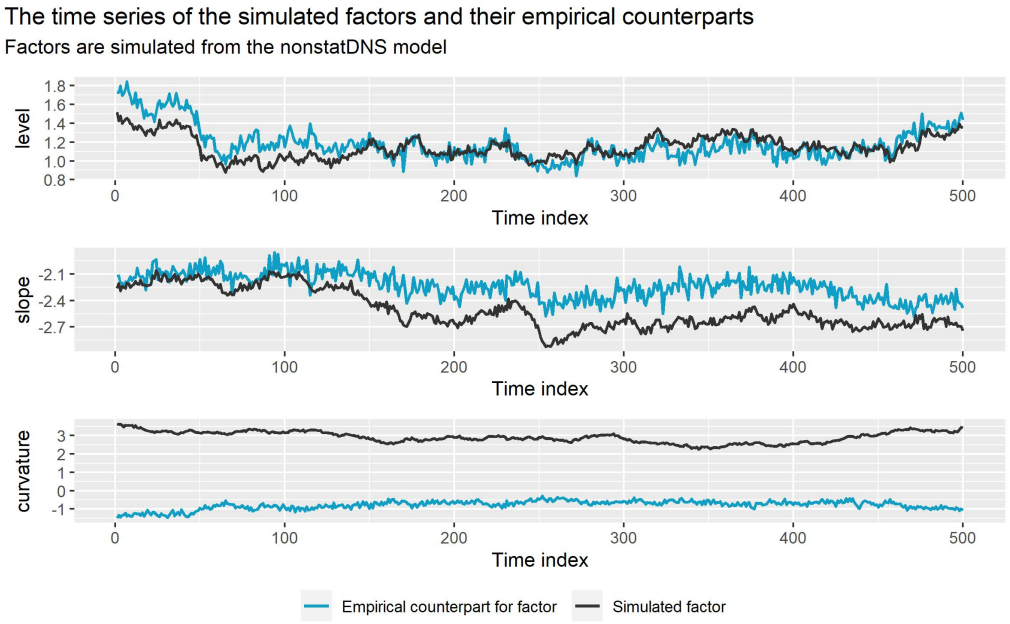


Figure 13: Plot of the series of empirical and simulated factors for one of the 36 simulations of the nonstatDNS model. Data is simulated under the assumption that the first-order differences of the true factors are stationary.

The stationarity of the empirical factors is further investigated by the results of the ADF tests. In Table 13 it is shown that most of the empirical factor series are not stationary. In the three rightmost columns of the table are the ADF test results for the factor series after a linear first-order difference transformation. These results imply that all transformed series follow a stationary autoregressive process. This implies that the stationarity of the empirical factors is a good proxy of the stationarity of the true factors.

	Level (%)	Slope (%)	Curvature (%)	Level (%) Differenced	Slope (%) Differenced	Curvature (%) Differenced
Stationary	0	33.3	5.56	100	100	100
Nonstationary	100	66.7	94.44	0	0	0

Table 13: The percentage of series of empirical factors that follow a stationary or nonstationary process given that the first order differences of the true factors follow a stationary autoregressive process of order one. Stationarity is determined by the ADF test with a critical value of -3.44 and a p-value lower than 0.05. The three rightmost columns contain the results for the series after a linear first-order difference transformation. All data is simulated from the nonstatDNS model, 100% corresponds to all 36 simulations.

The next step is to analyze the systematic relationship of the empirical factors through the ACF and PACF. It is assumed that the first-order differences of the true factors follow an AR(1) process. The ACF and PACF of the first-order differences of the empirical factors are compared to the ACF and PACF of an AR(1) process.

For this comparison, recall that the ACF of an AR(1) process tails off and the PACF of this process cuts off after one lag, as in Figure 12.

In Figure 14 are the ACF and PACF plots of the transformed empirical level series. Both the ACF and PACF tail off. That means that the ACF corresponds to an ACF of an AR(1) process. The PACF tails off and thus differs from the PACF of an AR(1) process, which cuts off after one lag. This means that the ACF and PACF of the empirical factors are not consistent with the true underlying structure of the true level factor from the nonstatDNS model, because the true factors are assumed to follow an AR(1) model. The ACF and PACF of the empirical factors correspond to an ARMA model.

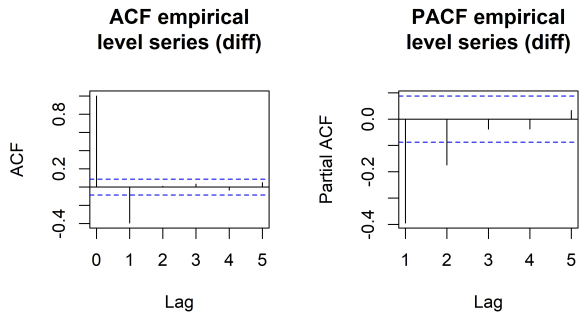


Figure 14: The ACF and PACF plot of the series of empirical counterparts for the level factor after a linear first-order difference transformation. Results for data simulated from the nonstatDNS model in which it is assumed that the first-order differences of the true factors follow an AR(1) process.

In order to keep the presentation of the results focused and clear, only one example of the time series, ACF, and PACF plots are presented in this section. The ACF and PACF plots of all simulations are checked for consistent results. All 36 ACF and PACF plots of the empirical factors for the statDNS model correspond to an AR(1) model. All 36 ACF and PACF plots of the transformed empirical level factor for the nonstatDNS model correspond to an ARMA model, where the number of lags differs between 2 and 4. This implies that the examples presented in this section are a good representation of the results of all simulations.

5.1.2 Correlation between two interest rates

The objective of this study is to investigate the relation between the mutual correlation of the empirical counterparts and the estimated DNS factors for two correlated interest rate curves, clarifying that the correlation being measured is between the factors themselves rather than between the two methods.
This study only consists of data simulation and analysis of the data structure. The model fitting and prediction are not included. For the simulations, I simulated 36 data sets from the sim-statDNS model and 36 data sets from the sim-nonstatDNS model. The empirical counterparts for the factors are computed for each simulated set of observations. The results are split into two cases, based on the stationarity of the true DNS factors that are used to simulate the data.

Stationary factor dynamics The cross-correlation of the empirical factors is presented in Figure 15. In these plots, it can be seen that there is significant cross-correlation between the series of empirical factors. However, these CCF plots do not correspond to the underlying dynamics of the true factors. The true factors are modeled from a model with cross-correlation up until the first lag. That is, the zeroth and first lag of both series (lag -1, lag 0, and lag 1) should have a significant correlation and all other lags should not. Although the empirical factors in Figure 15 show some cross-correlation, the correlation does not represent the structure of the true underlying factors.

Nonstationary factor dynamics In this setting, the first-order differences of the true DNS models follow a stationary process. According to the results of the previous simulation study, the DNS model for the factors in this setting will be transformed into a model on the first-order difference transformation. For that reason, the CCF of the transformed time series is evaluated. The cross-correlation of the transformed empirical factors is presented in Figure 16.
The results in this plot show that there is some significant cross-correlation between the transformed factors. But again, the cross-correlation in the CCF plots does not correspond to the structure of the true factors.

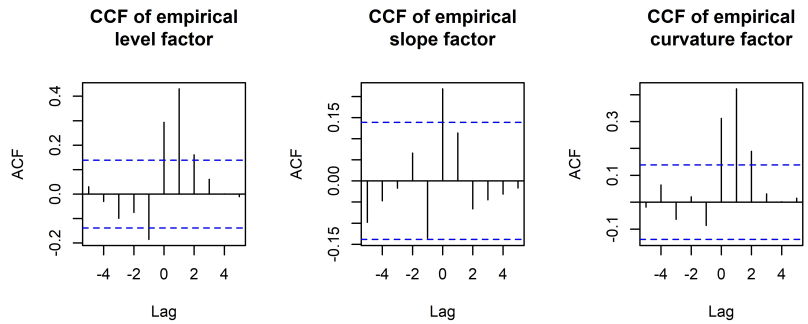


Figure 15: The cross correlation plots for the empirical factors. Results for data of which the true factors are stationary and mutually correlated.

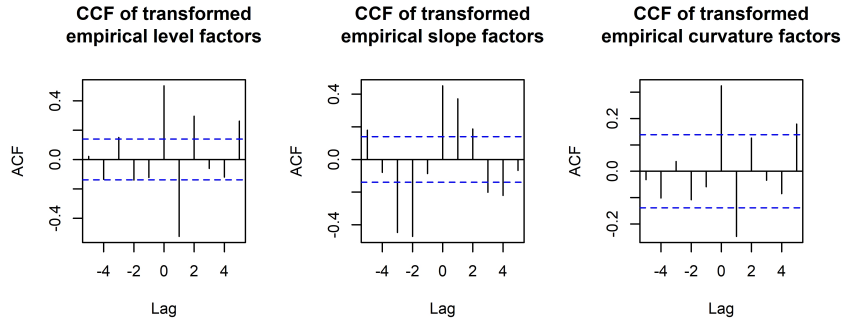


Figure 16: The cross correlation plots for the transformed empirical factors. Results for data of which the first order differences of the true factors are stationary and mutually correlated.

5.1.3 Intermediate conclusion for the empricial counterparts

Based on the simulation studies in this section, the relation between the empirical factors and the true factors is specified by the following.

- The actual values of the empirical factors are not equal to the true factor values (Figure 10, Figure 13).
- The stationarity of the empirical factors is a very good representation of the stationarity of the true factors. This conclusion holds for factors that are stationary and for factors that are nonstationary (Table 12, Table 13).
- If the true factors are a stationary autoregressive process, the autocorrelation of the empirical factors is a good representation of the autocorrelation of the true factors. The same holds for the partial autocorrelation (Figure 12).
- If the first-order differences of the true factors are a stationary autoregressive process, the autocorrelation of the empirical factors is not a good representation of the autocorrelation of the true factors. The same holds for the partial autocorrelation (Figure 14).
- The cross-correlation of the empirical factors is not a good representation of the cross-correlation of the true factors (Figure 15, Figure 16).

5.2 Model behavior

The second group consists of two simulation studies. The first study focuses on the estimation of $\hat{\lambda}_{MLE}$. The second study aims to learn the model behavior related to the parameter estimates for both λ and the factors $\beta_{t,i}$.

5.2.1 The estimation of $\hat{\lambda}_{MLE}$

The focus of this simulation study is on the estimation process for λ. The parameter λ in the DNS models defines the weight distribution on the factors. For that reason, λ is of major importance for the reconstruction of the predicted curves from the predicted factors. With this simulation study, I want to see how a value for λ is estimated by the DNS model.

For this simulation study, I simulate 36 different data sets from the statDNS model. For each of these data sets, I estimate a statDNS model. The estimation of λ is analyzed by looking at the final estimates for $\hat{\lambda}_{MLE}$ as well as looking at the intermediate estimates produced by the iterative BFGS optimization.

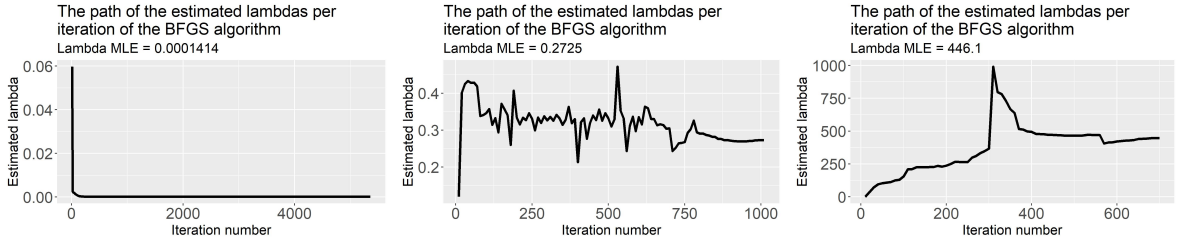
The estimated values The evaluation of the model behavior starts with the estimated values $\hat{\lambda}_{MLE}$. In Table 14 are the descriptive statistics for the set of 36 estimates in this simulation study. At least 50% of the estimated values is in the range 0.26-0.30, which is an overestimation of the true value of 0.1195. There are also values for $\hat{\lambda}_{MLE}$ that are extremely large or extremely small, for example the minimum and maximum values in Table 14.
The presence of multiple local optima in the likelihood surface is indicated by the inconsistency of the maximum likelihood estimate for λ across different data sets with the same true λ.

	Q1	Median	Q2	Min	Max
$\hat{\lambda}_{MLE}$	0.267	0.278	0.296	1.41×10^{-4}	4.46×10^{2}

Table 14: The descriptive statistics for the estimated values for λ. The true value for λ is 0.1195.

The estimation procedure Another result in this simulation is the path of λ throughout the BFGS algorithm iterations. In Figure 17 are three examples of the values for λ estimated by the BFGS iterations.
The leftmost plot is the path of λ in a simulation where $\hat{\lambda}_{MLE}$ is very small. In this plot, you can see that after a very few iterations, λ is already very small, and the value remains small for the rest of the iterations. Furthermore, the number of iterations is high relative to the other two plots. This result is consistent for all models that estimate λ small.
The rightmost plot is the path of λ in a simulation where $\hat{\lambda}_{MLE}$ is very large. In this plot, you can see that the value of λ increases fast in the first few iterations, and remains large for the rest of the iterations. This result is also consistent for all models that estimate λ large.
In the middle plot, you see that all values for λ stay in a range between 0.1-0.5. This path for λ represents all models where λ is estimated close to the true value.

Based on these results, it can be said that the order of the estimated value for λ is determined already by the first few iterations of the BFGS algorithm. By looking at the path for λ, it can be predicted if the value for λ will be very large, very small, or close to the value of λ_{true}.



(a) (b) (c)

Figure 17: The path of λ throughout the BFGS algorithm iterations for three different simulations.

In the results of this simulation study, there is no relation found between the simulated data sets and the estimated λ's for the three models. For one data set, it can happen that two models estimate λ close to the

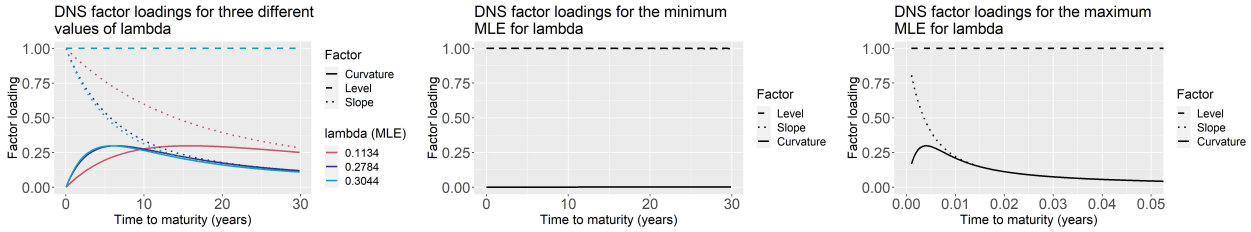
true value, but the third model estimates λ very small or very large.

5.2.2 Interaction between λ, the factors and the factor loadings

In the previous simulation study it has been showed that there can be the estimated λ's can take many different values. In this simulation, the effect of these estimates on the loadings, factors and interest rate curves will be investigated. That is, this simulation focuses on the interaction between the estimated $\hat{\lambda}$, the corresponding factor loadings, and the estimated factors $\hat{b}_t$. The objective is to relate the combination of λ and b_t to the model behavior and performance. By learning this relation, the model choices can be adjusted to the expected model behavior.

For this simulation study, I simulate 36 different data sets from the statDNS model. For each of these data sets, I estimate a statDNS model. The model behavior is analyzed by the procedure in Section 4.3.4. This includes a plot of the estimated factor loadings in combination with the estimated λ, a plot of the estimated factors $\hat{b}_t$, a plot of the interest rate curves implied by these factors and factor loadings, and all plots and tests for the error distribution.

The shape of the factor loadings $\hat{\lambda}_{MLE}$ has effect on the estimated and predicted interest rates through the factor loadings, see Equation 2. The value of λ determines the weight, or loading, on the factors for each time to maturity. In Figure 18 are plots of the shapes of these loadings for different values of λ.



(a) Three different values for $\hat{\lambda}_{MLE}$ (b) The minimum estimated $\hat{\lambda}_{MLE}$ (c) The maximum estimated $\hat{\lambda}_{MLE}$

Figure 18: The factor loadings for different values of the estimated $\hat{\lambda}_{MLE}$.

In Figure 18a is a plot of three different values for $\hat{\lambda}_{MLE}$. The shape of all loadings in this figure is equal to the shapes of the loadings in other literature (Diebold & Li, 2006), (Koopman et al., 2007).

In Figure 18b is the plot of the factor loadings for the minimum $\hat{\lambda}_{MLE}$ in this simulation study. As explained in Section 3.1.1, a very small λ implies a very slow decaying slope loading and a right skewed curvature loading. The slop factor is a flat line close to one, which means that the loading decays so slow that the slope loading is approximately one for all times to maturity between 0 and 30 years. The curvature loading is a flat line close to zero, which means that the skew of the loading is so far to the right that the curvature loading is approximately zero for all times to maturity in the range of 0 to 30 years.
The interest rate curves modeled with these loadings will not depend on the curvature loading and will depend on the level and slope factor with approximately equal weight for all times to maturity. This implies linear dependence in the level and slope factor, which leads to unreliable and unstable factor estimates and reduced model interpretability.

As explained in Section 3.1.1, a large λ implies a fast decay in the slope loading and a curvature loading that is right-skewed. In Figure 18c is the plot of the factor loadings for the maximum $\hat{\lambda}_{MLE}$ in this simulation study. The plot only shows the loadings for very small times to maturity of 0 to 0.05 years. In this plot it can be seen that the slope and curvature loadings are approximately zero for times to maturity larger than 0.04 years.
Zero weight on the slope and curvature factor implies that the interest rate only depends on the level factor. Because the weight on the level factor is constant across all times to maturity, the resulting interest rate curve will be flat and equal to the value of the level factor except for the very short end[9], which depends on the other two factors as well.

[9]The 'short end' of an interest rate curve is the part of the curve that has low values for the time to maturity. In this thesis, 'short' implies that the time to maturity is less than 3 years, and 'very short' implies that the time to maturity is less than 3 months.

The interaction with the estimated factors The shape of the interest rate curve in the DNS model is on one hand determined by λ and the factor loadings and on the other hand by the factors themselves. The next step in the evaluation of the model behavior is to look at the estimated factor values $\beta_{1,t}$, $\beta_{2,t}$ and $\beta_{3,t}$. In Table 15 are the descriptive statistics for the estimated factor values. To see if there is any interaction between the estimated $\hat{\lambda}_{MLE}$ and the factors, the estimated factors are distinguished by the estimated $\hat{\lambda}_{MLE}$ of the same model.

Factor	Quantile for $\hat{\lambda}_{MLE}$	Q1 $(\beta_{i,t})$	Median $(\beta_{i,t})$	Q2 $(\beta_{i,t})$	Min $(\beta_{i,t})$	Max $(\beta_{i,t})$
level	Q1	1.05	1.11	1.18	0.872	83.1
level	Q2	1.11	1.14	1.18	0.878	1.36
level	Q3	1.10	1.13	1.17	0.932	1.37
level	Q4	1.02	1.07	1.13	0.78	1.36
slope	Q1	-1.44	-1.38	-1.28	-83.1	-0.487
slope	Q2	-1.42	-1.4	-1.37	-1.60	-1.18
slope	Q3	-1.42	-1.39	-1.34	-1.65	-0.501
slope	Q4	-671	-32.2	-1.36	-2370	25.5
curvature	Q1	-0.429	-0.369	0.922	-4.51	384
curvature	Q2	-0.561	-0.528	-0.498	-0.823	-0.266
curvature	Q3	-0.626	-0.585	-0.549	-0.878	-0.279
curvature	Q4	-19	-3.18	-0.7	-274	-0.329

Table 15: The descriptive statistics for the estimated factors distinguished by the corresponding $\hat{\lambda}_{MLE}$: Q1 ($\lambda < 0.267$), Q2 ($0.267 \leq \lambda < 0.278$), Q3 ($0.278 \leq \lambda < 0.296$), Q4 ($\lambda \geq 0.296$).

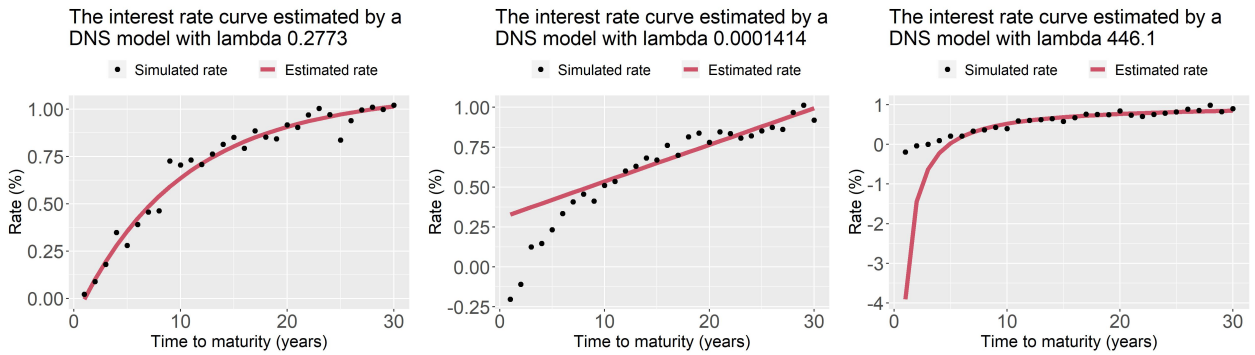
To start with the level factor, the results in Table 15 imply that almost all estimated level factors are in the range of 1.02-1.18. The results do not demonstrate a clear relationship between estimates for the level factor and the estimated $\hat{\lambda}_{MLE}$. The absence of this relationship can be explained by the factor loading for the level factor, which is equal to 1 and thus independent of λ.
The maximum estimate for the level factor of 83.1 in the case that $\hat{\lambda}_{MLE} \in Q1$ is remarkable, because it is much larger than all other estimates for the level factor. The model that estimated this factor value has a $\hat{\lambda}_{MLE}$ equal to 0.659×10^{-4}. This λ value implies factor loadings with a shape as in Figure 18b and thus leads to linear dependence in the level and slope factor and potentially unstable factor estimates. This explains the outlier for the level factor.

Next are the slope factors in the middle rows of Table 15. These results show that there is interaction between the estimated factors and the λ. If $\hat{\lambda}_{MLE} \in Q4$, i.e. if λ is large, the estimated slope factors become extremely negative. This again can be explained by the factor loadings. For large λ's, the factor loadings look like the plot in Figure 18c, which leads to very small weight on the slope factor. The slope factor reacts on this small weight by taking more extreme values.
Again, there is a remarkable value of -83.1 for the minimum slope factor in the case that $\hat{\lambda}_{MLE} \in Q1$. The model that estimated this factor is the same model that estimated the remarkable level factor of 83.1. This is another example where a very small λ results in unstable factor estimates.

The results for the curvature factors are in the bottom rows of Table 15. These results show that there is interaction between the estimated factors and the λ. For smaller λ's, i.e. $\hat{\lambda}_{MLE} \in Q1$, the curvature factors generally have larger estimates. The maximum value of 384 will be used to explain this interaction. If λ is very small, the curvature loading will be approximately zero, see Figure 18b. This leads to unstable estimates for the curvature factors.
For larger λ's, i.e. $\hat{\lambda}_{MLE} \in Q4$, the curvature factors generally have more negative estimates. This again can be explained by the factor loadings. For large λ's, the factor loadings look like the plot in Figure 18c, which leads to very small weight on the curvature factor. The curvature factor reacts on this small weight by taking more extreme values.

The shape of the interest rate curves The final step in the model behavior is the evaluation of the shape of the interest rate curves. In Figure 19 are three plots of examples of interest rate curves estimated by the statDNS model. The curve in Figure 19a is estimated by a model that has $\hat{\lambda}_{MLE}$ close to the median of all estimates. The shape of this curve closely follows the structure of the data points.

In Figure 19b is a plot of the curve estimated by the model that has the lowest estimate for $\hat{\lambda}_{MLE}$. This curve has the shape of a linear line and does not follow the structure of the data. This is due to the linear dependence of the level and slope factor and the zero weight on the curvature factor, all caused by the factor loadings. Such an estimated interest rate curve does not have normally distributed residuals. That means that



(a) The interest rate for $\hat{\lambda}_{MLE}$ close to the true value of λ

(b) The interest rate when $\hat{\lambda}_{MLE}$ is extremely small

(c) The interest rate when $\hat{\lambda}_{MLE}$ is extremely large

Figure 19: Three examples of interest rate curves estimated by the statDNS model.

the assumption on the residuals is violated.

The curve in Figure 19c is estimated by the model that has the largest estimate for $\hat{\lambda}_{MLE}$. This curve is close to the data except for the short end. The factor loadings for this model in Figure 18c show that the slope and curvature loading only imply weight on the very short end. The low weight is compensated by the estimation of larger slope and curvature factors, which was shown in Table 15. This compensation implies that the majority of the curve is close to the data. The deviation of the curve from the data on the short end is due to the increase in the factor loadings on this short end.

5.2.3 Intermediate conclusion for the model behavior

The model behavior and interaction between λ, the factor loadings, and the factors are summarized as follows.

- The estimates for $\hat{\lambda}_{MLE}$ can take a wide spread of values. This suggests that there are multiple local optima in the likelihood surface.
- The order of $\hat{\lambda}_{MLE}$ can be predicted after the first few iterations of the BFGS-algorithm. If the intermediate estimates for λ are very small or very large, the final estimate will also be very small or very large (Figure 17). That is, if the BFGS algorithm estimates λ at an extreme value once, all subsequent estimates throughout the algorithm will be extreme.
- For $\hat{\lambda}_{MLE}$ not extremely small or extremely large, the shape of the factor loadings is equal to the shape of the loadings in the article of Diebold and Li (2006). Furthermore, the shape of the estimated interest rate curve is close to the data and also equal to the shape of the curves in the article of Diebold and Li (2006). That implies that if $\hat{\lambda}_{MLE}$ is not extremely small or extremely large, it is validated that the statDNS produce reliable estimates for the interest rate curves.
- An extremely small $\hat{\lambda}_{MLE}$ implies that
  - The level and slope factor are linearly dependent (Figure 18b), which leads to unstable factor estimates (Table 15) and misleading results,
  - The loading on the curvature factor is approximately zero, leading to estimated interest rate curves that are flat,
  - The shape of the interest rate curves does not correspond to the data (Figure 19b), leading to not normally distributed residuals, and
  - The shape of the interest rate curve does not correspond to the shape of the interest rate curves that are mentioned in the articles of Diebold and Li (2006), (Koopman et al., 2007).

  For these reasons, models with an extremely small λ will not be used to estimate and predict the interest rate curves. In this thesis, all λ's smaller than 0.01 are categorized as extremely small.
- An extremely large $\hat{\lambda}_{MLE}$ implies that
  - There is approximately zero weight on the slope and curvature loading, except for the short end of the curve (Figure 18c),

- This is compensated by the estimates for the slope and curvature factors, that take more extreme values (Table 15),
- The compensation leads to a good fit of the curve to the data, except for the short end (Figure 19c).
- The shape of the interest rate curves deviates from the data at the short end (Figure 19c). This leads to heavy tailed residuals and thus a violation of the assumption of normal distributed errors.
- The shape of the interest rate curve does not correspond to the shape of the interest rate curves that are mentioned in the articles of Diebold and Li (2006), (Koopman et al., 2007).

For these reasons, models with an extremely large λ will not be used to estimate and predict the interest rate curves. In this thesis, all λ's larger than 10 are categorized as extremely large.

5.3 Model choices

The third group of simulation studies consists of three different studies that together have the objective to understand the effect of certain data preprocessing steps and initial value choices. The first study investigates the effect of the initial value for λ, the second study investigates the effect of missing observations, and the third study investigates the effect of preprocessing the bond observations with splines.
All three studies are performed using only the statDNS model and it is assumed that the results hold for the three other variants of the DNS models.[10]

5.3.1 Initial value for λ

One of the model choices is the initial value for λ. As explained in Section 3.4.3, there are various methods to choose the initial λ. The objective of this simulation study is to examine the effect of the initial value of λ on the estimated $\hat{\lambda}_{MLE}$ in the DNS models.
This study consists of the following steps: data simulation, model fitting and predicting, and assessment of the model behavior. For this simulation study, I use the same simulated data sets and estimated models as for the previous simulation study. That is, there are 12 different simulated data sets from the statDNS model with all three statDNS models estimated per data set. On each simulated data set, I estimate three statDNS models: one with an initial λ equal to the true λ, one with an initial λ that is two times larger than the true λ, and one with an initial λ that is half the value of the true λ. The estimated models are compared by their value for $\hat{\lambda}_{MLE}$.

The first results are the estimated values for $\hat{\lambda}_{MLE}$ per initial value $\lambda_{initial}$. In Table 16 are the first quantile, median, third quantile, minimum, and maximum for all estimated $\hat{\lambda}_{MLE}$, grouped by the initial value $\lambda_{initial}$. When compared to the true value of λ, it becomes apparent that the majority of the models tend to overestimate λ. This observation is supported by the median values of the estimated $\hat{\lambda}_{MLE}$. Looking at the first quantiles, a model with a larger initial λ exhibits a greater tendency for overestimation compared to the other models.

Not all estimated $\hat{\lambda}_{MLE}$'s are close to the true value of λ. The minimum and maximum values in Table 16 imply that the DNS models can estimate λ much smaller or larger than λ_{true}. Also, a small initial λ never leads to a very large $\hat{\lambda}_{MLE}$ and a large initial λ never leads to a very small $\hat{\lambda}_{MLE}$.

Initial value for λ	Scale factor w.r.t. λ_{true}	Q1 $\hat{\lambda}_{MLE}$	Median $\hat{\lambda}_{MLE}$	Q2 $\hat{\lambda}_{MLE}$	Min $\hat{\lambda}_{MLE}$	Max $\hat{\lambda}_{MLE}$
0.05975	0.5	0.1194	0.1218	0.266	0.0001414	0.2971
0.11950	1	0.1195	0.1223	0.2864	0.108	446.1
0.23900	2	0.1203	0.1945	0.2886	0.1163	137.1
All models	-	0.1197	0.123	0.2796	0.0001414	446.1

Table 16: The results for the estimated $\hat{\lambda}_{MLE}$'s estimated on data with $\lambda_{true} = 0.1195$. The results are split per value for $\lambda_{initial}$. The bottom row includes the results aggregated over all initial values for λ.

The second finding is related to the number of iteration of the optimization algorithm for the different initial λ's. In Figure 20 are the number of iterations for the BFGS algorithm per initial value of λ. The initial value of λ is indicated by the scale factor w.r.t. the true value of λ. In the plot, it can be seen that the median number of iterations is approximately equal for all initial λ's. But models with an initial λ not equal to the true value of λ have a larger number of iterations more often, which is indicated by the third quartile boundaries in the boxplot.
One iteration of the BFGS algorithm on my computer takes approximately 1 second for the models in this simulation. This implies that 1000 iterations take approximately 17 minutes, whereas 2000 iterations take more than half an hour. This implies that an initial value of λ close to the true value of λ can save a lot of time in the estimation of the model.

5.3.2 Missing observations

The real data misses observations when the market is closed, for example on weekend days. This simulation study aims to investigate the potential effect of the missing observations in the real data set on the model behavior and performance.
For this simulation study, I use the same simulated data sets and estimated models as for the previous simulation study. That is 12 different simulated data sets from the statDNS model and three statDNS models estimated per data set. Additionally, I construct 12 new data sets by removing all 'weekend observations' from the simulated

[10] Some informal simulation studies for the other three models are performed to check if this assumption holds. These results are not included in this thesis.

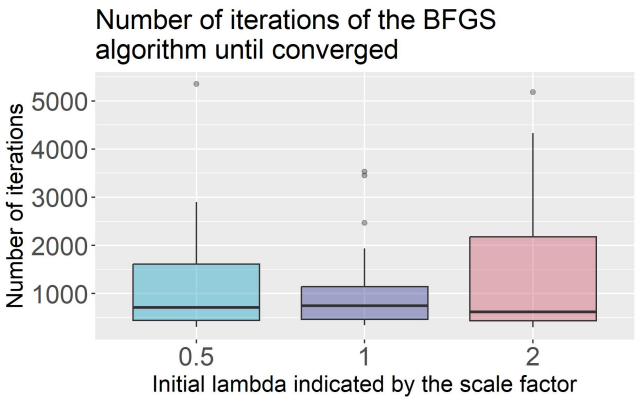


Figure 20: The number of iterations of the BFGS-optimization algorithm until the convergence criterion is met.

data sets. For each of these 12 data sets, I again estimate the statDNS models. The focus of the model behavior will be on the accuracy and uncertainty of the estimated and predicted factors and interest rates.

Model behavior For the models that are estimated on the full data sets, 7 out of 36 had an estimate of λ outside the range of 0.1-10. For the models estimated on the data sets excluding the weekend observations, there were also 7 out of 36 with λ in that range. This implies that missing observations have no significant effect on the ability of the model to estimate λ.

Model accuracy The second part of the evaluation is the accuracy of the estimated and predicted interest rates. In Table 17 are the rMSE results for the estimated and predicted interest rates.[11] In the table it can be seen that the rMSE of the rates estimated on the data set including weekends is in general slightly larger than the rMSE of the rates estimated on the data set excluding weekends.
The maximum rMSE of the models excluding the weekends is much larger relative to the rMSE in the case of a complete data set. This implies that, although it does not happen very often, the model estimated excluding the weekends could have a very bad accuracy.

An equal conclusion can be drawn for the predicted interest rates. The rMSE of the predicted rates is in general larger if there are no weekend observations. The difference between the rMSE of the two cases is larger for the predicted rates than for the estimated rates.

Rate	Data	Q1 (rMSE $\times 10^{-2}$)	Med (rMSE $\times 10^{-2}$)	Q3 (rMSE $\times 10^{-2}$)	Min (rMSE $\times 10^{-2}$)	Max (rMSE $\times 10^{-2}$)
Estimated	Incl. weekends	4.35	4.77	5.22	2.73	7.10
Estimated	Excl. weekends	4.38	4.82	5.30	2.71	305
Predicted	Incl. weekends	5.79	7.82	12.7	3.15	183
Predicted	Excl. weekends	6.07	9.27	17.6	3.15	302

Table 17: The descriptive statistics for the rMSE of the interest rate curves estimated and predicted by the statDNS model on the data sets including and excluding the weekend observations.

Model uncertainty The final step in the evaluation is the comparison of the uncertainty implied by the models. In Table 18 are the standard deviations for the interest rates, which are estimated by the Kalman filter. In Table 18 it can be seen that the standard deviation estimated by the model excluding the weekends is larger than the standard deviations estimated by the model including the weekends.
It is remarkable that more than 25% of the standard deviations in the case of missing weekend observations is estimated at a value of infinity. For the predicted interest rates, 100% of the standard deviations are estimated at a value of infinity in both the case of data including weekend observations and the case of data excluding weekend observations. That implies that the DNS model in this setting in general does not produce usable estimates for the standard deviation of the predictions. Therefore, there will be no prediction intervals constructed from the DNS models.

[11] All weekend dates are excluded from the results, because including them leads to an unfair comparison of the estimated rates.

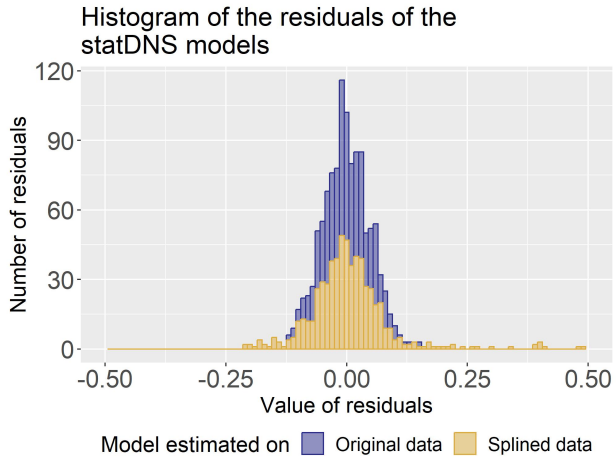
Data	Q1 ($\hat{\sigma}^2$)	Med ($\hat{\sigma}^2$)	Q3 ($\hat{\sigma}^2$)	Min ($\hat{\sigma}^2$)	Max ($\hat{\sigma}^2$)
Incl. weekends	5.67×10^{-3}	6.94×10^{-3}	7.60×10^{-3}	8.78×10^{-3}	1.96×10^{4}
Excl. weekends	7.69×10^{-3}	1.07×10^{-2}	∞	4.75×10^{-3}	∞

Table 18: The descriptive statistics for the variances for the interest rates estimated by the statDNS model for the data sets including and excluding the weekend observations.

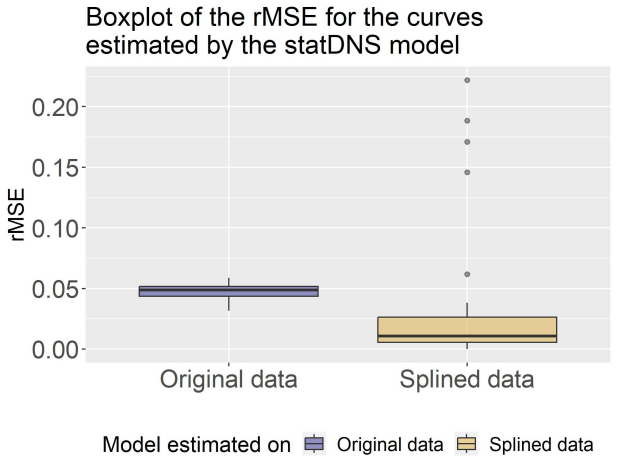
5.3.3 Distribution of the observations along the time to maturity axis

In Section 2.1.2 it was explained that one of the preprocessing steps for the empirical bond observations was to estimate the observations with a fixed time to maturity using splines. This choice was motivated by the computational issues when the observations have different times to maturity at each date. The objective of this simulation study is to investigate the effect of this choice on the model performance.
For this simulation study, 50 data sets are simulated from the statDNS model. The observations in this data set have various values for the time to maturity. The simulated observations will be preprocessed using splines, as described in Section 2.1, which result in a second data set. The statDNS model is estimated on both data sets. Because of the computational issues, the statDNS model can not be estimated on the set of observations with various values for the time to maturity. Therefore, a statDNS will be estimated on both data sets using the two-step procedure. The models will be compared by their fit on the simulated data set with various times to maturity, i.e. the first data set, to see the effect of splining.

In Figure 21a is a histogram of the residuals of the statDNS models. The blue bars represent the residuals of the statDNS models on the original data. The yellow bars represent the residuals of the statDNS models estimated on the splined data. In this plot it can be seen that both sets of residuals are normally distributed. The estimated mean and standard deviations of the residuals are presented in Table 19. Both distributions have a mean of approximately zero. The standard deviation for the residuals increases by more than 200% if the data is preprocessed by the splines.



(a) Histogram of the residuals of the statDNS models estimated on the two different data sets



(b) Boxplot of the rMSE of the curves estimated by the statDNS model on the two different data sets

Figure 21: Results of the statDNS model estimated on the original simulated data set and the data set after preprocessing.

	$\hat{\mu}$	$\hat{\sigma}$
Original data	0	0.0489
Splined data	0.004	0.107

Table 19: The mean and standard deviation of the residuals of the statDNS model.

Although the standard deviation of the residuals is larger if the data is splined, the rMSE of the estimated curves is not larger for the splined data set. In Figure 21b are the boxplots for the rMSE's of the estimated interest rate curves. In this plot it can be seen that the rMSE of the splined data is generally lower than the rMSE for the original data. In other words, although the variance of the errors increases by splining the data, the overall fit of the DNS model improves.

5.3.4 Intermediate conclusion

Based on the results in this simulation study, the effect of the initial λ on the estimation of $\hat{\lambda}_{MLE}$ is summarized by the following points.

- Different initial values for λ produces different estimates for $\hat{\lambda}_{MLE}$ for the same data set. This implies that there are multiple local maxima in the log-likelihood surface and the initial value for λ determines to which local maxima the BFGS-algorithm converges.
- A smaller initial value never leads to extremely large estimated values, and a larger initial value never leads to extremely small estimated values.
- A model with an initial λ close to the true λ has fewer iterations of the BFGS algorithm and thus a lower computational time, relative to a model with an initial λ further away from the true λ
- The intermediate estimates for λ throughout the iterative optimization can be used to predict whether or not the $\hat{\lambda}_{MLE}$ will take extremely small or extremely large values.

The missing weekend observations result in a slight increase in the rMSE of the estimated and predicted interest rates. Despite this increase, the weekend observations will not be imputed.
Furthermore, the uncertainty around the interest rate estimates increases by 25%. Independent of missing the weekend observations, the DNS models in this simulation study were not able to produce useful estimates for the variances of the predicted interest rate curves. All variances were estimated to be equal to infinity. That means that no prediction intervals can be constructed for the interest rate curves.

The preprocessing step where the data is splined in a set of fixed maturities has the following effect on the model performance. The variance of the residuals is larger for the model that is estimated on the splined data, relative to the model that is estimated on the original data. Despite the increase in variance, the overall accuracy of the model estimated on the splined data is actually improved, relative to the model estimated on the original data. Based on this result, the preprocessing step of splining the data will be included in the framework.

5.4 Model validation

Three new variants on the DNS model are introduced in this thesis in Section 3.3.2, Section 3.3.1, and Section 3.3.3. These models rely on certain assumptions, such as independent, normally distributed residuals. The objective of this fourth set of simulations is to confirm the validity of these assumptions when using each of these newly proposed models. There are four different simulation studies in this group.
The plots of the factor dynamics, curves, and residuals in this section are based on one single simulation per model. The results in these plots are consistent over all simulations within the simulation study for that model.

5.4.1 Validation of the sim-statDNS model

This simulation study aims to validate the sim-statDNS model by investigating the estimated parameters, the estimated factors, the shape of the interest rate curves, and the distribution, dependency, and variance of the residuals.
For this simulation study, I simulate 12 data sets of observations from the sim-statDNS model. On each simulated data set, a sim-statDNS model is estimated. These models have an initial value for λ equal to the true λ, in order to reduce computational time. The estimated models are assessed by all steps in Section 4.3.4, except for the prediction accuracy. The accuracy will be investigated in the next set of simulation studies.

The estimated values for λ The first result are the estimated values for $\hat{\lambda}_{MLE}$. All estimated values for λ were in the range 0.01-10. That means that all simulations are used for the validation.

Factor dynamics The first step in the validation procedure is the evaluation of the estimated and predicted factors. In Figure 22 is a plot of the simulated, estimated, and predicted level factor corresponding to the swap interest rate curves.
In the plot of the swap level factor, it can be seen that the estimated factors are almost equal to the simulated level factors. Furthermore, the predicted factors are stationary and the prediction intervals cover all simulated factor values. This implies that the sim-statDNS model is capable of estimating and predicting the latent factors for a simulated data set.

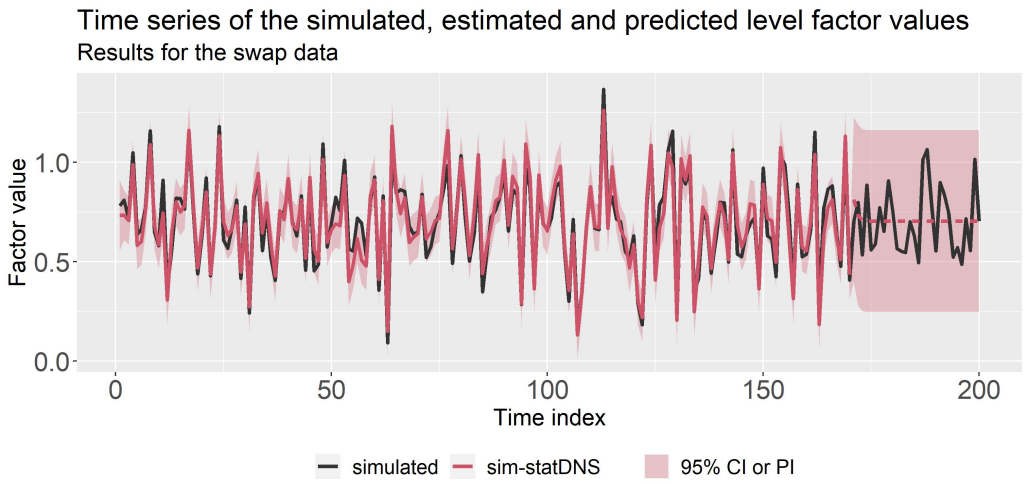


Figure 22: Plot of the series of simulated, estimated and predicted level factors for one of the simulations from the sim-statDNS model.

Interest rate curves The subsequent step in the validation procedure is the evaluation of the estimated and predicted interest rate curves. In Figure 23, there is one example of bond and swap rates estimated by the sim-statDNS model, and there are two examples of bond and swap rates predicted by the sim-statDNS model. The shape of estimated interest rate curves closely follows the structure of the data. The predicted curves slightly deviate from the data, due to over or under-prediction of the factors. Although this deviation of the curves from the data, the shape of the curve closely corresponds to the structure of the data. These results show that the sim-statDNS model is capable of estimating and predicting interest rate curves that capture the structure of the interest rate observations.

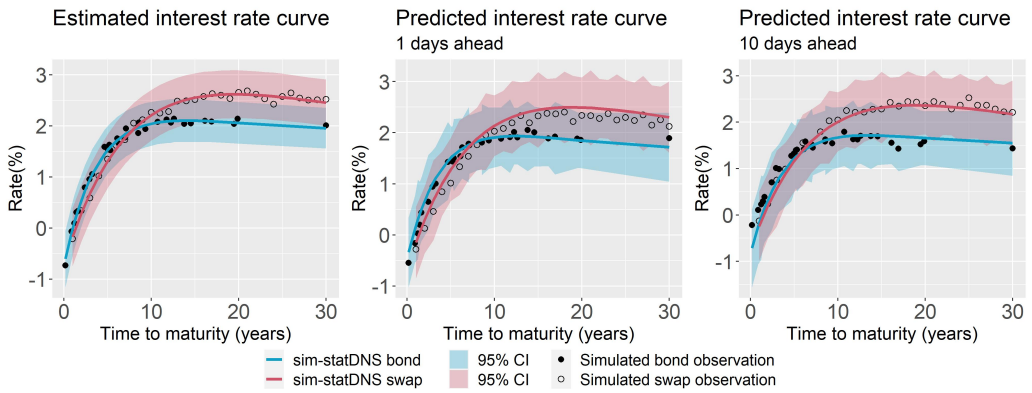


Figure 23: Plot of the estimated and predicted interest rate curves for one of the simulations from the sim-statDNS model.

Distribution, dependence, and variation of the residuals The final step in the model validation is the evaluation of the residuals of the interest rate curves. The residuals are assumed to be independent normally distributed. In Figure 24 are the QQ-plots for the residuals of the interest rate curves in Figure 23. In Table 20 are the results of the Shapiro-Wilk test for each of the estimated and predicted curves in this simulation study. The residuals in the QQ plots are not exactly equal to the theoretical normal distribution (red line), but they are close enough to not reject normality. Furthermore, the results of the Shapiro-Wilk test imply that more than 90% of all estimated and predicted curves have normal distributed error terms. Based on these results, the assumption of normally distributed errors is validated.

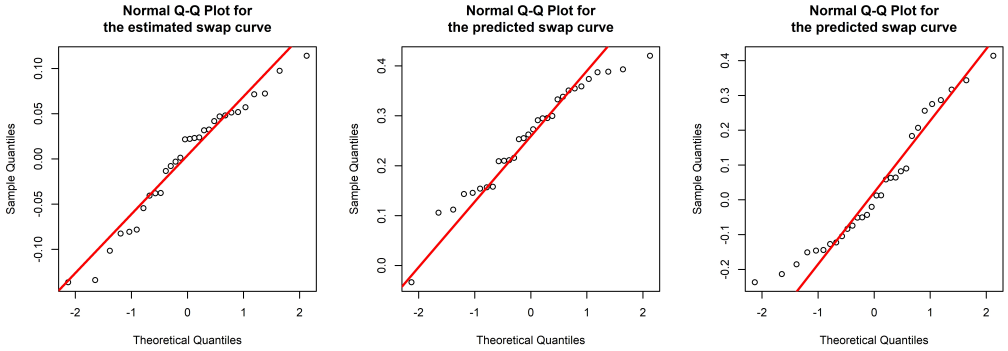


Figure 24: QQ plot of the residuals of the estimated and predicted swap curves in Figure 23.

	Instrument	Total number of curves	# curves that have Gaussian distributed residuals	% of curves that have Gaussian distributed residuals
Estimated	swap	2040	1946	95.4
Predicted	swap	360	325	90.3
Estimated	bond	2040	1924	94.3
Predicted	bond	360	331	91.9

Table 20: The number of interest rate curves estimated and predicted by the sim-statDNS model that have Gaussian distributed residuals according to the Shapiro-Wilk test with a p-value below 0.05.

The homoskedasticity of the errors is evaluated by the Breusch-Pagan test. The results of this test for each estimated and predicted curve are presented in Table 21. More than 90% of the estimated curves and more than 75% of the predicted curves have homoskedastic residuals. Based on this result, it is validated that the sim-statDNS has no heteroskedastic errors for the simulated data set.

	Instrument	Total number of curves	# curves with homoskedastic residuals	% curves with homoskedastic residuals
Estimated	swap	2040	1966	96.4
Predicted	swap	360	314	87.2
Estimated	bond	2040	1872	91.8
Predicted	bond	360	276	76.7

Table 21: The number of interest rate curves estimated and predicted by the sim-statDNS model that homoskedastic residuals according to the Breusch-Pagan test with a critical value of 5.991.

5.4.2 Misleading results of the statDNS model on nonstationary data

If the latent factors are assumed to follow a nonstationary autoregressive process, the assumptions of the statDNS are violated. Applying the model to this data is incorrect and may potentially yield misleading outcomes. This simulation study aims to explore the potential implications of estimating the statDNS model on nonstationary data.
For this study, 36 data sets are simulated from the nonstatDNS model. On each data set, the statDNS model is estimated. When estimating a stationary autoregressive process on a series that is not stationary, two things can happen:

- The estimated autoregressive process is stationary, i.e. the autoregressive coefficient is inside the unit circle. This indicates a case of spurious regression. In this situation, the stationary AR model may still provide parameter estimates and fit the data reasonably well, but the results would lack meaningful interpretation.
- The estimated autoregressive process is not stationary, i.e. the autoregressive coefficient is outside the unit circle. Such a model is not appropriate for analyzing the data and will lead to inaccurate predictions.

The implications of the statDNS model are assessed by their autoregressive processes for the factors that satisfy one of these two options.

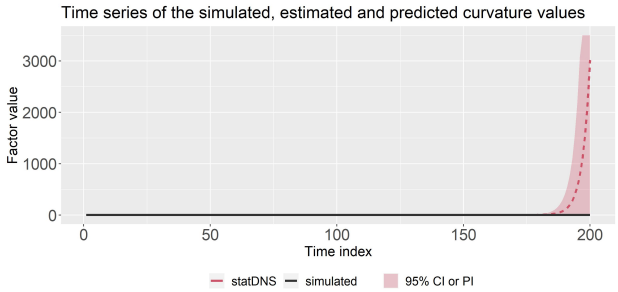
In Table 22 are the results for the estimated autoregressive coefficients for all AR(1) models in the simulations. The evaluation of these results is split in two parts. First, the case $|\hat{\phi}_{ii}| \geq 1$ is discussed, followed by the case $|\hat{\phi}_{ii}| < 1$.

	Total number of series	# series with $\lvert\hat{\phi}_{ii}\rvert < 1$	% of series with $\lvert\hat{\phi}_{ii}\rvert < 1$	# series with $\lvert\hat{\phi}_{ii}\rvert \geq 1$	% of series with $\lvert\hat{\phi}_{ii}\rvert \geq 1$
Level	17	13	76.5	4	23.5
Slope	17	8	47.1	9	52.9
Curvature	17	11	64.7	6	35.3
Total	51	32	62.7	19	37.3

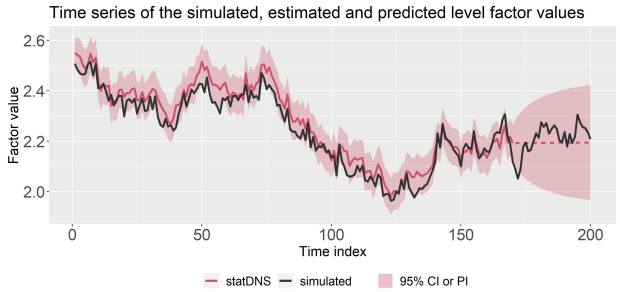
Table 22: The number and percentage of autoregressive coefficients $\hat{\phi}_{ii}$ estimated by the statDNS model that are inside or outside the unit circle. Results for the time series of factors for the data simulated from the nonstatDNS model.

Case of a nonstationary AR(1) model, $|\hat{\phi}_{ii}| \geq 1$: In the results in Table 22 it is shown that 37.3% of all estimated autoregressive coefficients lie outside the unit circle. AR(1) models with $|\hat{\phi}_{ii}| \geq 1$ imply nonstationary processes for the time series. Such a model produces predictions for the factors that diverge. An example of a factor modeled by a nonstationary AR(1) process is visualized in Figure 25a. In the plot it can be seen that the estimated factor values (time index smaller than 170) stay close to the simulated factors, whereas the predicted factors (time index larger than 170) diverge from the simulated factors. The behavior of the estimated factors is due to the update step in the Kalman filter, which rectifies the divergence of the predicted value. Since there is no update step for the predictions, the predicted factors diverge.
The factors estimated and predicted by a nonstationary AR(1) process lead to accurate estimated interest rate curves, but inaccurate predicted interest rate curves. This model should not be used for modeling factors with nonstationary dynamics.

Case of a stationary AR(1) model, $|\hat{\phi}_{ii}| < 1$: In the results in Table 22 it is shown that 62.7% of the estimated autoregressive coefficients lie inside the unit circle. The AR(1) models with such coefficients imply stationary processes. In Figure 25b is an example of a factor that is modeled by a stationary AR(1) model. In this plot it can be seen that the estimated factors closely follow the dynamics of the simulated factors and the predicted factors converge to a stationary value. Also, although the AR(1) model is stationary, the series



(a) Time series of a factor that is modeled by a nonstationary AR(1) model.



(b) Time series of a factor that is modeled by a stationary AR(1) model.

Figure 25: Plot of the time series of two factors that serve as example of the implications of estimating the statDNS on data with nonstationary factor dynamics.

of estimated factors is not stationary[12]. This is due to the update step of the Kalman filter, which rectifies the predicted factor values.
Factors estimated and predicted by a stationary AR(1) model are close to the simulated factors. Therefore they can lead to very accurate estimated and predicted interest rate curves. However, these results are misleading, since the AR(1) model is incorrect and may not be used for modeling factors with nonstationary dynamics.

5.4.3 Validation of the nonstatDNS model

This simulation study aims to validate the nonstatDNS model by investigating the estimated parameters, the estimated factors, the shape of the interest rate curves and the distribution, dependency and variance of the residuals.
The procedure is equal to the procedure for the model validation of the statDNS model, where the sim-statDNS model is replaced by the nonstatDNS model. 36 data sets are simulated from the nonstatDNS model. On each simulated data set, a nonstatDNS model is estimated. The models have an initial value for λ equal to the true λ, in order to reduce computational time.

The estimated values for λ In the simulation study for the nonstatDNS model, all estimated values for λ were in the range 0.01-10. That means that all simulations are used for the validation.

Factor dynamics In Figure 26 is a plot of the simulated, estimated, and predicted level factor. In this plot it can be seen that the estimated factor is close to the simulated factor, which implies that the nonstatDNS model can capture the true factor dynamics. The predicted level factor corresponds to the downward trend in the simulated factors. That means that the model predicts the factor as expected.

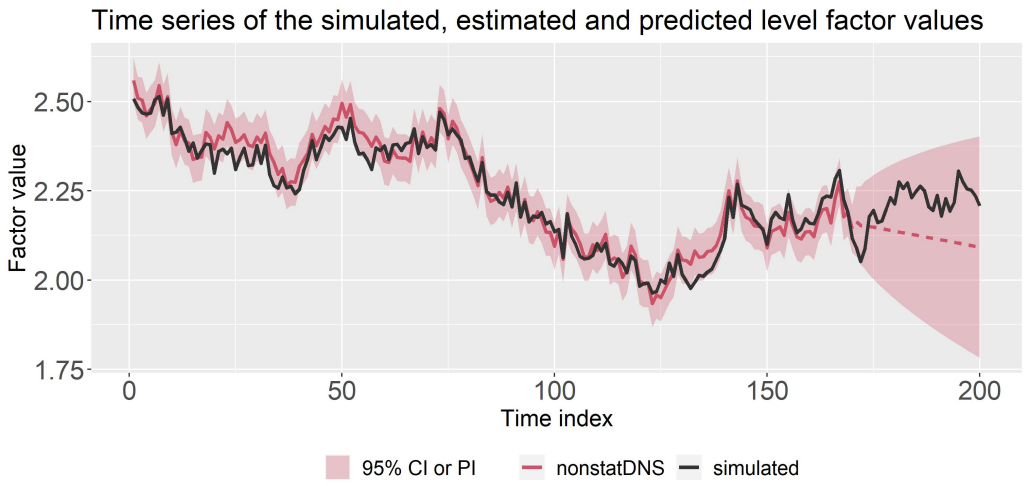


Figure 26: Plot of the series of simulated, estimated and predicted level factors for one of the simulations from the nonstatDNS model.

[12]The series of estimated factors in Figure 25b has an ADF test statistic of -2.24 with a p-value of 0.47, which implies nonstationarity

Interest rate curves In Figure 27 are the plots of the estimated (left) and predicted (middle and right) interest rate curves. The shape of these curves corresponds to the structure of the simulated observations. That implies that the nonstatDNS model is capable of estimating and predicting interest rate curves with a shape corresponding to the shape of the observations.

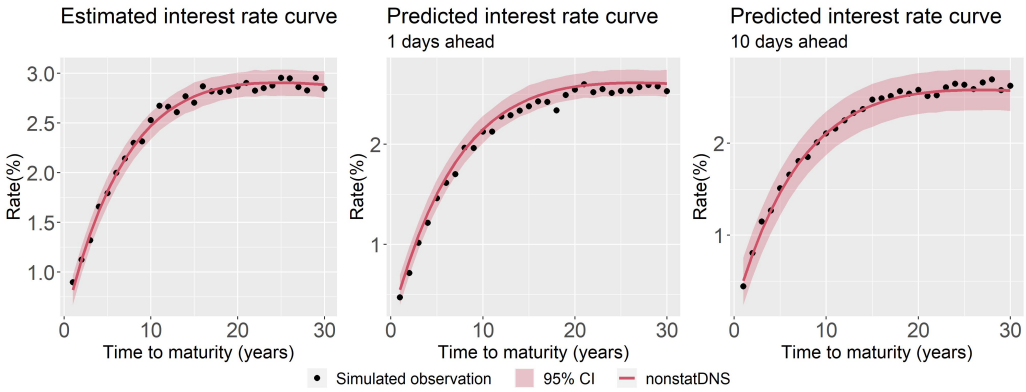


Figure 27: Plot of the estimated and predicted interest rate curves for one of the simulations from the nonstatDNS model.

Distribution, dependence and variation of the residuals The QQ-plots in Figure 28 of the residuals of the curves in Figure 27 can be used to validate the normal distribution of the error terms. In the right QQ-plot it can be seen that the residuals of the estimated curve could imply a light tail in the distribution of the errors, but the deviation is not heavy enough to reject normality.
Another result that implies a normal distribution for the errors, is the result of the Shapiro-Wilk test in Table 23. According to these results, more than 95% of the residuals are normally distributed.

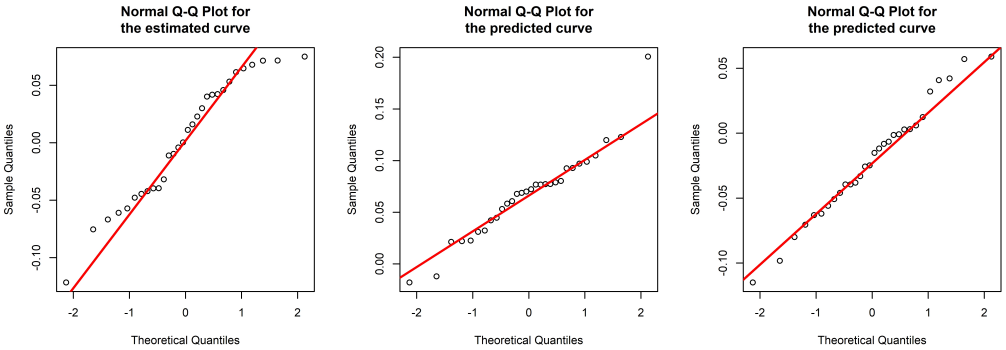


Figure 28: QQ plot of the residuals of the estimated and predicted swap curves in Figure 27.

	Total number of curves	# curves that have Gaussian distributed residuals	% of curves that have Gaussian distributed residuals
Estimated	6120	5813	95.0
Predicted	1080	1041	96.4

Table 23: The number of interest rate curves estimated and predicted by the nonstatDNS model that have Gaussian distributed residuals according to the Shapiro-Wilk test with a p-value below 0.05.

The variation of the residuals is examined with the Beusch-Pagan test, of which the results are presented in Table 24. These results imply that more than 90% of both the estimated and predicted curves have homoskedastic residuals.

	Total number of curves	# curves with homoskedastic residuals	% curves with homoskedastic residuals
Estimated	6120	5886	96.2
Predicted	1080	976	90.4

Table 24: The number of interest rate curves estimated and predicted by the nonstatDNS model that homoskedastic residuals according to the Breusch-Pagan test with a critical value of 5.991.

5.4.4 Validation of the sim-nonstatDNS model

This simulation study aims to validate the nonstatDNS model by investigating the same points as both previous model validation simulation studies. The procedure is equal to the procedure for the model validation of the statDNS model, where the sim-statDNS model is replaced by the sim-nonstatDNS model. The estimated models are assessed by all steps in Section 4.3.4, except for the prediction accuracy. The accuracy will be investigated in the next set of simulation studies.

The estimated values for λ In the simulation study for the sim-nonstatDNS model, all estimated values for λ were in the range 0.01-10. That means that all simulations are used for the validation in this section.

Factor dynamics In Figure 29 is a plot of the simulated, estimated and predicted level factor corresponding to the swap data. In this plot, it can be seen that the estimated factor values are close to the true factor values, which implies that the model is capable of estimating factors with the same structure as the true factors. Furthermore, the predicted factor values follow the slightly downward-sloping trend of the simulated values, implying that the predictions for the level factor are as expected.

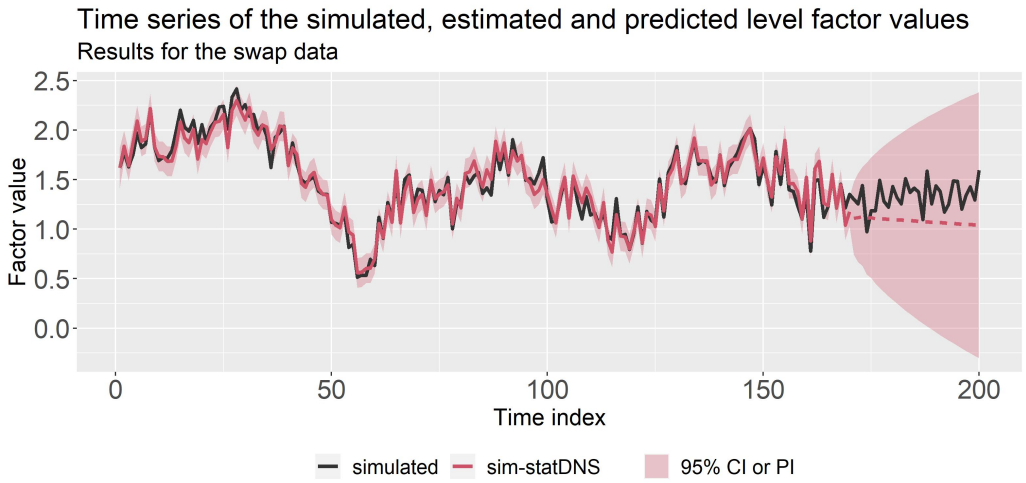


Figure 29: Plot of the series of simulated, estimated and predicted level factors for one of the simulations from the sim-nonstatDNS model.

Interest rate curves In Figure 30 are the plots of the estimated (left) and predicted (middle and right) interest rate curves. The shape of the estimated curves closely follows the structure of the observations. The shape of the predicted curves deviate from the observations due to prediction errors, but the shape of the curve still corresponds in the shape of the observations. This implies that the sim-nonstatDNS model can estimate and predict curves with a shape corresponding to the observations.

Distribution, dependence and variation of the residuals In Figure 31 are the QQ-plots for the residuals of the curves in Figure 30. These plots indicate that the residuals are normally distributed. This is confirmed by the results if the Shapiro-Wilk test in Figure 31. More than 95% of the estimated interest rate curves have normal distributed residuals. Although the percentage of curves with normal residuals is lower for the predicted interest rates, the majority of the predicted rates have normally distributed residuals.

The variation of the residuals is examined with the Beusch-Pagan test. The results for this test are presented in Table 26. The results for the estimated curves imply that more than 95% of the estimated curves have homoskedastic residuals. Only a bit more than 60% of the predicted curves have heteroskedastic residuals.

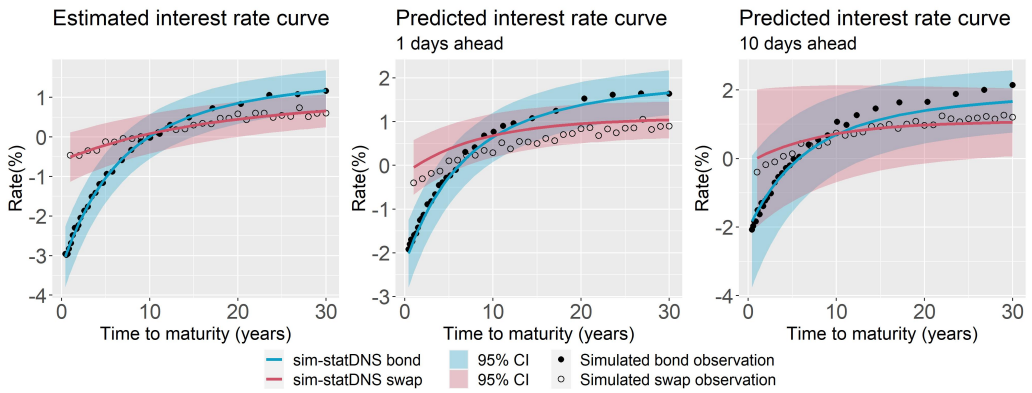


Figure 30: Plot of the estimated and predicted interest rate curves for one of the simulations from the sim-nonstatDNS model.

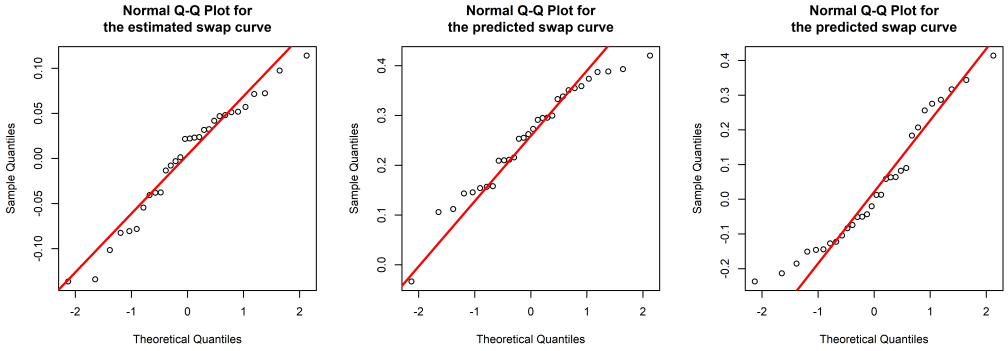


Figure 31: QQ plot of the residuals of the estimated and predicted swap curves in Figure 30.

	Instrument	Total number of curves	# curves that have Gaussian distributed residuals	% of curves that have Gaussian distributed residuals
Estimated	swap	3060	2918	95.4
Predicted	swap	540	343	63.5
Estimated	bond	3060	2906	95.0
Predicted	bond	540	380	70.4

Table 25: The number of interest rate curves estimated and predicted by the sim-nonstatDNS model that have Gaussian distributed residuals according to the Shapiro-Wilk test with a p-value below 0.05.

	Instrument	Total number of curves	# curves with homoskedastic residuals	% curves with homoskedastic residuals
Estimated	swap	3060	2955	96.6
Predicted	swap	540	341	63.1
Estimated	bond	3060	2951	96.4
Predicted	bond	540	350	64.8

Table 26: The number of interest rate curves estimated and predicted by the sim-nonstatDNS model that homoskedastic residuals according to the Breusch-Pagan test with a critical value of 5.991.

5.4.5 Intermediate conclusion for the model validation

In this group of simulation studies, the new variants on the DNS model were validated, based on their ability to estimate λ, their estimated factors, the shape, accuracy, and uncertainty of their estimated and predicted interest rates, and the assumptions on the residuals. For each of the sim-statDNS, the nonstatDNS, and the sim-nonstatDNS the following can be concluded.

- Each of the DNS variants is able to produce an estimate for $\hat{\lambda}_{MLE}$ that is not extremely large or extremely small. That means that the interest rate curves reconstructed from the factors are not violating assumptions on the error distribution.
- The estimated factors are close to the simulated factors, which implies that the models are capable of estimating the latent factors.

- The predicted factors are close to the simulated factors, which implies that the models are capable of predicting the latent factors.
- The estimated and predicted interest rate curves have the same shape as the observations and are close to the observations. That means that the DNS model variants can produce accurate interest rate curves.
- The residuals of the majority of the models are homoskedastic and normally distributed.

Based on these points, it can be concluded that the DNS model variants are valid models to estimate and predict the interest rate curves. All conclusions are drawn under the assumption that the observations have a structure that can be modeled by the specific model. That means these conclusions are only valid under that assumption.

Additional to the validation of the new DNS variants, this simulation study also investigated what could happen if the statDNS model, which assumes stationary factors, was estimated when the true factors are assumed to be nonstationary. In that case, the estimated autoregressive model can either be stationary ($|\phi_{ii}| < 1$) or nonstationary ($|\phi_{ii}|$). Both the stationary and the nonstationary models are theoretically incorrect due to the violation of the stationarity assumption. It can be concluded that the estimated time series model is stationary in 60% of the cases. Although they are incorrect, stationary models can produce very accurate predictions. Approximately 40% of the time series models were nonstationary. The predictions of these models were very inaccurate and not useful.
The proportion of models that were stationary (60%) or nonstationary (40%) depends on the amount of nonstationarity in the factor dynamics and thus also depends on the parameters of the model from which the data is simulated. In this thesis, the parameters were chosen such that the simulated data was a good representation of the real data. If other parameters are chosen to simulate the data, the proportion of stationary versus nonstationary models could change.

5.5 Model performance

The final set of simulations aims to assess the model performance of the DNS models in comparison with the benchmark model. This chapter is split into two parts: in the first part in Section 5.5.1 it is assumed that the data is simulated from a DNS model of which the factors follow a stationary VAR(1) model, and in the second part in Section 5.5.2 it is assumed that the data is simulated from a DNS model of which the first-order differences of the factors follow a stationary VAR(1) model.

The model performance is assessed by the ability of the models to estimate λ in the range of 0.01-10, the accuracy and uncertainty of the estimated swap, bond, and swap spread curves, and the accuracy and uncertainty of the predicted swap, bond, and swap spread curves.

5.5.1 Model performance in the case of stationary factor dynamics

Given that the data is simulated from a DNS model where the factors follow a stationary VAR(1) process, the statDNS and sim-statDNS models are estimated to obtain the estimated and predicted swap, bond, and swap spread curves. The results are presented in this section.

Ability to estimate λ The first result is the estimated values of $\hat{\lambda}_{MLE}$. The statDNS model estimated 4 out of 12 times a value for λ inside the range 0.01-10 for the swap data, and 6 out of 12 for the bond data. The remaining models estimated λ outside this range. For the sim-statDNS model, 12 out of 12 models estimated λ inside the range of 0.01-10. This implies that the sim-statDNS model is more robust in terms of estimating λ.

Accuracy of the estimated curves The second result is the accuracy of the estimated interest rate curves. In Table 27 are the descriptive statistics of the rMSE's of all estimated curves of all 12 simulations. In these results, it can be seen that the curves estimated by the sim-statDNS model have in general a lower rMSE relative to the curves estimated by the statDNS model. This holds for the rMSE of the bond, swap, and spread curves.

Instrument	Model	Q1 rMSE ($\times 10^{-2}$)	Med rMSE ($\times 10^{-2}$)	Q3 rMSE ($\times 10^{-2}$)	Min rMSE ($\times 10^{-2}$)	Max rMSE ($\times 10^{-2}$)
Swap	statDNS	6.45	7.41	14.2	3.84	82.8
Swap	sim-statDNS	6.23	6.90	7.63	3.86	13.8
Bond	statDNS	11.2	12.2	14.1	8.69	22.9
Bond	sim-statDNS	6.2	6.84	7.58	4.21	14.6
Spread	statDNS	54.9	73.2	87.4	10.1	207
Spread	sim-statDNS	9.80	10.5	11.3	7.49	21.9

Table 27: The descriptive statistics of the rMSE's for all estimated interest rate curves and all estimated spread curves aggregated over all simulations. Data is simulated under the assumption that the factors follow a stationary VAR(1) process.

Accuracy of the predicted curves The third result focuses on the comparison of the predicted curves of the two stationary models based on their accuracy. Before measuring the prediction performance by the rMSE, the predicted curves are compared visually. In Figure 32 are the predictions for the swap and bond curves in the top two plots, and the predictions for the spread curves in the bottom two plots. The prediction intervals are left out of the plot to improve visibility. For the prediction intervals, see the plots in the previous paragraphs.

In Figure 32 it can be seen that the swap curve (red) estimated by the statDNS model (top-left plot) is very close to the data. However, due to the lower accuracy of the bond curve (blue), the resulting swap spread prediction in the bottom left plot is not very accurate.
In contrast, the swap and bond curves predicted by the sim-statDNS model (top-right) both have good accuracy, which results in a more accurate swap spread curve for the sim-statDNS model in the bottom-right plot.

The accuracy is again measured by the rMSE. First, the models are mutually compared by their prediction rMSE. Table 28 contains the number of times that one of the models outperforms the other two models based on the rMSE. The results are separated for the swap and bond curves and the table contains the results for the 1-day ahead prediction and all 30 days ahead predictions combined. In these results, it can be seen that the sim-statDNS most often outperforms the statDNS and the random walk model. The sim-statDNS model has the best accuracy for more than 70% of the predictions.

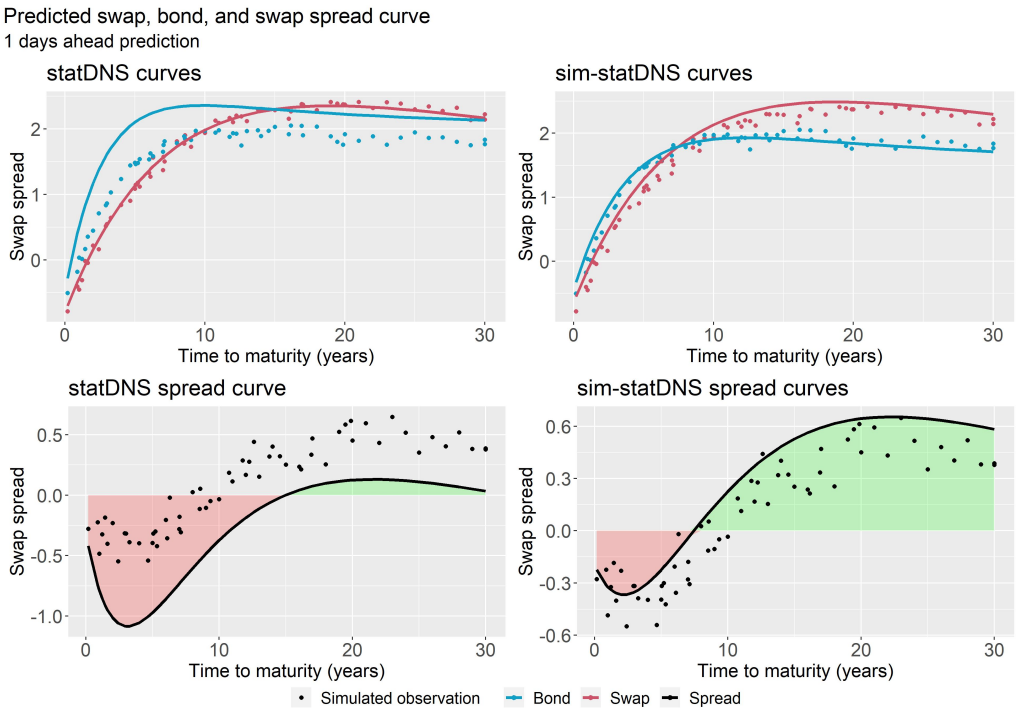


Figure 32: Example of the 1 day ahead predicted swap, bond, and spread curves for the statDNS model and the sim-statDNS model. Data is simulated under the assumption that the factors follow a VAR(1) process.

Furthermore, the statDNS model not often outperforms the sim-statDNS and the random walk model. Although there are some swap curves from the statDNS model that outperform the other models, none of the spread curves predicted by the statDNS model gave the best accuracy among all three models.

		# 1 day ahead	% 1 day ahead	# all predictions	% all predictions
Swap	Total	12	100	360	100
Swap	statDNS	2	16.7	21	5.8
Swap	sim-statDNS	10	83.3	252	70
Swap	Random walk	0	0	87	24.2
Bond	Total	12	100	360	100
Bond	statDNS	0	0	0	0
Bond	sim-statDNS	7	58.3	258	71.7
Bond	Random walk	5	41.7	102	28.3
Spread	Total	12	100	360	100
Spread	statDNS	0	0	0	0
Spread	sim-statDNS	8	66.7	279	77.5
Spread	Random walk	4	33.3	81	22.5

Table 28: Table with the number and percentage of the times that one certain model outperforms both other models based on the rMSE. Results for data simulated under the assumption that the factors follow a VAR(1) process.

It is not only important to know how often one model outperforms the other models, but also to quantify this outperformance. In Table 29 are the descriptive statistics for the rMSE of the one-day ahead predicted curves for all 12 simulations. In this table, it can be seen that the statDNS model not only has worse accuracy than the other two models, but the actual values of the rMSE are also much higher than the rMSE of the other two models. For example, the median rMSE of the statDNS model on the swap data is more than 17 times larger than the median rMSE of the sim-statDNS model on the same data.
Also, the median of the sim-statDNS model is more than 3 times smaller than the random walk model on the swap data. That implies that the sim-statDNS model has not just a smaller rMSE, but a much smaller rMSE. The one-day ahead rMSE for the spread curves implies that the sim-statDNS model reduces the rMSE by almost 50%, relative to the random walk model. This result also holds for the one-day ahead predictions of the bond and swap curves.

Instrument	Model	Q1 rMSE	Med rMSE	Q3 rMSE	Min rMSE	Max rMSE
Swap	statDNS	0.143	2.78	7.37	5.99×10^{-2}	1.23×10^{3}
Swap	sim-statDNS	0.137	0.163	0.282	7.12×10^{-2}	0.691
Swap	Random walk	0.303	0.540	0.736	0.179	0.737
Bond	statDNS	0.557	0.614	0.942	0.525	1.78
Bond	sim-statDNS	0.232	0.253	0.341	0.192	0.437
Bond	Random Walk	0.199	0.255	0.341	0.192	0.437
Spread	statDNS	1.64	3.85	27.4	0.475	4.53×10^{3}
Spread	sim-statDNS	0.177	0.271	0.474	9.46×10^{-2}	0.985
Spread	Random walk	0.288	0.561	0.767	0.137	1.46

Table 29: The rMSE results for the 1 day ahead predictions of the statDNS, sim-statDNS and random walk model aggregated over all simulations with the assumption that the factors follow a stationary VAR(1) process.

Uncertainty The final step in the comparison of the models is the uncertainty of their estimations. In Table 30 are the results for the variances of the estimated interest rates. The statDNS model has a much larger uncertainty than the sim-statDNS model. For the swap curves, the statDNS model variance is much larger than the variance estimated by the sim-statDNS model. For the bond data, the lowest 50% of the variances has comparable size for both models, but the upper 50% of the statDNS model is much larger than the upper 50% of the sim-statDNS model.

Instrument	Model	Q1 $(\hat{\sigma^2}\times10^{-2})$	Med $(\hat{\sigma^2}\times10^{-2})$	Q3 $(\hat{\sigma^2}\times10^{-2})$	Min $(\hat{\sigma^2}\times10^{-2})$	Max $(\hat{\sigma^2}\times10^{-2})$
			Estimated			
swap	statDNS	7.93	2.58×10^{4}	3.29×10^{5}	6.45	1.59×10^{8}
swap	sim-statDNS	5.66	6.15	6.96	5.01	61.7
bond	statDNS	1.83	6.69	7.56×10^{3}	1.23	4.81×10^{4}
bond	sim-statDNS	5.15	6.49	8.74	4.08	30.2
			Predicted			
swap	statDNS	9.33	9.89×10^{17}	5.10×10^{48}	6.45	7.57×10^{125}
swap	sim-statDNS	6.91	7.75	8.89	5.01	19.5
bond	statDNS	7.32	8.19	36.3	1.23	2.89×10^{34}
bond	sim-statDNS	7.67	9.36	11.5	4.09	20.0

Table 30: The results for the estimate of the variance for the estimated and predicted interest rate curves per instrument and per model.

5.5.2 Model performance in the case of nonstationary factor dynamics

Given that the data is simulated from a DNS model where the first-order differences of the factors follow a stationary VAR(1) process, the nonstatDNS, and sim-nonstatDNS models are estimated to obtain the estimated and predicted swap, bond, and swap spread curves. The results will be presented in this section.

Ability to estimate λ For the nonstatDNS model on the bond data, 14 out of 18 models have an estimate for λ inside the range of 0.01-10. For the nonstatDNS model on the swap data, all 18 out of 18 models have λ in that range. For the sim-nonstatDNS model, also all 18 out of 18 models have λ in the range 0.01-10 for both the bond and swap λ's.

Accuracy of the estimated curves In Table 31 are the rMSE results for the estimated interest rate curves and the estimated swap spread curves for all simulations. In this table, it can be seen that the rMSE on the swap data is almost equal for both the nonstatDNS and the sim-nonstatDNS models. For the bond curves, the sim-nonstatDNS model has a lower rMSE in general.
Our main interest is the rMSE of the swap spread in the bottom two rows of Table 31. The nonstatDNS has a maximum rMSE for the swap spread of 2.34%, which is significantly larger than the maximum rMSE of 0.145 obtained from the sim-nonstatDNS model. Not only the maximum, but all descriptive statistics for the rMSE of the spread are lower for the sim-nonstatDNS model relative to the nonstatDNS model. This means that the sim-nonstatDNS model estimates the spread more accurately than the nonstatDNS model.

Accuracy of the predicted curves The next part of the analysis is the accuracy of the predicted curves. The first step is to visualize the predicted swap curve, bond curve and swap spread curve. In Figure 33 are the plots of one example of a 1-day ahead prediction for all three curves and both models. In the top row

Instrument	Model	Q1 rMSE ($\times 10^{-2}$)	Med rMSE ($\times 10^{-2}$)	Q3 rMSE ($\times 10^{-2}$)	Min rMSE ($\times 10^{-2}$)	Max rMSE ($\times 10^{-2}$)
Swap	nonstatDNS	6.11	6.68	7.30	3.87	9.86
Swap	sim-nonstatDNS	6.11	6.69	7.31	3.86	9.89
Bond	nonstatDNS	8.70	12.2	16.6	3.97	62.8
Bond	sim-nonstatDNS	6.04	6.63	7.25	3.83	9.51
Spread	nonstatDNS	12.5	17.3	25.6	8.55	234
Spread	sim-nonstatDNS	9.69	10.4	11.1	7.04	14.5

Table 31: The descriptive statistics of the rMSE's for all estimated interest rate curves and all estimated spread curves aggregated over all simulations. Data is simulated under the assumption that the first-order differences of the factors follow a stationary VAR(1) process.

plots are the swap curve and bond curve and in the bottom row plots are the swap spread curves for both the nonstatDNS (left) and sim-nonstatDNS (right) models.
The plot visualizes the relation between the accuracy of the swap and bond curve and the accuracy of the spread curve. For example, the predicted bond curves (blue) have a very accurate fit for small times to maturity. But due to the worse fit of the predicted swap curves (red), the predicted spread curve (black) is not as accurate as the predicted bond curve. This implies that, in order to have an accurate spread prediction, both the swap and bond curves must be accurately predicted.

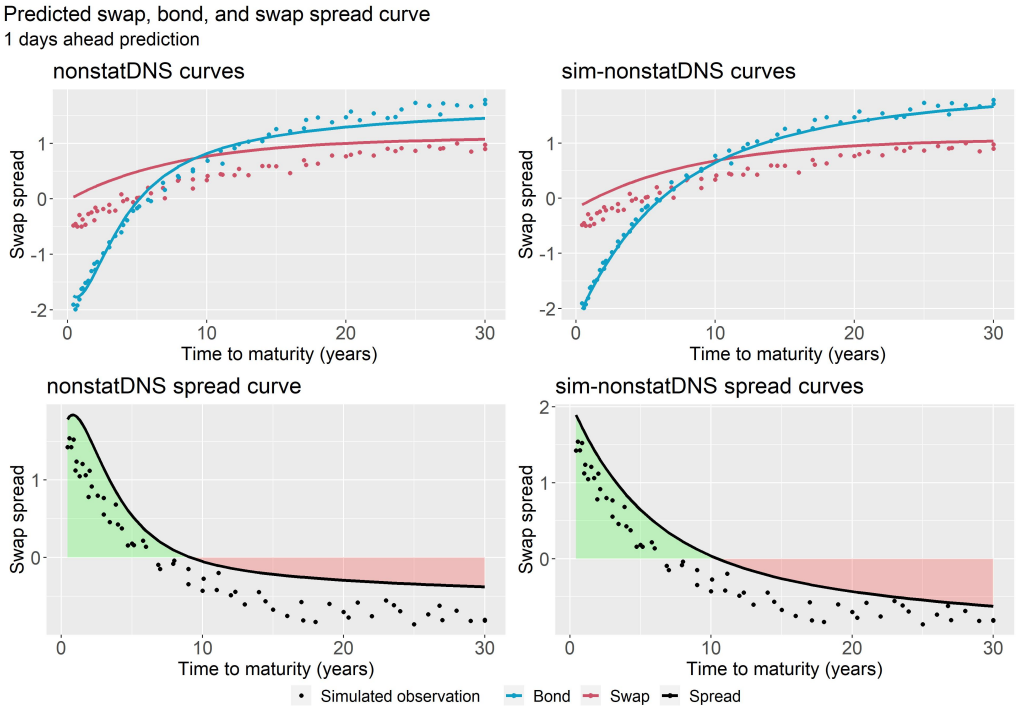


Figure 33: Example of the 1 day ahead predicted swap, bond, and spread curves for the nonstatDNS model and the sim-nonstatDNS model. Data is simulated under the assumption that the first-order differences of the factors follow a VAR(1) model.

The next step is to compare the predicted curves by their rMSE. In Table 32 are the predictions of the nonstatDNS, the sim-nonstatDNS, and the random walk models compared by their rMSE. The table contains the number of times one model outperforms the other two models. In this table, it can be seen that the sim-statDNS model outperforms the other models most often for the swap, bond, and spread curves. The random walk model has approximately one-third of the times the most accurate prediction and the nonstatDNS model has approximately one-third of the times the most accurate prediction.

To quantify the accuracy, in Table 33 are the descriptive statistics for the rMSE of all one-day ahead predicted interest rate curves. The rMSE of the sim-nonstatDNS model is generally the lowest among all three models. The rMSE of the nonstatDNS model does not reduce the rMSE relative to the random walk model for the spread curve predictions.

		# 1 day ahead	% 1 day ahead	# all predictions	% all predictions
Swap	Total	18	100	540	100
Swap	nonstatDNS	3	16.7	141	26.1
Swap	sim-nonstatDNS	11	61.1	207	38.3
Swap	Random walk	4	22.2	192	35.6
Bond	Total	18	100	540	100
Bond	nonstatDNS	2	11.1	131	24.3
Bond	sim-nonstatDNS	13	72.2	202	37.4
Bond	Random walk	3	16.7	207	38.3
Spread	Total	18	100	540	100
Spread	nonstatDNS	0	0	53	9.8
Spread	sim-nonstatDNS	13	72.2	278	51.5
Spread	Random walk	5	27.8	209	38.7

Table 32: Table with the number and percentage of the times that one certain model outperforms both other models based on the rMSE. Results for data simulated under the assumption that the first-order differences of the factors follow a VAR(1) process.

Instrument	Model	Q1 rMSE ($\times 10^{-2}$)	Med rMSE ($\times 10^{-2}$)	Q3 rMSE ($\times 10^{-2}$)	Min rMSE ($\times 10^{-2}$)	Max rMSE ($\times 10^{-2}$)
Swap	nonstatDNS	10.3	24	37.2	4.93	47.4
Swap	sim-nonstatDNS	10.4	13.8	18.4	7.61	41.2
Swap	Random walk	18.7	30.6	36.6	10.4	48.3
Bond	nonstatDNS	20.6	25	44.1	13.7	63.8
Bond	sim-nonstatDNS	10.8	18.6	30.6	8.9	51.6
Bond	Random walk	26.8	35.5	46.7	14.9	92.2
Spread	nonstatDNS	23.8	27.8	49.3	17.7	240
Spread	sim-nonstatDNS	16.2	22.3	34.8	11.4	56.6
Spread	Random walk	18.0	23.0	41.9	14.6	77.9

Table 33: The rMSE results for the 1 day ahead predictions aggregated over all simulations with the assumption that the first-order differences of the factors follow a stationary VAR(1) process.

Uncertainty The last result is the uncertainty of the estimated interest rate curves, measured by the estimate for the variance of the interest rate estimation. In Table 34 are the results for the estimated variances by the DNS models. For the swap data, the nonstatDNS has a larger variance and thus a larger uncertainty. For the bond data, the sim-nonstatDNS has a larger variance. This means, there is no significant difference in the uncertainty of the estimations of the nonstatDNS and the sim-nonstatDNS models.

Instrument	Model	Q1 ($\hat{\sigma^2} \times 10^{-2}$)	Med ($\hat{\sigma^2} \times 10^{-2}$)	Q3 ($\hat{\sigma^2} \times 10^{-2}$)	Min ($\hat{\sigma^2} \times 10^{-2}$)	Max ($\hat{\sigma^2} \times 10^{-2}$)
			Estimated			
swap	nonstatDNS	6.90	8.07	10.6	5.35	1.80×10^{3}
swap	sim-nonstatDNS	5.39	6.19	7.66	4.3	3.23×10^{3}
bond	nonstatDNS	6.91	7.97	9.62	5.75	1.01×10^{3}
bond	sim-nonstatDNS	7.58	9.94	12.3	5.51	1.30×10^{3}
			Predicted			
swap	nonstatDNS	62.3	127	231	5.11	1.29×10^{5}
swap	sim-nonstatDNS	34.0	58.8	90.2	4.30	375
bond	nonstatDNS	50.1	99.0	243	5.07	3.36×10^{44}
bond	sim-nonstatDNS	30.5	53.5	90.2	5.11	257

Table 34: The results for the estimate of the variance for the estimated and predicted interest rate curves per instrument and per model for the nonstationary data.

5.5.3 Intermediate conclusion

If the factors are assumed to follow a stationary process, the sim-statDNS model outperforms both the statDNS and the random walk model by means of rMSE of the estimated and predicted swap and bond curves in more than 50% of the cases. The sim-statDNS model outperforms both other models by means of rMSE of the estimated and predicted swap spead curves in more than 60% of the cases. The statDNS has the lowest accuracy in terms of rMSE of all models and could not compete with the random walk model.
The uncertainty of the predicted interest rates can not be investigated. To compare the uncertainty, I look at the estimated variances for the estimated interest rate curves. In these results, the sim-statDNS model has a lower uncertainty than the statDNS model.

If the first-order differences of the factors are assumed to follow an AR(1) process, the sim-nonstatDNS model slightly outperforms the random walk model on the accuracy measured by the rMSE for the swap and bond curves. The sim-nonstatDNS outperforms both other models by means of rMSE of the predicted spread curve in more than 50% of the cases. The uncertainty of the sim-nonstatDNS model is comparable to the uncertainty of the nonstatDNS model.

5.6 Methodology derived from the simulation studies

The simulation studies are combined into the following methodology that specifies how to predict the swap spread using the DNS model.

1. **Preprocess the data set of observations**
   Follow the preprocessing method for the bond and swap data as described in Section 2.1 and Section 2.2.

2. **The analysis of the empirical factors**
   Compute the time series empirical factors of both rates from the data set of observations following the method specified in Section 2.1.4. Perform the ADF test to test the stationarity of the factors. If the series are not stationary, perform the ADF test on the series after a linear first-order difference transformation.

3. **Choose the correct DNS model variant**
   If the series of empirical factors are all stationary processes, choose the statDNS and the sim-statDNS. If all series are nonstationary and all series after a linear first-order difference transformation are stationary, choose the nonstatDNS and the sim-nonstatDNS models. This framework does not include the case where some of the factor series are stationary and some are not.

4. **Choose the initial value for λ**
   Choose an initial value of λ. This can be any value between 0.01 and 10.

5. **Estimate the DNS models**
   Estimate the DNS models with the initial value for λ following the method in Section 3.4.3. Keep tracking the intermediate estimates for λ. If the estimates for $\hat{\lambda}_M LE$ are too small (< 0.01) or too large (> 10), the DNS model can not be used to estimate and predict the swap spread.

6. **Plot the factor dynamics and the interest rate curves**
   The structure and dynamics of the factors should correspond to the structure and dynamics of the empirical factors. If not, the model probably does not produce accurate results. This is assessed by looking at the plot of the empirical factors and the estimated factors.

7. **Expectation of the results**
   In the case of stationary factor dynamics, it is expected that the sim-statDNS outperforms the random walk model and the statDNS model on both accuracy and uncertainty. It is also expected that the statDNS model does not outperform the random walk model.
   In the case of nonstationary factor dynamics, it is expected that the sim-nonstatDNS outperforms the random walk model and the nonstatDNS model on both accuracy and uncertainty. It is also expected that the nonstatDNS model outperforms the random walk model in accuracy.

8. **Result analysis**
   The estimated rates are analyzed by their accuracy (rMSE), their uncertainty ($\hat{\sigma}^2$), and the distribution (Shapiro-Wilk test) and variability (Breusch-Pagan test) of the residuals. The predicted interest rates are analyzed by the same measures except for the uncertainty because the models can not produce predictions for this uncertainty.

6 Case study

In this section, the framework to predict the swap spread will be applied to a real data set. The source, structure, characteristics, and preprocessing procedure of the data set are already described in Section 2 and will not be repeated in this section. The structure of this section corresponds with the steps in the framework.

6.1 Analysis of the empirical factors

The empirical factors for the real data are computed according to the method described in Section 2.1.4. In Figure 34 is a plot of the empirical counterparts of the level factor. The plot for the slope and curvature factors are attached in Figure 54. The time series in these plots do not look stationary.
Following the framework, all three time series are transformed by a linear first-order difference transformation. The plots of these transformed time series are in Figure 35 for the level factors and in Figure 55 for the slope and curvature factors. The series in these plots look stationary.

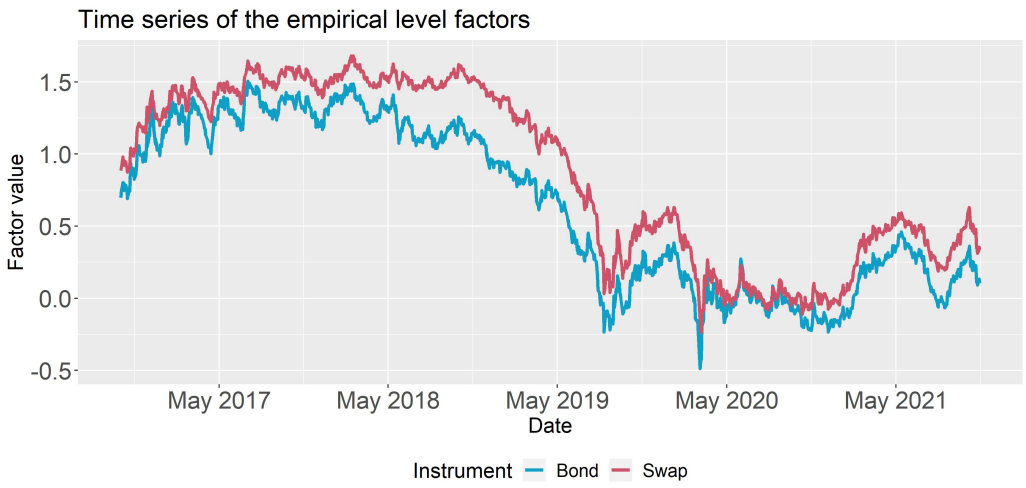
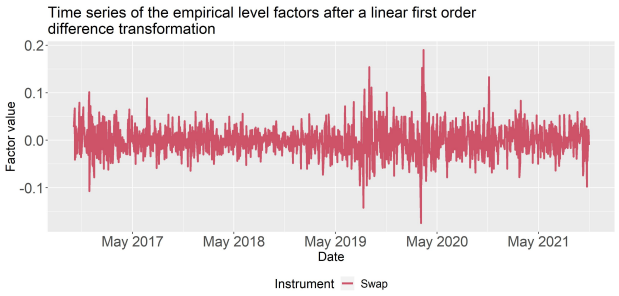
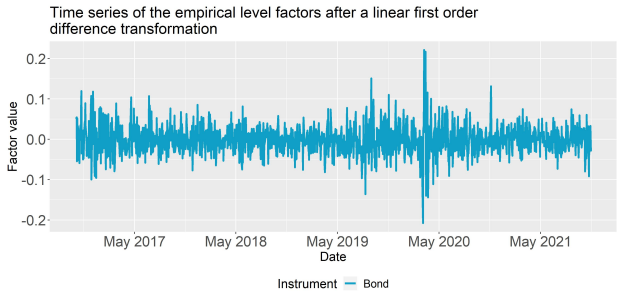


Figure 34: Time series of the empirical level factors computed from the real data.



(a) The transformed empirical level factor for the swap data



(b) The transformed empirical level factor for the bond data

Figure 35: Time series after a first order difference transformation of the empirical level factors computed from the real data.

The stationarity of the empirical time series is statistically evaluated with the ADF test. The results of these tests are presented in Table 35. A series is said to be nonstationary if the ADF test statistic is below the critical value of -3.44 and the p-value is below 0.05. That implies that the series without a transformation are all nonstationary and the series after a first-order difference transformation are stationary. Based on these results, the factors of the DNS model are assumed to be stationary after a linear first-order difference transformation. According to the framework, the interest rate curves will be predicted using the nonstatDNS model and the sim-nonstatDNS model.

		No transformation		Transformed series	
Instrument	Factor	ADF test statistic	ADF p-value	ADF test statistic	ADF p-value
Swap	Level	-2.55	0.344	-10.6	< 0.01
Swap	Slope	-2.42	0.399	-10.9	< 0.01
Swap	Curvature	2.22	0.486	-10.8	< 0.01
Bond	Level	-2.95	0.175	-11.0	< 0.01
Bond	Slope	-2.90	0.198	-11.0	< 0.01
Bond	Curvature	-2.96	0.173	-13.9	< 0.01

Table 35: Results of the ADF test for the empirical factor series before and after a linear first order difference transformation.

6.2 Results of the case study

Both the nonstatDNS and the sim-nonstatDNS models are estimated on the train data set. In this section, the results of these models are presented.

6.2.1 The estimated parameters and factors

The estimates for all parameters are included in Section H.1. In this section, I will only point out the most remarkable results.
The first result is the estimated value for λ in Table 36. The estimated λ for the nonstatDNS model on the bond data is smaller than 0.01. According to the framework, this model leads to inaccurate estimated and predicted bond curves and should not be used. Because it is interesting to see if the results of the simulation studies also hold for the real data, also the model with a λ smaller than 0.01 will be presented. The other three estimated λ's are in the range of 0.01-10 and thus can be used to predict the interest rate curves.

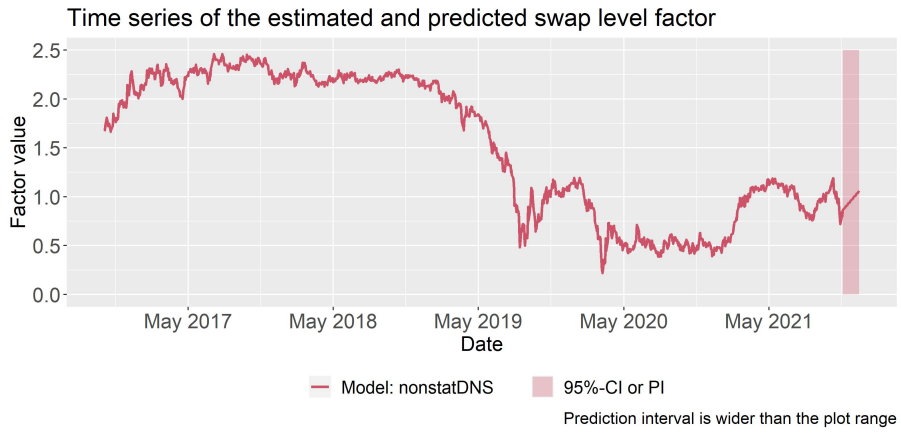
Instrument	Model	Initial λ	$\hat{\lambda}_{MLE}$
Swap	nonstatDNS	0.1195	0.346
Swap	sim-nonstatDNS	0.1195	0.441
Bond	nonstatDNS	0.1195	0.00987
Bond	sim-nonstatDNS	0.1195	0.347

Table 36: The estimates for $\hat{\lambda}_{MLE}$ for the real data

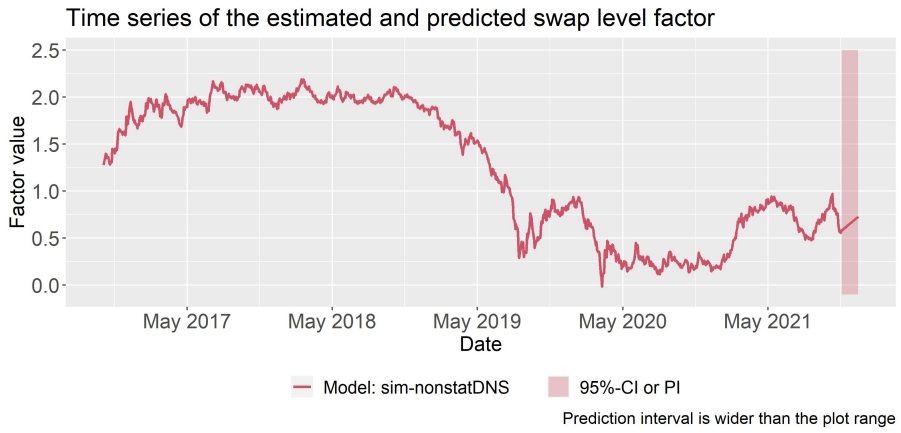
The subsequent results are the factors estimated by each of the models. In Figure 36 are the plots of the estimated level factors. In Appendix H are the plots of the estimated slope and curvature factors. The first step in the analysis is to look at the values of the factors. The factor values for the swap level factors of both models take values approximately between 0 and 2.5. The factor values for the bond level factors estimated by the sim-nonstatDNS model take values between -0.5 and 2. These values are comparable to the simulation study and are not extreme. The factor values for the bond level factor estimated by the nonstatDNS model take values between -200 and 300. These values are more extreme, which is due to the quite small estimate for λ.
Another remark on the results in Figure 36 is that the prediction intervals are much larger than the prediction intervals seen in the simulation studies.

6.2.2 The estimated bond and swap curves

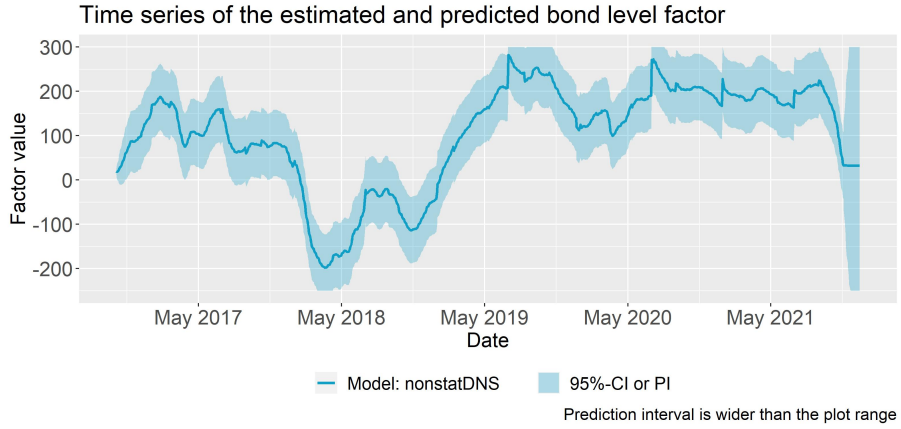
After the evaluation of the estimated and predicted factor values, the estimated curves are analyzed to see if the model is able to reproduce the shape of the observations. In Figure 37 are the plots of the interest rate curves estimated by the different models. The two plots in the top row are the swap curves estimated by the nonstatDNS model (left) and the sim-nonstatDNS model (right). The two plots in the bottom row are the bond curves estimated by the nonstatDNS model (left) and the sim-nonstatDNS model (right). The shaded area in the plot represents the 95%-confidence interval. The range of the plot is smaller than the confidence interval in order to see the fit of the estimated bond and swap curves. To see the whole confidence interval, see Figure 61 in the appendix.
The bond curve that is estimated by the nonstatDNS model (bottom left) is linear. This is exactly what happens with the curves that are modeled by a λ that is very small, as shown in the simulation in Section 5.2.2. All other curves are close to the observations, which is also in line with the results in the simulation studies for models with a λ in the range of 0.01-10.



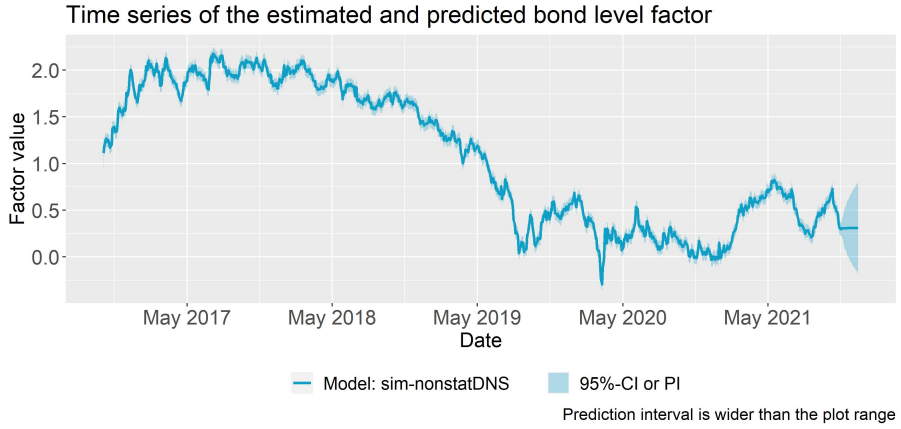
(a) Caption



(b) Caption



(c) Caption



(d) Caption

Figure 36: The level factors estimated and predicted by the nonstatDNS (left) and sim-nonstatDNS (right) models for the real data.

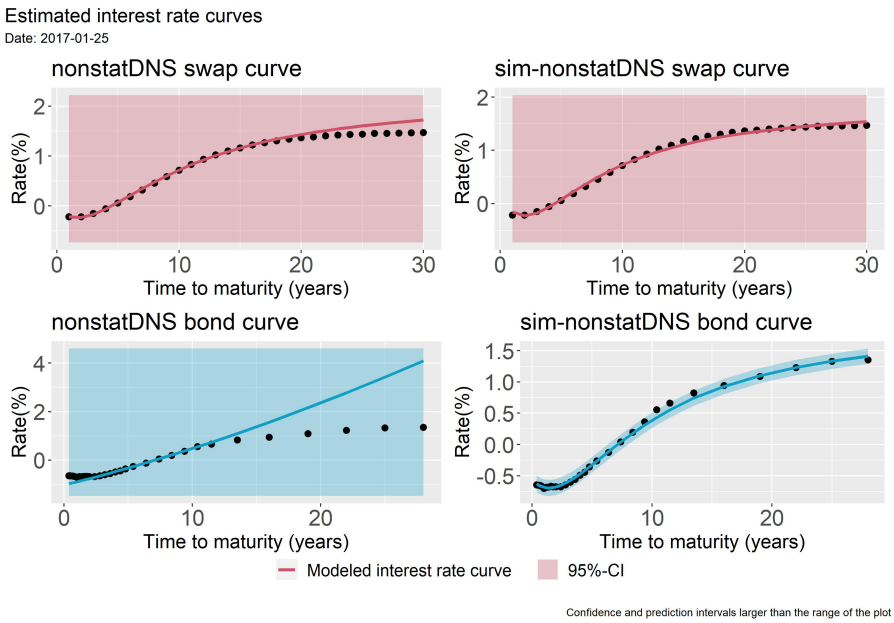
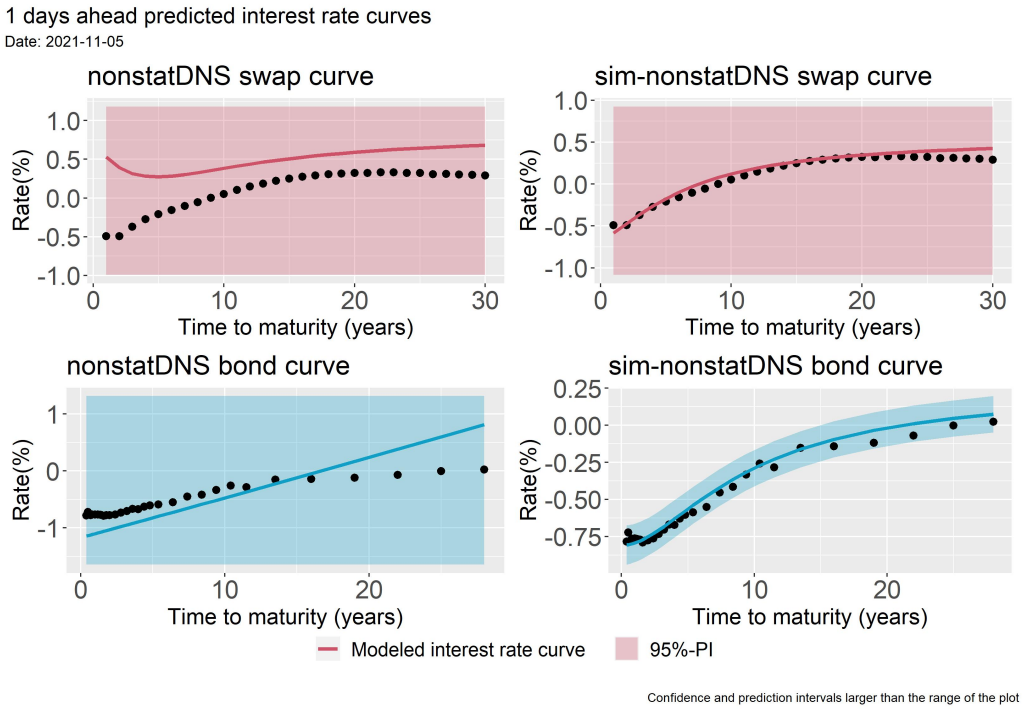


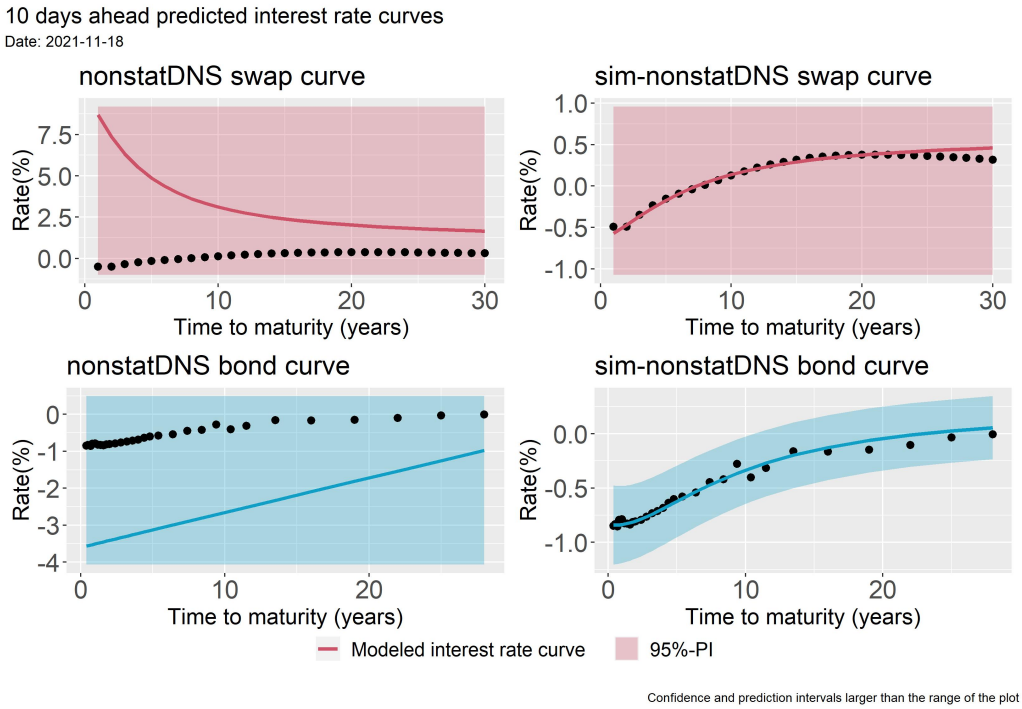
Figure 37: Example of the estimated interest rate curves for the nonstatDNS and the sim-nonstatDNS models.

Subsequently, the predicted curves will be analyzed. In Figure 38a and Figure 38b are the 1 day and 10 days ahead predicted bond and swap curves. Again, the 95%-prediction intervals are larger than the range of the plot. To see the full prediction intervals, see Figure 61 in the appendix.
The swap curve predicted by the nonstatDNS model is plotted in the top left of both figures. In these plots, it can be seen that the predicted swap curve is not close to the observations. The shape of the predicted curve is also not the same as the shape implied by the observations. The shape of the curve can be explained by the prediction of the slope factors, which are increasing relative to their value one day earlier, see Appendix H. This effect holds for the 1-day ahead prediction and is even more present for the 10-days ahead prediction. Thus, in this case, a bad prediction for the slope factor leads to a swap curve that does not fit the structure of the data.
The bond curve predicted by the nonstatDNS model is plotted in the bottom left of both Figure 38a and Figure 38b. The shape of these curves is again a linear line due to the very small estimate for $\hat{\lambda}_{MLE}$.

The swap curve that is predicted by the sim-nonstatDNS model is plotted in the top right of the plots. Although there is a deviation from the data for high times to maturity, both predicted swap curves are in general close to the data. Also, the 95% prediction intervals are very large.
The bond curves predicted by the sim-nonstatDNS model are plotted in the bottom right of the plots. These curves are close to the data at every time to maturity. The prediction interval is smaller than the prediction interval for the predicted swap curves.



(a) The one day ahead predicted interest rate curves for the nonstatDNS (left) and the sim-nonstatDNS (right) models.



(b) The 10-days ahead predicted interest rate curves for the nonstatDNS (left) and the sim-nonstatDNS (right) models.

Figure 38: Examples of bond and swap curves that are predicted using the DNS models on the real data.

6.2.3 Performance analysis

So far, the shape and accuracy of the estimated and predicted curves are visually evaluated. The next step is to compare the models by their rMSE. As explained in Section 3.5, it is not possible to determine the rMSE of the DNS models for the swap spread curve, hence the performance accuracy is measured by the rMSE of the swap curve and the bond curve separately.
The results for the rMSE of the estimated bond and swap curves are presented in Table 37. In this table, it can be seen that the sim-nonstatDNS model produces bond and swap curves that have a lower rMSE, relative to the curves estimated by the nonstatDNS model. The curves with the lowest and highest rMSE's were plotted to detect any pattern in the estimation accuracy. However, no clear pattern could be found in the shape of these curves. See Figure 62 and Figure 63 in the Appendix for these plots.

Instrument	Model	Q1	Med	Q2	Min	Max
Swap	nonstatDNS	0.0897	0.0975	0.105	0.0576	0.150
Swap	sim-nonstatDNS	0.0327	0.0352	0.0386	0.0254	0.0709
Bond	nonstatDNS	0.305	0.376	0.464	0.124	1.34
Bond	sim-nonstatDNS	0.0228	0.0273	0.0331	0.0112	0.548

Table 37: Descriptive statistics of the rMSE for all swap and bond curves estimated by the DNS models on the real data.

The rMSE of the one, ten, and thirty days ahead predicted curves are presented in Table 38. In this table, it can be seen that for the swap curves, the random walk model has the lowest rMSE for each predicted curve. For the bond data, the sim-nonstatDNS model outperforms both the nonstatDNS and the random walk model.

Instrument	Model	rMSE 1 day ahead	rMSE 10 days ahead	rMSE 30 days ahead
Swap	nonstatDNS	0.426	3.58	10.7
Swap	sim-nonstatDNS	0.0639	0.0530	0.169
Swap	Random walk	0.0587	0.0308	0.0605
Bond	nonstatDNS	0.315	2.40	7.51
Bond	sim-nonstatDNS	0.0421	0.0413	0.0964
Bond	Random walk	0.0472	0.0784	0.103

Table 38: Results for the rMSE of the swap and bond curves predicted by the DNS models and the random walk model.

6.2.4 Residual analysis

The final part of the result section is the analysis of the residuals for the estimated bond and swap curves. In both the nonstatDNS and the sim-nonstatDNS models, it is assumed that the errors are i.i.d. normally distributed. The normality of the residuals is tested by the Breusch-Pagan (BP) test. The results for the BP tests are presented in Table 39. The results for the nonstatDNS models imply that the normality assumption is violated for all estimated and for all predicted interest rate curves. The results for the sim-nonstatDNS model imply that most of the residuals of the estimated and predicted curves are normally distributed.

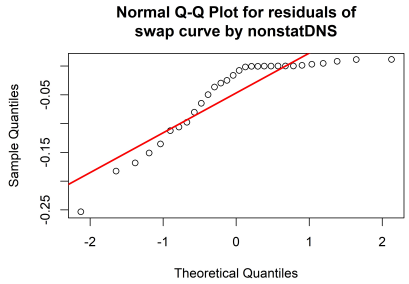
Instrument	Model	total # curves	# normally distributed residuals	% normally distributed residuals
Residuals of the estimated curves				
Swap	nonstatDNS	1300	0	0
Swap	sim-nonstatDNS	1300	883	67.9
Bond	nonstatDNS	1300	0	0
Bond	sim-nonstatDNS	1300	809	62.2
Residuals of the predicted curves				
Swap	nonstatDNS	30	0	0
Swap	sim-nonstatDNS	30	25	83.3
Bond	nonstatDNS	30	0	0
Bond	sim-nonstatDNS	30	18	60

Table 39: The number of estimated and predicted curves with normally distributed residuals according to the Shapiro-Wilk test with a p-value higher than 0.05.

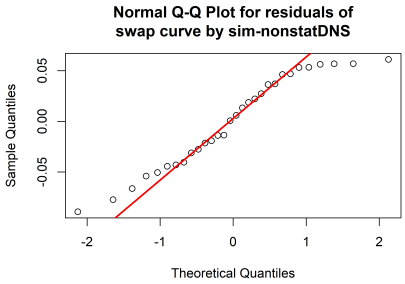
To visualize the distribution of the residuals, in Figure 39 are the Q-Q plots of the residuals of each model for one of the estimated curves. To see where the errors presented in the Q-Q plots in Figure 39 are large or

small, the corresponding estimated swap and bond curves are plotted in Figure 40.

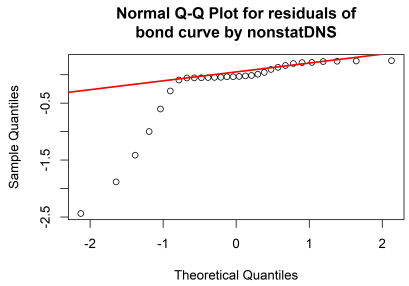
The Q-Q plot in Figure 39a confirms the results of the BP test: the residuals of the swap curve estimated by the nonstatDNS model are not normally distributed. In Figure 40a it can be seen that there are larger errors for very small or very large times to maturity.
The same can be said about the Q-Q plot for the bond curve estimated by the nonstatDNS model in Figure 39c. This Q-Q plot indicates a heavily skewed distribution for the residuals. The distribution of these residuals can be explained by the shape of the bond curve in Figure 40c, which diverges from the data for large times to maturities.
The Q-Q plots in Figure 39b and Figure 39d represent the residuals for the swap and bond curves estimated by the sim-nonstatDNS model. Although there is a deviation of the points from the red line, the results in the Q-Q plots indicate that the residuals are indeed normally distributed. In Figure 40b and Figure 40d it can be seen that the curves are close to the data over the whole range of times to maturities.



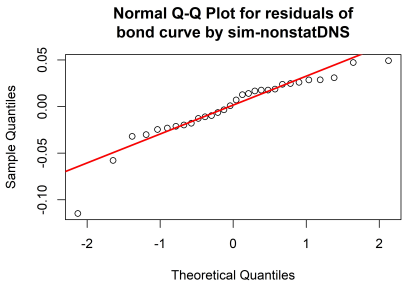
(a) Residuals of a swap curve estimated by the nonstatDNS model



(b) Residuals of a swap curve estimated by the sim-nonstatDNS model



(c) Residuals of a bond curve estimated by the nonstatDNS model

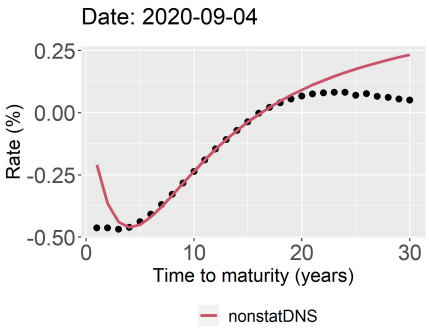


(d) Residuals of a bond curve estimated by the sim-nonstatDNS model

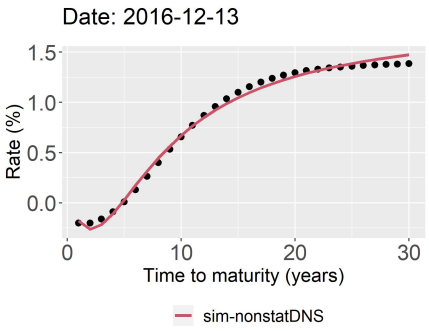
Figure 39: The Q-Q plots of the residuals of one of the estimated bond and swap curves for both the nonstatDNS and the sim-nonstatDNS models.

The results of the Shapiro-Wilk test in Table 39 are directly related to the results of the Q-Q plots and the shapes of the curves. The curves estimated by the nonstatDNS model do not follow the structure of the observations. For the bond data, this is due to very small $\hat{\lambda}_{MLE}$. The small λ results in a linear curve, which can not follow the data structure. In that case, the residuals will be skewed, because the curve deviates from the data for large times to maturity, as in Figure 40c. This directly causes the non-normality of the residuals. This problem was discussed in the simulation studies in Section 5.2.3.
The non-normality of the residuals of the swap curves estimated by the nonstatDNS model is caused by the deviation of the curve from the observations for small or large times to maturity. This can not be explained by the estimate for λ, because the estimate is not very large or very small. The distribution of the residuals is due to the disability of the model to follow the data structure.

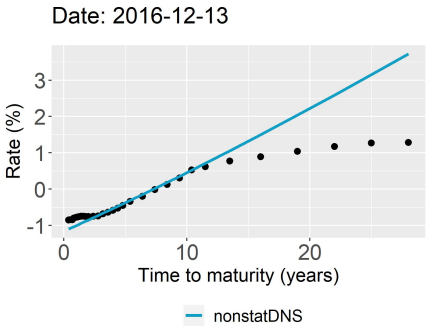
The homoskedasticity of the residuals is tested with the Breusch-Pagan (BP) test. In Table 40 are the results of the BP test for the estimated and predicted swap and bond curves. These results show that there is heteroskedasticity present for the residuals of the sim-nonstatDNS model. That means that there is variability in the swap and bond rate that is not explained by the three factors.
It is remarkable that the results in Table 40 are in contrast with the results for the sim-nonstatDNS model from the simulation study, see Table 26. In the simulations, more than 90% of the estimated curves and more than 60% of the predicted curves from the sim-nonstatDNS model have homoskedastic residuals. This is significantly



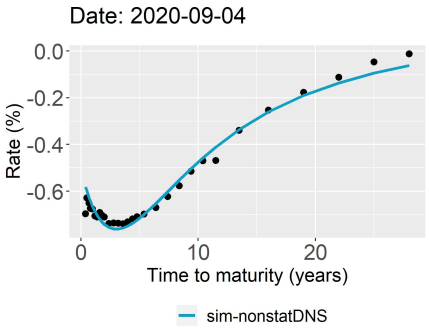
(a) Swap curve corresponding to the Q-Q plot in Figure 39a



(b) Swap curve corresponding to the Q-Q plot in Figure 39b



(c) Bond curve corresponding to the Q-Q plot in Figure 39c



(d) Bond curve corresponding to the Q-Q plot in Figure 39d

Figure 40: The swap and bond curve that correspond to the Q-Q plots in Figure 39

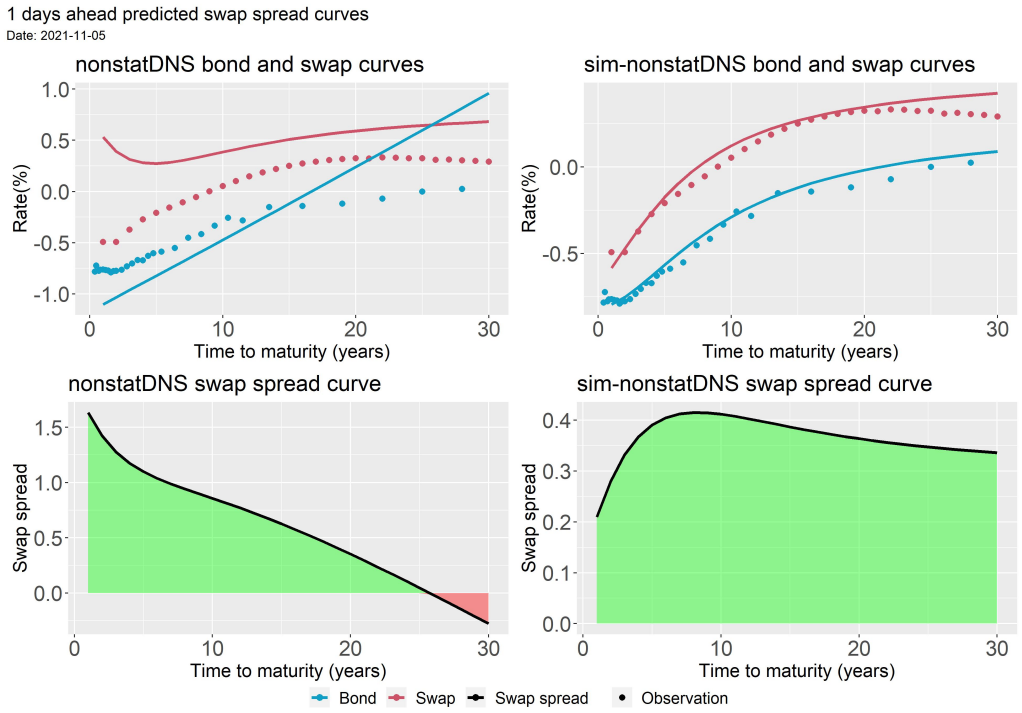
more often than in the case study, in which fewer than 50% of the estimated curves and fewer than 30% of the predicted curves have homoskedastic residuals. This will be discussed in the discussion section.

Instrument	Model	Total # curves	# curves with homoskedastic residuals	% curves with homoskedastic residuals
Residuals of the estimated curves				
Swap	nonstatDNS	1300	1300	100
Swap	sim-nonstatDNS	1300	204	15.7
Bond	nonstatDNS	1300	1300	100
Bond	sim-nonstatDNS	1300	595	45.8
Residuals of the predicted curves				
Swap	nonstatDNS	30	30	100
Swap	sim-nonstatDNS	30	4	13.3
Bond	nonstatDNS	30	1	3.33
Bond	sim-nonstatDNS	30	16	53.3

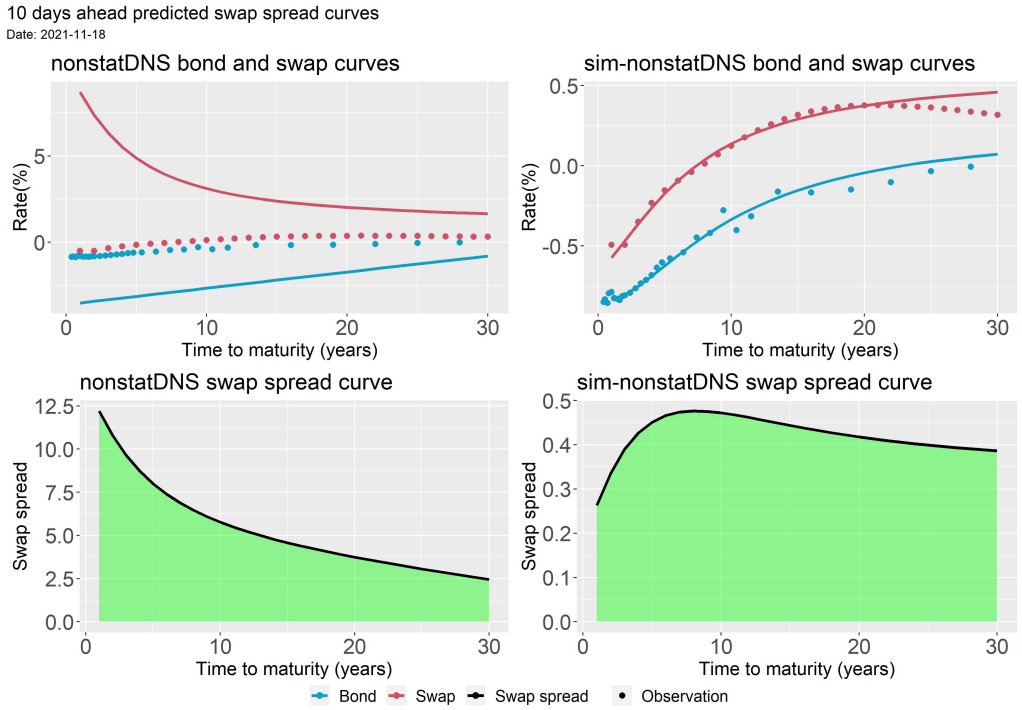
Table 40: Homoskedasticity of the residuals. Results for the Breusch-Pagan test with a critical value of 5.991.

6.2.5 Estimated and predicted swap spread curves

The main objective of this thesis was to predict the swap spread curve, i.e. the difference between the bond curve and the swap curve. In Figure 41 are the plots of the 1-day and the 10-days ahead predicted swap spread curves. The top plots in the Figure 41a and Figure 41b contain the bond and swap curves predicted by the nonstatDNS (left) and the sim-nonstatDNS (right). These are the same predicted curves as shown in Figure 38, but then for the swap and bond curves combined into one plot per model. The plots in the bottom row of Figure 41a and Figure 41b contain the predicted swap spread curves for both the nonstatDNS (left) and the sim-nonstatDNS (right) models.



(a) The one day ahead predicted swap spread curves for the nonstatDNS (left) and the sim-nonstatDNS (right) models.



(b) The 10-days ahead predicted swap spread curves for the nonstatDNS (left) and the sim-nonstatDNS (right) models.

Figure 41: Examples of spread curves that are predicted using the DNS models on the real data. The green and red shade indicate a positive or negative spread respectively.

The shape of the swap spread curve in Figure 38 is different for both the DNS models. The swap spread curve for the nonstatDNS model is a downward-sloping curve, whereas the swap spread curve predicted by the sim-nonstatDNS model has a bumped shape.
From the results in this section, it is known that the nonstatDNS model has not a good performance in the prediction of the swap and bond curves. That means that the swap spread in Figure 41 predicted by the nonstatDNS model is not a good prediction for the swap spread.
On the other hand, the results have shown that the sim-nonstatDNS model has a good performance in predicting the swap spread. This is confirmed by the predicted swap and bond curves in Figure 41. Based on this result, the swap spread curve predicted by the sim-nonstatDNS model is a good prediction for the swap spread.
The performance of the sim-nonstatDNS model is comparable to the performance of the random walk model. However, the random walk model only predicts certain bond and swap observations with different times to maturities. The random walk model can not be used to predict the swap spread. Therefore, among the three models in this thesis, the predictions from the sim-nonstatDNS model, produce the only good prediction for the swap spread.

7 Conclusion

The primary objective of a pension fund is to align current investments with future pension payments. However, this can be a challenging endeavor as the return on current investments is influenced by the bond rate, while the value of future pension payments is contingent upon the swap rate. In situations where these two rates undergo disparate fluctuations, a discrepancy arises between future pension payments and investment returns, potentially leading to significant issues for pensioners. To mitigate current and future potential mismatches, it is crucial to estimate and predict the spread between the bond rate and the swap rate. This spread is commonly known as the swap spread.

The goal of this thesis was to develop a methodology to predict the swap spread. The data consists of swap and bond observations with a mismatch in time to maturity. This implied the need for a model that models the swap curve and the bond curve as a continuous function of the time to maturity, which is the main motivation to use the DNS model in the prediction of the swap spread.

The application of DNS models before this thesis was limited to the estimation and prediction of one single interest rate. For the swap spread, it is necessary to estimate and predict two interest rates that are highly correlated: the bond rate and the swap rate. A new variant of the DNS model has been proposed to incorporate the correlation of these two rates. This variant models the latent factors of both curves in one restricted VAR model, in contrast to the original DNS model that models the latent factors by several univariate AR models.
In the simulation study, it has been shown that this sim-statDNS variant outperforms both the statDNS and the benchmark model for at least 70% of the one-day ahead prediction and at least 70% of the multiple-days ahead predictions.

The original DNS model as well as the sim-statDNS variant assumes the latent factors to follow a stationary process. Using the interpretation of the latent factors, it has been shown that the stationarity of the empirical factors computed from the data set is a good representation of the stationarity of the actual latent factors. Because the empirical factors of the real data were not stationary, modeling their latent factors by a stationary process would be incorrect.
This motivates the development of the second and third variants of the DNS model in this thesis. The second variant, nonstatDNS, independently predicts the swap and bond rate by modeling the first-order differences of the latent factors by univariate AR(1) models. The third variant, sim-nonstatDNS, simultaneously predicts the swap and bond rate by modeling the first-order differences of the latent factors by a restricted VAR(1) model. .
In the simulation study, it has been shown that the nonstatDNS model has the highest accuracy in only 13% of the cases for the 1-day ahead prediction and only 25% of the cases for the multiple-days ahead predicted curves. For the one-day ahead prediction, the sim-nonstatDNS outperforms both the benchmark model and the nonstatDNS model in 67% of the cases. For the multiple-days ahead prediction, the sim-nonstatDNS model has the highest accuracy among all three models in 38% of the cases. This implies that the sim-nonstatDNS model has higher accuracy than the benchmark model.

The nonstatDNS model and the sim-nonstatDNS model were applied to the real data set in the case study. following the methodology implied by the simulations. On the real data, the nonstatDNS has a much worse performance than both the sim-nonstatDNS and the benchmark model. This confirms the results in the simulations. The sim-nonstatDNS model outperforms the benchmark model on the prediction accuracy of the bond curves. However, the benchmark model has a better performance on the predicted swap curves.
The benchmark model only predicts the swap and bond curve for certain times to maturity, and can not be used to compute a prediction for the swap spread. The DNS model variants predict the swap and bond curve as a continuous function for the time to maturity and can therefore be used to compute the swap spread prediction. Despite the fact that the sim-nonstatDNS model does not always outperform the benchmark model, the sim-nonstatDNS model is the best model to predict the swap spread.

In addition to this main conclusion, the simulation study and the case study lead to several other key conclusions, which will be formulated in the sequel of this section.

The first conclusion is the comparison of the model behavior in the simulation studies, relative to the model behavior in the case study. The error terms in the DNS model are assumed to be normally distributed. This assumption is not violated in the simulation studies, but the residuals of the nonstatDNS model in the case study violate this assumption. Furthermore, the curves estimated and predicted by the sim-nonstatDNS model have heteroskedastic residuals, whereas this is not the case in the simulation studies. This will be discussed more in-depth in the discussion section.

A part of the selection process is the assumption of the stationarity of the latent factors. If these factors follow a nonstationary process, it is incorrect to model the latent factors by a stationary autoregressive model. Therefore it is of great importance to analyze the stationarity of the latent factors. In the simulation studies, it is investigated to what extent the interpretation of the DNS factors can be used to represent the stationarity of the latent factors. It is shown that the stationarity of the empirical DNS factors is an accurate representation of the stationarity of the true DNS factors.
This leads to the key conclusion for selecting a DNS model variant to predict the swap spread. If the empirical factors are stationary, the swap spread is predicted using a DNS model that assumes a restricted VAR(1) model on the latent factors. If the first-order differences of the empirical factors are stationary, the swap spread is predicted using a DNS model that assumes a restricted VAR(1) model on the first-order differences of the latent factors.

The second conclusion concerns the empirical factors. Although the stationarity of the empirical factors is a good representation of the stationarity of the latent factors, other systematic relationships of the true latent factors are not captured in these empirical factors. The ACF, PACF, and CCF of the empirical factors deviate from the ACF, PACF, and CCF of the true latent factors. The analysis of these systematic relationships will therefore not be included in the methodology.

The third conclusion concerns the estimate for the parameter λ in the DNS models. If the estimate $\hat{\lambda}_{MLE}$ is extremely small, the factors in the DNS model become linearly dependent. This leads to unstable factor estimates and inaccurate or incorrect predictions for the interest rate curves. If the estimate for $\hat{\lambda}_{MLE}$ is extremely large, the weight on the slope and curvature factor is approximately zero except for very small times to maturity. This results in interest rate curves that can not capture the structure of the actual interest rates. These conclusions are captured in the methodology by rejecting all DNS models that have an estimate for λ that is smaller than 0.01 or larger than 10.

Different initial values for λ result in different estimates for $\hat{\lambda}_{MLE}$ for the same set of observations. This implies that there are multiple local maxima in the log-likelihood surface and the initial value for λ in combination with the BFGS optimization algorithm determines in which local maximum the final parameter estimates are specified. Although there is a relation between the initial value and the final estimate, the specific relation can not be determined from the simulation studies. For that reason, a specific methodology to choose the initial value is absent. Any initial value between 0.01 and 10 can be used.

The real data misses observations on days that the market was closed. It can be concluded that these missing observations cause a decrease in accuracy and an increase in uncertainty of the predicted interest rates. Since the effect on the accuracy and uncertainty is not significantly large, it is decided that in the preprocessing of the data, these missing observations will not be imputed and will be treated as missing observations. The Kalman filter is able to handle missing observations and therefore this choice does not lead to a violation of model assumptions.

The bond data have observations with various times to maturity, which leads to computational issues in the parameter estimation of the DNS models. For that reason, the bond data is preprocessed using smoothing splines to obtain a set of observations with equal times to maturity for each date in the data set. Based on the results of the simulation studies it can be concluded that this preprocessing step does not decrease the accuracy of the DNS model.

8 Discussion

This chapter outlines the points for discussion that have arisen from the study performed in this thesis. Some of these points can be considered topics for further research.

First of all, the accuracy of the models for the swap spread is measured by the accuracy of the bond rate and swap rate separately. The choice for this performance measure is motivated by the absence of a true swap spread due to two data sets with different times to maturity. However, this performance measure for the swap spread could be a misleading representation of the actual performance. For example, if both the swap curve and bond curve are overestimated by the same amount, the difference between the rates is very accurate. On the other hand, if one of the curves is over-estimated and the other curve is under-estimated, the accuracy of the swap spread prediction is worse than the accuracy measured by the rMSE on both rates separately. For this reason, it is recommended to develop a benchmark model that can be used to measure the accuracy of the swap spread predictions.

The second limitation concerns the interpretation and computation of the empirical factors. Although the stationarity of the empirical factors is an accurate representation of the stationarity of the true factors, other systematic relationships of the true factors are not captured in the empirical factors. The autocorrelation, partial autocorrelation, and cross-correlation functions of the empirical factors did not correspond to the true underlying factor dynamics.
The absence of information about the systematic relationships of the latent factors is a limitation of the information that can be used to choose the best-fitting time series model for the latent factors. The choice to model the dynamics by autoregressive processes with only one lag could be a simplification of the true process of the latent factors. Although this results in less accurate results, this choice does not lead to a violation of model assumptions. A way to improve the DNS variants in this thesis is to find other representations for DNS factors that capture the ACF, PACF, and CCF of the true DNS factors.

The DNS model variants in this thesis are limited to two types of stationarity for the factor dynamics. The factor dynamics are either assumed to be stationary or the first-order differences are assumed to be stationary. In practice, however, differencing a nonstationary time series does not always lead to a stationary time series. It could be the case that other transformations of the time series are needed in order to make the series of factors stationary. If such a transformation is a linear transformation, the model can be rewritten as a state space model using the same method as the method used in this thesis.

Subsequently, the parameter λ in the DNS models is assumed to be constant over time. The estimate for λ determines the weight of the factors and thereby determines the shape of the interest rate curves. It was shown that the shape of the interest rate curves is not constant over time, which suggests a time-varying parameter λ. Koopman et al. (2007) have developed a variant of the original DNS model with such a time-varying λ. A first recommendation for further research is to investigate the possibilities for a time-varying λ for the new variants of the DNS model presented in this thesis.

Also, it is important to remark that the estimation of λ served as the main bottleneck in the application of all DNS models. In the simulation studies as well as in the case study, the estimate for λ was unstable and could take extremely large or extremely small values, leading to a violation of the model assumptions. In this thesis, the BFGS algorithm with a finite difference approximation for the gradients was used to obtain the maximum likelihood estimates. The resulting estimates for λ suggest that this is not the best method to use for this application. For that reason, it is recommended to perform further research on what optimization algorithm to use to estimate the DNS model parameters.
A more specific recommendation for improving the BFGS algorithm is as follows. In the section related to the BFGS algorithm, it was mentioned that the BFGS algorithm could be improved by taking into account the theoretical gradient functions (Durbin & Koopman, 2012a) for several parameters including λ. This will improve the algorithm by determining the change in the log-likelihood value with respect to λ. Another option is to determine the boundaries for λ that specify the values for which the model can not be used. These boundaries can be used to develop a bounded optimization algorithm for λ. Finally, Koopman et al. (2007) used the EM algorithm to obtain parameter estimates for the original DNS model. The usage of the EM algorithm could also be investigated as a suitable optimization method for the DNS models in this thesis.

Throughout all simulations, it was assumed that the true data has the structure of a DNS model. It is not investigated how the DNS variants perform on a data set with a different underlying model. In addition to the validation in this thesis, it is recommended to perform several simulation studies where the data has an underlying model that is not a DNS model.

In this thesis, the main objective was to predict the swap spread curve. The spread is reconstructed from the bond and swap curves. The application of the Kalman filter makes it possible to construct a prediction interval for the bond and swap curves. However, it is not investigated how to construct a prediction interval for the spread curves. Since the uncertainty for the spread is very important, this should be investigated as well.

The formulation of the DNS model as a linear Gaussian state space model assumes that the state vectors are i.i.d. normally distributed, conditionally on the previous observations. The state space form also assumes that interest rate observations are i.i.d. normally distributed conditionally on the previous observations. This is a limitation of the model. If the errors are not i.i.d. normally distributed, the model could be extended to a model with correlated errors (Shumway & Stoffer, 2016a). The model can also be extended to a model with non-normal errors (Durbin & Koopman, 2012a).

The residuals of sim-nonstatDNS model predictions for the bond and swap curves in the case study did not pass the test for homoskedasticity. Less than 50% of the estimated curves and less than 30% of the predicted curves have homoskedastic residuals. In all other cases, there is significant proof for heteroskedastic residuals. Heteroskedastic residuals suggest that there is a variance in the interest rate observations that is not explained by the three factors. The heteroskedasticity is an important result for the application of the DNS models because heteroskedasticity could lead to biased factor estimates and misleading factor interpretation. This result is in contrast with the simulation studies, where the majority of the interest rate curves have homoskedastic errors. That means that, if the data actually has the structure of the DNS model, the three factors capture all variability in the interest rates.
A possible way to overcome heteroskedasticity is to apply variance-stabilizing transformations on the factor loadings. A commonly used transformation for this purpose is a log or power transformation. Such a transformation can be included in the state space model by reformulating the factor loadings as a function of the parameter λ. Since this function is already nonlinear, a nonlinear transformation would not harm the structure of the state space model.
Another option is to use a more advanced model, such as the NSS-model (Svensson, 1994). Finally, the restriction on λ to be constant over time could be retrieved in order to make the model more advanced (Koopman et al., 2007).

The final recommendation concerns the preprocessing of the data. The decision to spline the data and not impute the missing observations result in a decreased accuracy of the DNS model. That suggests that the DNS model could be improved by other choices in the preprocessing steps.

References

Berk, J., & DeMarzo, P. (2017). *Corporate finance*. Pearson.

Björk, T., & Christensen, B. J. (1999). Interest rate dynamincs and consistent forward rate curves. *Mathematical Finance*, *9*(4), 323-348.

Christensen, J. H. E., Diebold, F. X., & Rudebusch, G. D. (2009). An arbitrage-free generalized nelson-siegel term structure model. *Econometrics Journal*, *12*(3), 33-64.

Christensen, J. H. E., Diebold, F. X., & Rudebusch, G. D. (2011). The affine arbitrage-free class of nelson-siegel term structure models. *Journal of Econometrics*, *164*(1), 4-20.

Cottle, R. W., & Thapa, M. N. (2017). *Linear and nonlinear optimization*. Springer.

Cox, J. C., Ingersoll, J. E., & Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica*, *53*(2), 385-408.

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

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

Durbin, J., & Koopman, S. (2012a). *Time series analysis by state space methods*. Oxford University Press.

Durbin, J., & Koopman, S. (2012b). *Time series analysis by state space methods*. Oxford University Press.

Fernandes, M., & Vieira, F. (2019). A dynamic nelson-siegel model with forward-looking macroeconomic factors for the yield curve in the us. *Journal of Economic Dynamics and Control*, *106*.

Filipović, D. (2009). *Term-structure models - a graduate course*. Springer.

Guidolin, M., & Pedio, M. (2019). Forecasting and trading monetary policy effects on the riskless yield curve with regime switching nelson–siegel models. *Journal of Economic Dynamics and Control*, *107*.

Hamilton, J. D. (1994). *Time series analysis*. Princeton Universtiy Press.

Hull, J., & White, A. (1990). Pricing interest-rate-derivative securities. *Review of Financial Studies*, *3*(4), 573-592.

Koopman, S. J., Mallee, M. I. P., & Van der Wel, M. (2007). Analyzing the term structure of interest rates using the dynamic nelson-siegel model with time varying parameters. *Tinbergen Institute Discussion Paper*, *4*.

Nelson, C. R., & Siegel, A. F. (1987). Parsimonious modeling of yield curves. *The Journal of Business*, *60*(4), 473-489.

Shaw, F., Murphy, F., & O'Brien, F. (2014). The forecasting efficiency of the dynamic nelson siegel model on credit default swaps. , *30*, 348-368.

Shumway, R. H., & Stoffer, D. S. (2016a). *Time series analysis and its applications*.

Shumway, R. H., & Stoffer, D. S. (2016b). *Time series analysis and its applications*.

Shumway, R. H., & Stoffer, D. S. (2016c). *Time series analysis and its applications*.

Svensson, L. E. (1994). Estimating and interpreting forward interest rates: Sweden, 1992-1994. *International Monetary Fund*.

Van Dijk, D., Koopman, S. J., Van der Wel, M., & Wright, J. H. (2014, October). Forecasting interest rates with shifting endpoints. *Journal of Applied Econometrics*, *29*, 693-712.

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

Xiang, J., & Zhu, X. (2013). A regime-switching nelson-siegel term structure model and interest rate forecast. *Journal of Financial Econometrics*, *11*(3), 522-555.

A List of economic definitions

Issuer	Entity, government or organization that offers or sells financial securities or instruments to investors in order to raise capital
Zero-coupon bond	A zero-coupon bond is a contract where an investor lends money to the issuer in exchange for a future payment.
Coupon bond	A coupon bond is a contract where an investor lends money to the issuer in exchange for multiple future interest payments.
Investor	Entity or individual that buys financial securities to obtain a return on its investments
Interest rate swap	An interest rate swap contract is a contract between two parties that defines a scheme where a payment stream at a fixed interest rate is exchanged for a payment stream of a floating interest rate.
Maturity date	The time when the bond issuer must repay the face value bond value to the bond holder.
Time to maturity	The time until the maturity date
Face value	The amount of money repaid to the holder at the maturity date
Coupon payment	Multiple interest payments of the bond issuer to the bondholder at a prespecified set of coupon dates
Coupon	The interest rate for the coupon payments
Yield to maturity	The return, positive or negative, of a bond contract when held until the maturity date.
Swap rate	The fixed interest rate of a swap contract

Table 41: List of economic definitions

B Data

B.1 Bootstrapping

Bootstrapping is a commonly employed method to derive a zero coupon yield curve from observed market data (Berk & DeMarzo, 2017). The same method can be used to obtain a zero coupon swap curve. It is important to have zero coupon observations, because that implies that the only cash flow for the observation is at the maturity date. In this part of the appendix, I will explain the bootstrap method that will be used to obtain zero coupon observations for the bond and swap data.

B.1.1 Bootstrap method for coupon bonds

The price of a coupon bond is equal to the sum of discounted cash flows. The price is related to the yield and the coupon payments through the following formula (Berk & DeMarzo, 2017)

$$P(t,T) = \sum_{i=1}^{n} c \cdot \left(\frac{1}{(1+y_i)^{\tau_i}} \right) + FV \cdot \left(\frac{1}{(1+y_n)^{\tau_n}} \right), \tag{10}$$

where $P(t,T)$ is the price at date t for a bond with maturity date T, c is the fixed coupon payment, y_i is the yield of a bond with maturity date τ_i in years, and τ_i corresponds to the time till the coupon dates. FV is the face value of the bond, τ_n is the time to maturity and y_n is the corresponding zero coupon yield of the bond.
In Equation 10, the price $P(t,T)$, the value for the coupon payments c, the time till the coupon dates y_i for $i = 1, \ldots, n$, and the face value FV are observed. Only the zero coupon yields y_i for $i = 1, \ldots, n$ are unknown.

The bootstrapping method aims to compute the zero coupon yield y_n for the maturity date τ_n for all observed coupon bonds. If the values for y_i, $i = 1, \ldots, n-1$ are known, the zero coupon yield y_n can be computed using Equation 10. The values for y_i, $i = 1, \ldots, n$ are obtained from the zero coupon bond observations through linear interpolation.
The iterative process to compute the zero coupon yields for all observed coupon bonds is as follows. Choose the coupon bond observation with the smallest time to maturity τ_n. For this observation, determine the dates of the annual coupon payments τ_i. For each coupon date, determine the zero coupon yield from the zero coupon bond observations. If there is a zero coupon bond observation with a time to maturity equal to τ_i, the yield of this observation can be used directly. If there is no zero coupon bond observation in the data set that has a matching time to maturity, I use the two nearest points along the time to maturity axis and interpolate the yield by linear interpolation.
When bootstrapping, the clean price is typically used in the calculations to ensure consistency in valuing different bonds and deriving the implied yields. It provides a more accurate representation of the market's perception of the bond's value at a specific point in time.

B.1.2 Bootstrap method for interest rate swaps

The bootstrapping method for the swap contract is explained in this part of the appendix. The bootstrapping process starts with a 1-year swap contract that has cash flow dates as in Table 42, where $r_{0.5}$ is the 6M-Euribor rate at $t = 0$ over 1 year, r_1 is the 6M-Euribor rate at $t = 0.5$ over a period of 1 year, and q_1 is the swap rate quoted in the data set. By the definition of a swap rate, the net present value of these cashflows is equal to

	Fixed lag	Floating lag
$t = 0.5$		$r_{0.5} \cdot FV$
$t = 1$	$q_1 \cdot FV + FV$	$r_1 \cdot FV + FV$

Table 42: Cash flows of a 1-year swap contract

zero, i.e.

$$0 = d_{0.5} \cdot r_{0.5} + d_1 \left(r_1 - q_1 \right),$$

where $d_{0.5}$ is the 6M-Euribor at $t = 0$ over 0.5 years, d_1 is the 6M-Euribor rate at $t = 0.5$ over a period of 1 year, and FV is left out of the equation. By definition, it holds that $d_1 = r_1$ and $(d_{0.5})^2 = r_{0.5}$. Furthermore, for the bootstrapping procedure, it is assumed that the 6M-Euribor at t=0 is equal to the 6M-Euribor at $t = 0.5$. This simplifies the cashflow equation to

$$0 = d_1^2 + d_1 - q_1$$

Since the value of q_1 is known from the data, this equation can be used to obtain the value for d_1. Then, the zero coupon swap rate is the rate that sets the present values of the cash flows without the coupon payments

equal to zero. Thus, the zero coupon swap rate $\tilde{q}_1$ can be computed from

$$0 = d_1 \left(r_1 - \tilde{q}_1\right),$$

where we use $d_1 = r_1$.

For a 2-year swap contract, the cash flows are given in Table 43. The net present value equation for the cash

	Fixed lag	Floating lag
$t = 0.5$		$r_{0.5} \cdot FV$
$t = 1$	$q_1 \cdot FV$	$r_1 \cdot FV$
$t = 1.5$		$r_{1.5} \cdot FV$
$t = 2$	$q_2 \cdot FV + FV$	$r_2 \cdot FV + FV$

Table 43: Cash flows of a 1-year swap contract

flows is

$$0 = d_{0.5} \cdot r_{0.5} + d_1 \left(r_1 - q_1\right) + d_{1.5} \cdot r_{1.5} + d_2 \left(r_2 - q_2\right)$$

where $(d_{0.5})^2 = r_{0.5}$, $d_1 = r_1$, $(d_{1.5})^{\frac{1}{1.5}} = r_{1.5}$, and $(d_2)^{\frac{1}{2}} = r_2$. The values for $d_{0.5}$ and d_1 are obtained from the 1-year swap contract. It is assumed that $d_{1.5}$ can be linearly interpolated between d_2 and d_1. That is, $d_{1.5} = d_1 + (1.5 - 1)\frac{d_2 - d_1}{2-1}$. Since the swap rates q_1 and q_2 are known from the data, the value of d_2 is the only unknown in this equation. The equation is solved for d_2. Then, the zero coupon 2-year swap rate $\tilde{q}_2$ can be computed by

$$0 = d_2(r_2 - \tilde{q}_2),$$

where we use that $(d_2)^{\frac{1}{2}} = r_2$. Following the same procedure for swap rates with a time to maturity of $3, 4, 5, \ldots$ years results in a zero-coupon swap curve.

B.2 Approximation of the bond observation using smoothing splines

The bond observations are preprocessed by estimating the observations for a fixed set of maturities. In this part of the appendix, the procedure for this preprocessing step will be explained.
For a set of interest rate observations (y_i, τ_i) the cubic smoothing spline estimate $f(\tau)$ is defined by the minimizer of

$$\sum_{i=1}^{n} \{y_i - \hat{f}(\tau_i)\}^2 + \lambda \int \hat{f}''(\tau)^2 d\tau.$$

The splines that are estimated to obtain the bond observations use the 4 observations with a time to maturity that is closest to the time to maturity of the objective observation. This procedure is visualized in Figure 42. The splines in this thesis are estimated with the function `smooth.spline` of the `stats`-package.

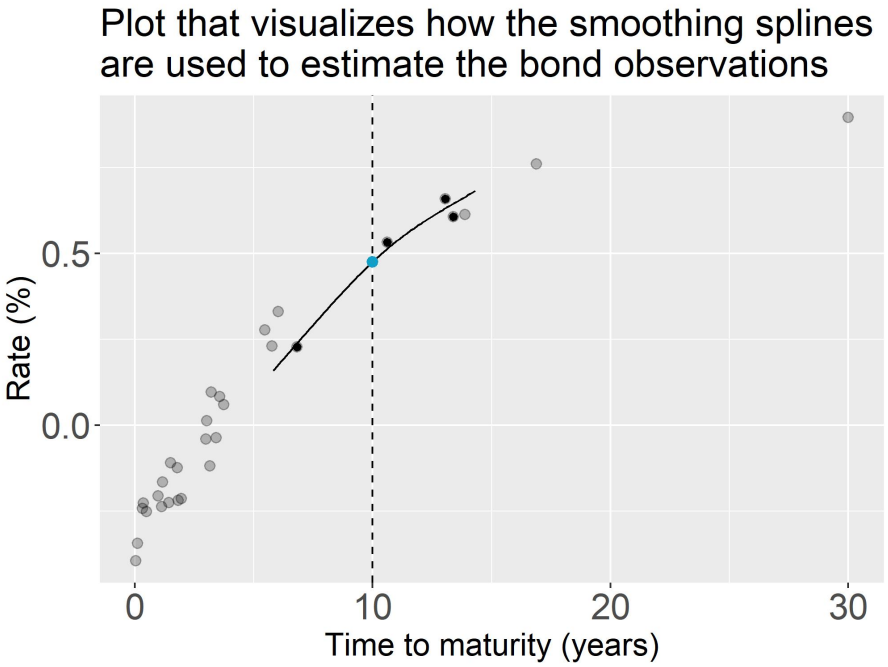


Figure 42: A visualization of the observation estimated by a smoothing spline

C Proofs

C.1 Convergence of the factor loadings

C.1.1 Convergence of the β_2 loading

The convergence of the second loading to zero for infinitely large values of τ:

$$\lim_{\tau\to\infty}\left(\frac{1-e^{-\tau\lambda}}{\tau\lambda}\right)=0 \tag{11a}$$

The convergence of the second loading to one for infinitely small values of τ:

$$\lim_{\tau\to 0}\left(\frac{1-e^{-\tau\lambda}}{\tau\lambda}\right)\overset{L.H.}{=}\lim_{\tau\to 0}\left(\frac{\lambda\cdot e^{-\lambda\tau}}{\lambda}\right)=\lim_{\tau\to 0}e^{-\lambda\tau}=1, \tag{11b}$$

where $L.H.$ indicates that L'Hôpitals rule is applied.

C.1.2 Convergence of the β_3 loading

The convergence of the third loading to zero for infinitely large values of τ:

$$\lim_{\tau\to\infty}\left(\frac{1-e^{-\tau\lambda}}{\tau\lambda}-e^{-\tau\lambda}\right)=\lim_{\tau\to\infty}\left(\frac{1-e^{-\tau\lambda}}{\tau\lambda}\right)-\lim_{\tau\to\infty}\left(e^{-\tau\lambda}\right)=0-0=0 \tag{12a}$$

The convergence of the third loading to zero for infinitely small values of τ:

$$\begin{aligned}\lim_{\tau\to 0}\left(\frac{1-e^{-\tau\lambda}}{\tau\lambda}-e^{-\tau\lambda}\right)&=\lim_{\tau\to 0}\left(\frac{1-e^{-\tau\lambda}}{\tau\lambda}\right)-\lim_{\tau\to 0}\left(e^{-\tau\lambda}\right)\\&\overset{L.H.}{=}\lim_{\tau\to 0}\left(\frac{\lambda\cdot e^{-\tau\lambda}}{\lambda}\right)-\lim_{\tau\to 0}\left(e^{-\tau\lambda}\right)\\&=1-1\\&=0,\end{aligned} \tag{12b}$$

where $L.H.$ indicates that L'Hôpitals rule is applied.

C.2 The interpretation of the factors β_1 and β_2 as level and slope

C.2.1 The level factor

The interpretation of β_1 as the level factor:

$$\begin{aligned}\lim_{\tau\to\infty}y(\tau)&=\lim_{\tau\to\infty}\left(\beta_1+\beta_2\cdot\left(\frac{1-e^{-\tau\lambda}}{\tau\lambda}\right)+\beta_3\cdot\left(\frac{1-e^{-\tau\lambda}}{\tau\lambda}-e^{-\tau\lambda}\right)\right)\\&=\beta_1+\beta_2\cdot\lim_{\tau\to\infty}\left(\frac{1-e^{-\tau\lambda}}{\tau\lambda}\right)+\beta_3\cdot\lim_{\tau\to\infty}\left(\frac{1-e^{-\tau\lambda}}{\tau\lambda}-e^{-\tau\lambda}\right)\\&\downarrow\text{ Equation 11a and Equation 12a}\\&=\beta_1+\beta_2\cdot 0+\beta_3\cdot(0-0)\\&=\beta_1.\end{aligned} \tag{13}$$

C.2.2 The slope factor

The interest rate for infinitely small values of τ is given by

$$\begin{aligned}\lim_{\tau\to 0}y(\tau)&=\lim_{\tau\to 0}\left(\beta_1+\beta_2\cdot\left(\frac{1-e^{-\tau\lambda}}{\tau\lambda}\right)+\beta_3\cdot\left(\frac{1-e^{-\tau\lambda}}{\tau\lambda}-e^{-\tau\lambda}\right)\right)\\&=\beta_1+\beta_2\cdot\lim_{\tau\to 0}\left(\frac{1-e^{-\tau\lambda}}{\tau\lambda}\right)+\beta_3\cdot\lim_{\tau\to 0}\left(\frac{1-e^{-\tau\lambda}}{\tau\lambda}-e^{-\tau\lambda}\right)\\&\downarrow\text{ Equation 11b and Equation 12b}\\&=\beta_1+\beta_2\cdot 1+\beta_3\cdot(1-1)\\&=\beta_1+\beta_2.\end{aligned} \tag{14}$$

Combine the results in Equation 13 and Equation 14 to find the interpretation of β_2 as the slope factor:

$$\lim_{\tau\to\infty}y(\tau)-\lim_{\tau\to 0}y(\tau)=\beta_1-(\beta_1+\beta_2)=-\beta_2$$

D Statistical definitions, tests and methodology

D.1 Definitions for time series analysis

All definitions in this part of the appendix are defined in Shumway and Stoffer (2016b), except if specified differently.

D.1.1 Stationarity

Definition D.1 (Weakly stationary or stationary)**.** A time series x_t is a finite variance process such that

1. the mean value function, μ_t is constant and does not depend on time t, and
2. the autocovariance function, $\gamma(s,t)$, depends on s and on t only through their difference $|s-t|$.

Theorem D.1 (Stationary AR(1) process)**.** Consider a first-order autoregression

$$\beta_t = c + \phi\beta_{t-1} + \eta_t.$$

When $|\phi| \geq 1$, there does not exist a stationary process for β_t with finite variance. When $|\phi| < 1$, there is a stationary process for β_t satisfying the AR(1) model. Hamilton (1994)

D.1.2 Autocovariance and autocorrelation

Definition D.2 (Autocovariance)**.** The autocovariance function measures the linear dependece between x_s and x_t by

$$\gamma_x(s,t) = cov\left(x_s, x_t\right) = \mathbb{E}\left[\left(x_s - \mu_s\right)\left(x_t - \mu_t\right)\right]$$

The autocovariance is estimated from the data by the sample autocovariance function

$$\hat{\gamma}(h) = \frac{1}{n}\sum_{t=1}^{n-h}\left(x_{t+h} - \overline{x}\right)\left(x_t - \overline{x}\right)$$

Definition D.3 (Autocorrelation (ACF))**.** The autocorrelation function measures the predictability of the series at time t using only the value x_s by

$$\rho(s,t) = \frac{\gamma(s,t)}{\sqrt{\gamma(s,s)\gamma(t,t)}}$$

The autocorrelation function is estimated from the data by the sample autocorrelation function

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$$

D.1.3 Partial autocorrelation

If the process is an autoregressive process, the ACF alone tells us little about the orders of dependence. Therefore, partial autocorrelation will be used, which is the correlation between x_s and x_t with the linear effect of everything in the middle removed.

Definition D.4 (Partial autocorrelation (PACF) (Shumway & Stoffer, 2016c))**.** The partial autocorrelation function (PACF) of a stationary process x_t, denoted ϕ_{hh}, for $h = 1, 2, \ldots$, is

$$\begin{aligned}\phi_{11} &= \operatorname{corr}\left(x_{t+1}, x_t\right) = \rho(1),\\ \phi_{hh} &= \operatorname{corr}\left(x_{t+1} - \hat{x}_{t+1}, x_t - \hat{x}_t\right), \quad h \geq 2\end{aligned}$$

D.1.4 Cross covariance and cross correlation

Definition D.5 (Cross covariance)**.** The cross-covariance function between two series x_t and y_t is

$$\gamma_{xy}(s,t) = \text{Cov}\left[x_s, y_t\right] = \mathbb{E}\left[\left(x_s - \mu_{xs}\right)\left(y_t - \mu_{yt}\right)\right]$$

The cross-covariance is estimated from the data by the sample cross-covariance function

$$\hat{\gamma}_{xy} = \frac{1}{n}\sum_{t=1}^{n-h}\left(x_{t+1} - \overline{x}\right)\left(y_t - \overline{y}\right)$$

Definition D.6 (Cross-correlation function (CCF))**.** The cross-correlation function (CCF) of time series x_t and y_t is defined as

$$\rho_{xy}(s,t) = \frac{\gamma_{xy}(s,t)}{\sqrt{\gamma_x(s,s)\gamma_y(t,t)}},$$

where we have $-1 \leq \rho_{xy}(h) \leq 1$.

The cross-correlation is estimated from the data by the sample cross-correlation function

$$\hat{\rho}_{xy}(h) = \frac{\hat{\gamma}_{xy}(h)}{\sqrt{\hat{\gamma}_x(0)\hat{\gamma}_y(0)}}$$

D.2 Interpretation of ACF and PACF plots

The following table specifies the behavior of the ACF and PACF for ARMA models (Shumway & Stoffer, 2016c).

	AR(p)	MA(q)	ARMA(p,q)
ACF	Tails off	Cuts off after lag q	Tails off
PACF	Cuts off after lag p	Tails off	Tails off

Table 44: The interpretation of ACF and PACF plots used to choose the best time series models.

D.3 Akaikes Information Criterion

Akaikes Information Criterion (AIC) measures the relative goodness of fit for a particular model on the data. The AIC balances its measure between the error of the model and a penalized number of parameters. The AIC measure is defined by

$$AIC = 2 \cdot K - 2 \cdot \ln\left(L\right),$$

where K denotes the number of independent variables in the model and L denotes the log likelihood that can be achieved by the model on a given data set.

The AIC is a measure that provides the relative goodness of fit of the model on a data set. The rule of thumb is that one model has an AIC of two units lower than another, the model is a significantly better fit for the specific data set.

To determine the AIC of the models on the data in this thesis, we used the function `VARselect` from the R-package `vars`. This function takes the time series data as input data and returns the number of lags for the autoregressive model that has the lowest AIC value.

D.4 Augmented Dicky-Fuller (ADF) test

The Augmented Dicky-Fuller (ADF) test is a statistical test that is used to determine whether or not time series data is covariance stationary. The ADF-test tests is a unit root test for

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \ldots + \delta_{p-1}\Delta y_{t-p+1} + \varepsilon_t,$$

where y_t is an observation from the time series data, α is a constant trend, β the time-dependent trend and p is the order of the autoregressive model, i.e. the number of lagged values. To use the ADF test, the number of autoregressive lags need to be determined, for example using AIC.

If $\gamma < 0$, the time series is considered as a stationary time series.
The null hypothesis and the alternative hypothesis are defined as

$$H_0 : \gamma = 0, \text{ i.e. the time series is not stationary}$$
$$H_A : \gamma < 0, \text{ i.e. the time series is stationary}$$

The ADF test statistic is given by

$$DF_\tau = \frac{\hat{\gamma}}{\text{SE}\hat{\gamma}}$$

A more negative test statistic DF_τ gives a stronger rejection of the null hypothesis. If the calculated test statistic is more negative than a critical value, the null hypothesis is rejected. Notice that in order to reject H_0, the p-value of the ADF test should be less than a critical value.

The ADF test is implemented in R in the function `adf.test` in the `tseries`-package. In this thesis we have used this implementation to determine whether or not our time series data is a stationary process.

D.5 Q-Q plot

A Q-Q plot is a method to compare two probability distributions. In the application within this thesis, the probability distribution of the residuals is compared with a theoretical normal distribution. A point in the Q-Q plot corresponds to one of the quantiles of the theoretical normal distribution on the x-axis plotted against the same quantile of the residuals distribution on the y-axis.

If the residuals have a normal distribution, the points should lie the straight line $x = y$. This line is represented by the red line in the Q-Q plot. See Figure 43 for an example of a Q-Q plot. There is a pre-implemented Q-Q plot function available in R: `qqplot`.

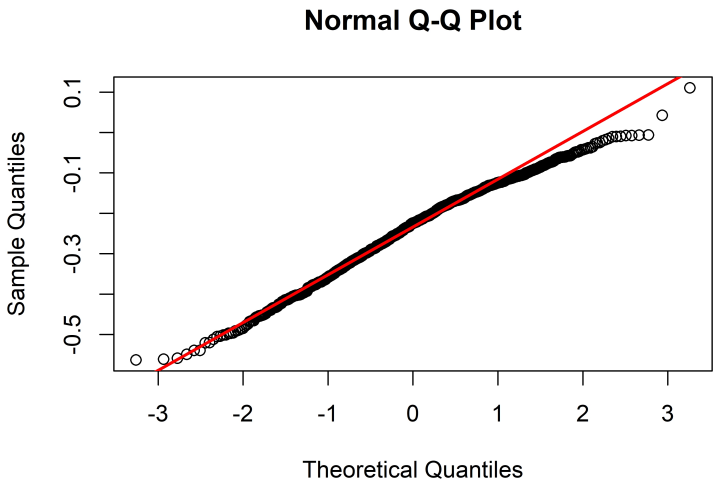


Figure 43: Example of a Q-Q plot for the residuals

D.6 Shapiro-Wilk test

The Shapiro-Wilk (SW) test is a test for normality and will be used to assess the distribution of a set of residuals $r_1, \ldots, r_n$. The SW test uses a zero hypothesis H_0: the sample $r_1, \ldots, r_n$ comes from a normally distributed population. The test statistic for the SW test is given by

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)}{\sum_{i=1}^{n} (x_i - \overline{x})^2}$$

where $a_i := \frac{m^T V^{-1}}{||V^{-1} m||}$, m is the expected value and V is the covariance matrix of the order statistics of i.i.d. standard normal random variables.
For the SW test the rule of thumb is to reject H_0 if the p-value is less than a certain α-level. In this thesis, H_0 is rejected if the p-value is less than 0.05. The SW test is pre-implemented in R by the function `shapiro.test`.

D.7 Breusch-Pagan test

In order to assess the presence of heteroskedasticity, or variance of errors, in the Nelson-Siegel model, I conducted the Breusch-Pagan test. This test examines whether the variability of the residuals in the model was constant or varied across different levels of the independent variables. The test tests whether the variance of the errors from a regression is dependent on the values of the independent variables. In that case, heteroskedasticity is present. The Breusch-Pagan test is an hypothesis test where H_0 assumes homoskedasticity and the alternative hypothesis assumes heteroskedasticity.

The Breusch-Pagan test statistic BP is obtained by following these steps for one estimated interest rate curve:

1. Obtain the estimated interest rates from the estimated DNS model,
   $\hat{y}_t(\tau_i) = \hat{\beta}_{1,t} + \hat{l}_{2,i} \cdot \hat{\beta}_{2,t} + \hat{l}_{3,i} \cdot \hat{\beta}_{3,t}$,
   where $\hat{l}_{2,i}$ and $\hat{l}_{3,i}$ are the factor loadings in Equation 2 corresponding to the interest rate with a time to maturity of τ_i years, using the estimated $\hat{\lambda}$, and $\beta k, t$, for $k = 1, 2, 3$ are the estimated factors.

2. Compute the residuals of the DNS model,
   $\hat{\varepsilon}_{i,t} = y_t(\tau_i) - \hat{y}_t(\tau_i)$

3. Compute the dependent variable g_i defined by
   $g_{i,t} = \frac{\hat{\varepsilon}^2_{i,t}}{\hat{\sigma}^2_t}$, where $\hat{\sigma}^2_t = \sum_{i=1}^{n} \frac{\hat{\varepsilon}^2_{i,t}}{n}$,
   for n the number of observations at date t

4. Estimate the auxiliary regression
   $g_{i,t} = \gamma_1 + \hat{l}_{2,i} \cdot \gamma_2 + \hat{l}_{3,i} \cdot \gamma_3$,
   where $\hat{l}_{2,i}$ and $\hat{l}_{3,i}$ are the factor loadings equal to the factor loadings used in step 1.

5. Compute the test statistic BP by
   $BP_t = \frac{1}{2}(TSS_t - SSR_t)$,
   where TSS_t is the sum of squared deviations of the $g_{i,t}$ from 1 and SSR_t is the sum of squared residuals from the auxiliary regression.

6. Compare the obtained test statistic BP_t to the critical value corresponding to the critical value of a χ^2_1 distribution, i.e. the Chi-squared distribution with 2 degrees of freedom. If the test statistic BP is below the critical value, there is not enough evidence to reject H_0, i.e. there is not enough evidence to reject homoskedasticity.

E DNS models as state space models

In this part of the appendix, the proposed variations of the DNS model are formulated in the state space representation given in Equation 8a-8e. For convenience, the state space model is repeated below,

$$
\begin{aligned}
b_t &= c + \Phi b_{t-1} + \eta_t \\
y_t &= \Lambda_t b_t + \varepsilon_t \\
\eta_t &\sim \mathcal{N}(0, Q) \\
\varepsilon_t &\sim \mathcal{N}(0, H_t) \\
b_0 &\sim \mathcal{N}\left(\bar{b}_0, P_0\right).
\end{aligned}
$$

E.1 DNS variation with non stationary factor dynamics as state space model

The DNS variation where the factors transformed by two times the linear difference operator are modeled as a VAR(1) model (Equation 6) in state space representation is

$$
\begin{pmatrix} \beta_{1,t} \\ \beta_{2,t} \\ \beta_{3,t} \\ \beta_{1,t-1} \\ \beta_{2,t-1} \\ \beta_{3,t-1} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \tilde{\Phi} + I & -\tilde{\Phi} \\ I & \emptyset \end{pmatrix} \begin{pmatrix} \beta_{1,t-1} \\ \beta_{2,t-1} \\ \beta_{3,t-1} \\ \beta_{1,t-2} \\ \beta_{2,t-2} \\ \beta_{3,t-2} \end{pmatrix} + \begin{pmatrix} \eta_{1,t} \\ \eta_{2,t} \\ \eta_{3,t} \\ 0 \\ 0 \\ 0 \end{pmatrix} \tag{15a}
$$

$$
\begin{pmatrix} y_t(\tau_1) \\ y_t(\tau_2) \\ \vdots \\ y_t(\tau_n) \end{pmatrix} = \begin{pmatrix} 1 & \frac{1-e^{-\tau_1\lambda}}{\tau_1\lambda} & \frac{1-e^{-\tau_1\lambda}}{\tau_1\lambda} - e^{-\tau_1\lambda} & 0 & 0 & 0 \\ 1 & \frac{1-e^{-\tau_2\lambda}}{\tau_2\lambda} & \frac{1-e^{-\tau_2\lambda}}{\tau_2\lambda} - e^{-\tau_2\lambda} & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\tau_n\lambda}}{\tau_n\lambda} & \frac{1-e^{-\tau_n\lambda}}{\tau_n\lambda} - e^{-\tau_n\lambda} & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \beta_{1,t} \\ \beta_{2,t} \\ \beta_{3,t} \\ \beta_{1,t-1} \\ \beta_{2,t-1} \\ \beta_{3,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \vdots \\ \varepsilon_{n,t} \end{pmatrix} \tag{15b}
$$

where $\tilde{\Phi}$ is a 3×3 diagonal matrix with diagonal elements ϕ_{11}, ϕ_{22} and ϕ_{33}, and I is the 3×3 identity matrix. We have initial conditions

$$
\eta_t \overset{iid}{\sim} \mathcal{N}(0, Q_t) \tag{15c}
$$

$$
\varepsilon_t \overset{iid}{\sim} \mathcal{N}(0, H_t) \tag{15d}
$$

$$
\begin{pmatrix} \beta_{1,t} \\ \beta_{2,t} \\ \beta_{3,t} \\ \beta_{1,t-1} \\ \beta_{2,t-1} \\ \beta_{3,t-1} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \bar{b}_0 \\ \bar{b}_0 \end{pmatrix}, \begin{pmatrix} P_0 & 0 \\ 0 & P_0 \end{pmatrix} \right), \tag{15e}
$$

with Q_t a 3×3 diagonal matrix, H_t a $n \times n$ diagonal matrix, $\bar{b}_0$ a 3×1 vector and P_0 a 3×3 diagonal matrix.

E.2 DNS variation with factor dynamics of two rates in one model as state space model

The DNS model that models the factors of both the swap and bond curves in one VAR model (Equation 5) is

$$\begin{pmatrix}\beta_{1,t,s}\\ \beta_{2,t,s}\\ \beta_{3,t,s}\\ \beta_{1,t,b}\\ \beta_{2,t,b}\\ \beta_{3,t,b}\end{pmatrix} = \begin{pmatrix}c_{1,s}\\ c_{2,s}\\ c_{3,s}\\ c_{1,b}\\ c_{2,b}\\ c_{3,b}\end{pmatrix} + \begin{pmatrix}\phi_{11} & & & \phi_{14} & & \\ & \phi_{22} & & & \phi_{25} & \\ & & \phi_{33} & & & \phi_{36}\\ \phi_{41} & & & \phi_{44} & & \\ & \phi_{52} & & & \phi_{55} & \\ & & \phi_{63} & & & \phi_{66}\end{pmatrix} \begin{pmatrix}\beta_{1,t-1,s}\\ \beta_{2,t-1,s}\\ \beta_{3,t-1,s}\\ \beta_{1,t-1,b}\\ \beta_{2,t-1,b}\\ \beta_{3,t-1,b}\end{pmatrix} + \begin{pmatrix}\eta_{1,t}\\ \eta_{2,t}\\ \eta_{3,t}\\ \eta_{4,t}\\ \eta_{5,t}\\ \eta_{6,t}\end{pmatrix} \tag{16a}$$

$$\begin{pmatrix}y_{t,s}(\tau_1)\\ y_{t,s}(\tau_2)\\ \vdots\\ y_{t,s}(\tau_{n_s})\\ y_{t,b}(\tau_1)\\ y_{t,b}(\tau_2)\\ \vdots\\ y_{t,b}(\tau_{n_b})\end{pmatrix} = \begin{pmatrix}1 & \frac{1-e^{-\tau_1\lambda_s}}{\tau_1\lambda_s} & \frac{1-e^{-\tau_1\lambda_s}}{\tau_1\lambda_s}-e^{-\tau_1\lambda_s} & 0 & 0 & 0\\ 1 & \frac{1-e^{-\tau_2\lambda_s}}{\tau_2\lambda_s} & \frac{1-e^{-\tau_2\lambda_s}}{\tau_2\lambda_s}-e^{-\tau_2\lambda_s} & 0 & 0 & 0\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ 1 & \frac{1-e^{-\tau_n\lambda_s}}{\tau_n\lambda_s} & \frac{1-e^{-\tau_n\lambda_s}}{\tau_n\lambda_s}-e^{-\tau_n\lambda_s} & 0 & 0 & 0\\ 0 & 0 & 0 & 1 & \frac{1-e^{-\tau_1\lambda_b}}{\tau_1\lambda_b} & \frac{1-e^{-\tau_1\lambda_b}}{\tau_1\lambda_b}-e^{-\tau_1\lambda_b}\\ 0 & 0 & 0 & 1 & \frac{1-e^{-\tau_2\lambda_b}}{\tau_2\lambda_b} & \frac{1-e^{-\tau_2\lambda_b}}{\tau_2\lambda_b}-e^{-\tau_2\lambda_b}\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ 0 & 0 & 0 & 1 & \frac{1-e^{-\tau_n\lambda_b}}{\tau_n\lambda_b} & \frac{1-e^{-\tau_n\lambda_b}}{\tau_n\lambda_b}-e^{-\tau_n\lambda_b}\end{pmatrix} \begin{pmatrix}\beta_{1,t,s}\\ \beta_{2,t,s}\\ \beta_{3,t,s}\\ \beta_{1,t,b}\\ \beta_{2,t,b}\\ \beta_{3,t,b}\end{pmatrix} + \begin{pmatrix}\varepsilon_{1,t}\\ \varepsilon_{2,t}\\ \vdots\\ \varepsilon_{n_s,t}\\ \varepsilon_{n_s+1,t}\\ \varepsilon_{n_s+2,t}\\ \vdots\\ \varepsilon_{n_s+n_b,t}\end{pmatrix}, \tag{16b}$$

where the subscripts s and b denote the elements related the swap and bond models respectively. This model has the initial conditions

$$\eta_t \overset{iid}{\sim} \mathcal{N}(0, Q_t) \tag{16c}$$

$$\varepsilon_t \overset{iid}{\sim} \mathcal{N}(0, H_t) \tag{16d}$$

$$\begin{pmatrix}\beta_{1,t,s}\\ \beta_{2,t,s}\\ \beta_{3,t,s}\\ \beta_{1,t,b}\\ \beta_{2,t,b}\\ \beta_{3,t,b}\end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix}\bar{b}_{0,s}\\ \bar{b}_{0,b}\end{pmatrix}, \begin{pmatrix}P_{0,s} & 0\\ 0 & P_{0,b}\end{pmatrix}\right), \tag{16e}$$

with Q_t a 6×6 diagonal matrix, H_t a $(n_s+n_b)\times(n_s+n_b)$ diagonal matrix, $\bar{b}_{0,s}$ and $\bar{b}_{0,s}$ two 3×1 vectors and $P_{0,s}$ and $P_{0,s}$ two 3×3 diagonal matrices.

E.3 DNS variation with the

The DNS model that models the factors of both the swap and bond curves in one VAR model (Equation 5) is

$$\begin{pmatrix} \beta_{1,t,s} \\ \beta_{2,t,s} \\ \beta_{3,t,s} \\ \beta_{1,t,b} \\ \beta_{2,t,b} \\ \beta_{3,t,b} \\ \beta_{1,t-1,s} \\ \beta_{2,t-1,s} \\ \beta_{3,t-1,s} \\ \beta_{1,t-1,b} \\ \beta_{2,t-1,b} \\ \beta_{3,t-1,b} \end{pmatrix} = \begin{pmatrix} c_{1,s} \\ c_{2,s} \\ c_{3,s} \\ c_{1,b} \\ c_{2,b} \\ c_{3,b} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \overline{\Phi} + I & -\overline{\Phi} \\ I & \emptyset \end{pmatrix} \begin{pmatrix} \beta_{1,t-1,s} \\ \beta_{2,t-1,s} \\ \beta_{3,t-1,s} \\ \beta_{1,t-1,b} \\ \beta_{2,t-1,b} \\ \beta_{3,t-1,b} \\ \beta_{1,t-2,s} \\ \beta_{2,t-2,s} \\ \beta_{3,t-2,s} \\ \beta_{1,t-2,b} \\ \beta_{2,t-2,b} \\ \beta_{3,t-2,b} \end{pmatrix} + \begin{pmatrix} \eta_{1,t} \\ \eta_{2,t} \\ \eta_{3,t} \\ \eta_{4,t} \\ \eta_{5,t} \\ \eta_{6,t} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \tag{17a}$$

$$\begin{pmatrix} y_{t,s}(\tau_1) \\ y_{t,s}(\tau_2) \\ \vdots \\ y_{t,s}(\tau_{n_s}) \\ y_{t,b}(\tau_1) \\ y_{t,b}(\tau_2) \\ \vdots \\ y_{t,b}(\tau_{n_b}) \end{pmatrix} = \begin{pmatrix} \overline{\Lambda}_s & O & O & O \\ O & O & O & \overline{\Lambda}_b \end{pmatrix} \begin{pmatrix} \beta_{1,t,s} \\ \beta_{2,t,s} \\ \beta_{3,t,s} \\ \beta_{1,t,b} \\ \beta_{2,t,b} \\ \beta_{3,t,b} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \vdots \\ \varepsilon_{n_s,t} \\ \varepsilon_{n_s+1,t} \\ \varepsilon_{n_s+2,t} \\ \vdots \\ \varepsilon_{n_s+n_b,t} \end{pmatrix}, \tag{17b}$$

$$\overline{\Phi} := \begin{pmatrix} \phi_{11} & & & \phi_{14} & & \\ & \phi_{22} & & & \phi_{25} & \\ & & \phi_{33} & & & \phi_{36} \\ \phi_{41} & & & \phi_{44} & & \\ & \phi_{52} & & & \phi_{55} & \\ & & \phi_{63} & & & \phi_{66} \end{pmatrix}$$

$$\overline{\Lambda}_s := \begin{pmatrix} 1 & \frac{1-e^{-\tau_1\lambda_s}}{\tau_1\lambda_s} & \frac{1-e^{-\tau_1\lambda_s}}{\tau_1\lambda_s} - e^{-\tau_1\lambda_s} \\ 1 & \frac{1-e^{-\tau_2\lambda_s}}{\tau_2\lambda_s} & \frac{1-e^{-\tau_2\lambda_s}}{\tau_2\lambda_s} - e^{-\tau_2\lambda_s} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\tau_n\lambda_s}}{\tau_n\lambda_s} & \frac{1-e^{-\tau_n\lambda_s}}{\tau_n\lambda_s} - e^{-\tau_n\lambda_s} \end{pmatrix} \tag{17c}$$

$$\overline{\Lambda}_b := \begin{pmatrix} 1 & \frac{1-e^{-\tau_1\lambda_b}}{\tau_1\lambda_b} & \frac{1-e^{-\tau_1\lambda_b}}{\tau_1\lambda_b} - e^{-\tau_1\lambda_b} & \\ 1 & \frac{1-e^{-\tau_2\lambda_b}}{\tau_2\lambda_b} & \frac{1-e^{-\tau_2\lambda_b}}{\tau_2\lambda_b} - e^{-\tau_2\lambda_b} & \\ & \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\tau_n\lambda_b}}{\tau_n\lambda_b} & \frac{1-e^{-\tau_n\lambda_b}}{\tau_n\lambda_b} - e^{-\tau_n\lambda_b} & \end{pmatrix},$$

where O is a 30×3-zero-matrix.

F Kalman filter

F.1 Background theory

In this section of the appendix, it is explained how the Kalman filter obtains MSE optimal estimates and predictions for the state vectors and the observations in a state space model. The theory in this section originates from Durbin and Koopman (2012a). The full derivation of the Kalman filter can also be found in Durbin and Koopman (2012a).

Consider the state space system in Equation 8a-8e, repeated below

$$\begin{aligned} b_{t+1} &= c + \Phi b_t + \eta_t, & \eta_t &\sim \mathcal{N}(0, Q) \\ y_t &= \Lambda_t b_t + \varepsilon_t, & \varepsilon_t &\sim \mathcal{N}(0, H_t) \\ & & b_0 &\sim \mathcal{N}(\overline{b}_0, P_0) \end{aligned}$$

Furthermore, define $Y_t = (y_1', \ldots, y_t')'$, the observations obtained until time t. The Kalman filter aims to find a series of MSE optimal estimators and MSE optimal predictors for the state vectors b_t and the observations y_t in this state space system. The MSE optimal estimators for the state vectors b_t and b_{t+1} are given by the conditional expectation conditioned on Y_t,

$$\begin{aligned} \hat{b}_{t|t} &= \mathbb{E}[b_t|Y_t] \\ \hat{b}_{t+1|t} &= \mathbb{E}[b_{t+1}|Y_t], \end{aligned}$$

where $\hat{b}_{t|t}$ is the MSE optimal estimate and $\hat{b}_{t+1|t}$ is the MSE optimal predictor for the state vector. Using this notation, the Kalman filter produces the series of estimates $\hat{b}_{t|t}$ and $\hat{b}_{t+1|t}$ for these state vectors for $t = 1, \ldots, T$.
In Durbin and Koopman (2012a), it is proved that the distributions of b_t and b_{t+1} conditioned on Y_t are multivariate normal

$$\begin{aligned} b_t|Y_t &\sim \mathcal{N}(\overline{b}_{t|t}, P_{t|t}) \\ b_{t+1}|Y_t &\sim \mathcal{N}(\overline{b}_{t+1|t}, P_{t|t}) \end{aligned}$$

and due to these multivariate normality, the MSE optimal estimates $\hat{b}_{t|t}$ and $\hat{b}_{t+1|t}$ are equal to the mean vectors of the conditional distributions, i.e.

$$\begin{aligned} \hat{b}_{t|t} &= \mathbb{E}[b_t|Y_t] = \overline{b}_{t|t} \\ \hat{b}_{t+1|t} &= \mathbb{E}[b_{t+1}|Y_t] = \overline{b}_{t+1|t}. \end{aligned}$$

The Kalman filter obtains the estimates $\overline{b}_{t|t}$ and $\overline{b}_{t+1|t}$ in step 1(a) and step 2(c) respectively.

For the MSE optimal predictors for the observations y_t, the same reasoning is applied (Durbin & Koopman, 2012a). The MSE optimal estimators for the observations are given by

$$\hat{y}_{t+1|t} = \mathbb{E}[y_{t+1}|Y_t].$$

The state space system implies that the observations are normally distributed conditional on Y_t,

$$y_{t+1}|Y_t \sim \mathcal{N}(\Lambda_t \overline{b}_{t|t-1}, F_t),$$

which implies

$$\hat{y}_{t+1|t} = \mathbb{E}[y_{t+1}|Y_t] = \Lambda_t \overline{b}_{t|t-1}.$$

The MSE optimal prediction for y_t is obtained in step 1(c) of the Kalman filter.

F.2 Derivation of the log likelihood function

In this section we derive the log likelihood function of the state space model in Equation 8a-8e as a function of the model parameters $\Theta := \{\overline{b}_0, P_0, \Phi, c, Q_t, \Lambda_t, H_t\}$.

Suppose we have a sequence of observations $y_1, y_2, \ldots, y_T$. The joint density of these observations satisfy

$$p(y_1, y_2, \ldots, y_T) = \prod_{t=1}^{T} p(y_t|Y_{t-1}), \tag{18a}$$

where $p\left(y_t|Y_{t-1}\right)$ is the density of y_t conditioned on all previous observations $Y_{t-1} = \{y_1, y_2, \ldots, y_{t-1}\}$. The here we use $p(y_1|Y_0) = p(y_1)$. The joint density for y_t, $t = 1, \ldots, T$ depends on parameter Θ. The likelihood for Θ given the observations $y_1, y_2, \ldots, y_T$ is given by

$$L\left(\Theta\right) = p\left(y_1, y_2, \ldots, y_T; \Theta\right) = \prod_{t=1}^{T} p\left(y_t|Y_{t-1}; \Theta\right). \tag{18b}$$

The log likelihood is

$$\ell\left(\Theta\right) := \log L\left(\Theta\right) = \sum_{t=1}^{T} \log p\left(y_t|Y_{t-1}; \Theta\right), \tag{18c}$$

where in the last two equations we used $p(y_1|Y_0) = p(y_1)$. In **??** we have shown that the y_t is multivariate normal distributed given Y_{t-1} with mean $\overline{y}_t$ and covariance matrix F_t. The probability density function is given by

$$\begin{aligned} p\left(y_t|Y_{t-1}\right) &= \left(\left(2\pi\right)^{\frac{n_t}{2}} \sqrt{|F_t|}\right)^{-1} \exp\left(-\frac{1}{2}\left(y_t - \overline{y}_t\right)' F_t^{-1}\left(y_t - \overline{y}_t\right)\right) \\ &= \left(\left(2\pi\right)^{\frac{n_t}{2}} \sqrt{|F_t|}\right)^{-1} \exp\left(-\frac{1}{2}\left(v_t\right)' F_t^{-1}\left(v_t\right)\right) \end{aligned} \tag{18d}$$

where n_t is the number of elements in the observation vector y_t, and $v_t := y_t - \overline{y}_t$ is the prediction error. From the state space system we know that the prediction error v_t and its covariance matrix F_t depend on the parameters in the system Θ. Therefore we write $v_t\left(\Theta\right)$ and $F_t\left(\Theta\right)$. Filling in Equation 18d into Equation 18c gives

$$\begin{aligned} \ell\left(\Theta\right) &= \sum_{t=1}^{T} \log\left(\left(\left(2\pi\right)^{\frac{n_t}{2}} \sqrt{|F_t\left(\Theta\right)|}\right)^{-1} \exp\left(-\frac{1}{2}\left(v_t\left(\Theta\right)\right)' F_t\left(\Theta\right)^{-1}\left(v_t\left(\Theta\right)\right)\right)\right) \\ &\downarrow \text{ Take the natural logarithm on both sides and assume that } n_t = N \text{ for all } t \\ \ln \ell\left(\Theta\right) &= -\frac{NT}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\left(\ln|F_t\left(\Theta\right)| + \left(v_t\left(\Theta\right)\right)' F_t\left(\Theta\right)^{-1}\left(v_t\left(\Theta\right)\right)\right) \end{aligned} \tag{18e}$$

F.3 Missing observations

A main advantage of estimating the factors b_t in a state space model with the Kalman filter is its allowance for missing observations. In this subsection we explain how the estimates of the factors and their covariance matrices are constructed when there are missing observations.

Suppose the Kalman Filter given in Algorithm 3.1 has obtained estimates for $b_1, \ldots, b_{j-1}$ using the observations $y_t, \ldots, y_{j-1}$. If the whole observation vector y_j is missing, we can obtain the predictions for $\overline{b}_{j|j-1}$ and $P_{j|j-1}$ by the general Kalman filter equations. The update to $\overline{b}_{j|j}$ and $P_{j|j}$ and the prediction to $\overline{b}_{j+1|j}$ and $P_{j+1|j}$ can not be obtained by the Kalman filter equations without observation y_j. Instead, we use the following recursive equations,

$$\overline{b}_{j|j} = \mathbb{E}\left[b_j|Y_j\right] = \mathbb{E}\left[b_j|Y_{j-1}\right] = \overline{b}_{j|j-1} \tag{19a}$$

$$P_{j|j} = \mathrm{Var}\left[b_j|Y_j\right] = \mathrm{Var}\left[b_j|Y_{t-j}\right] = P_{j|j-1} \tag{19b}$$

$$\overline{b}_{j+1|j} = \mathbb{E}\left[b_{j+1}|Y_j\right] = \mathbb{E}\left[\Phi_j b_t j + \eta_j\right] = \Phi_j b_{j|j-1} \tag{19c}$$

$$P_{j+1|t} = \mathrm{Var}\left[b_{j+1}|Y_t\right] = \mathrm{Var}\left[\Phi_j b_j + \eta_t|Y_{j-1}\right] = \Phi_t P_{j|j-1}\Phi_j' + Q_j. \tag{19d}$$

These equations hold are used when more than one consecutive observation is missing. The algebraïc proof of these equation is given in the book of Durbin and Koopman (Durbin & Koopman, 2012b).

F.4 Predicting the observations with the Kalman Filter

For the prediction of the factors, we treat the future observation as missing observations. That is, we observe $y_t, \ldots, y_T$ and have missing observations $y_{T+1}, \ldots, y_{T+k}$, for $k = 1, 2, \ldots$. The prediction of the factors $b_{T+1}, \ldots, b_{T+k}$ and their Covariance matrices $P_{T+1}, \ldots, P_{T+k}$ are obtained by equation Equation 19a-19d.

In this thesis, we are not so much interested in the predicted factors, but mainly in the predictions for the observations $y_{T+1}, \ldots, y_{T+k}$. In **??** we showed that the distribution of y_t given Y_{t-1} is multivariate normal with mean $\overline{y}_t$ and covariance matrix F_t. Therefore, the minimum mean square error forecast of y_{T+k} given $Y_T = \{y_1, \ldots, y_T\}$ is the conditional mean $\overline{y}_{T+k} = \mathbb{E}\left[y_{T+k}|Y_T\right]$.

The one step ahead prediction for the observation y_{T+1} and its covariance matrix F_{T+1} given Y_T are obtained by

$$\overline{y}_{T+1} = \Lambda_{T+1}\mathbb{E}\left[b_{T+1}|Y_T\right] = \Lambda_{T+1}\bar{b}_{T+1|T} \tag{20a}$$

$$F_{T+1} = \mathbb{E}\left[\left(\overline{y}_{T+1} - y_{T+1}\right)\left(\overline{y}_{T+1} - y_{T+1}\right)' |Y_T\right] = \Lambda_{T+1}P_{T+1}\Lambda_{T+1} + H_{T+1} \tag{20b}$$

The k step ahead prediction is given by

$$\overline{y}_{T+k} = \Lambda_{T+k}\mathbb{E}\left[b_{T+k}|Y_T\right] = \Lambda_{T+k}\bar{b}_{T+k|T+k-1}, \tag{20c}$$

where $\bar{b}_{T+k|T+k-1}$ is obtained recursively by Equation 19a and 19c. The corresponding covariance matrix is obtained by

$$F_{T+k} = \Lambda_{T+k}P_{T+k|T+k-1}\Lambda_{T+k} + H_{T+k}, \tag{20d}$$

where $P_{T+k|T+k-1}$ is obtained recursively from Equation 19b and 19d.

Equation 20a-20d give the recursive equations from which we construct the predictions for the observations and their corresponding covariance matrices.

F.5 The 2-step approach to estimate the parameters and factors in the DNS model

F.5.1 The 2-step approach

In this part of the appendix, it is explained how to estimate the parameters for the two DNS models using the 2-step approach. 2-step stands for estimating the factors in the first step, and then estimating the time series model parameters in the second step.

Estimating the factors starts with choosing a fixed value for λ. With this λ, the factors $\beta_{1,t}$, $\beta_{2,t}$, and $\beta_{3,t}$ in Equation 2 can be estimated for each day t using ordinary least squared optimization. This results in a series of estimates for each of the factors. This part of the estimation procedure is the so-called first step.
For the nonstatDNS and the sim-nonstatDNS models, the three series of factors are transformed into a stationary time series using a first-order difference transformation.

The second step in the procedure is to estimate the parameters of the time series model. The parameters for the time series model are the autoregressive coefficients $\hat{\phi}_{ii}$ and the unconditional mean vector c. The parameters in the time series model are also estimated by Maximum likelihood estimation. The specific procedure on how to estimate the parameters of an autoregressive model by MLE is specified in Shumway and Stoffer (2016a).

F.5.2 The usage of the 2-step approach as initial parameters for the 1-step approach

The 2-step approach explained in the previous section of the appendix is used to obtain the initial parameters for the MLE estimation of the DNS model parameters in the state space representation. In this part of the appendix, it is specified how to use the estimated parameters as initial parameters.

The value for λ that was used in the 2-step approach serves as the initial λ for the state space model. Also, the estimated autoregressive coefficients and unconditional mean can be used directly as initial parameters for the state space model. The initial values for the factors in $\bar{b}_0$ are set to the initial values that were estimated in the first step of the 2-step approach. The initial covariance P_0 matrix is a diagonal matrix with on the diagonal the standard deviations of the estimated factors. The residuals of the OLS estimation in the first step and the residuals of the autoregressive models are used to obtain the initial values for the residual distribution Q and H.

In the case of missing weekend observations, the missing values are deleted from the data. That means it is assumed that Monday directly follows on Friday and there are no days without observations. Notice that this assumption is only used to obtain initial parameters. In the MLE optimization procedure with the Kalman filter, missing observations are treated as missing observations.

G Simulation studies

G.1 Parameters for the simulation studies

G.1.1 Parameters for the statDNS model

$$
\begin{aligned}
b_0 &= \begin{pmatrix} 1 & -2 & 3.5 \end{pmatrix}^T \\
P_0 &= \operatorname{diag}\begin{pmatrix} sqrt0.02 & sqrt0.02 & sqrt0.02 \end{pmatrix} \\
c &= \begin{pmatrix} 1.5 & -2 & 1 \end{pmatrix}^T \\
\Phi &= \begin{pmatrix} -0.4 & & \\ & -0.5 & \\ & & -0.25 \end{pmatrix} \\
\lambda &= 0.1195 \\
Q &= \operatorname{diag}\begin{pmatrix} \sqrt{0.004} & \sqrt{0.004} & \sqrt{0.002} \end{pmatrix}^T \\
H &= \operatorname{diag}\begin{pmatrix} \sqrt{0.0025} & \dots & \sqrt{0.0025} \end{pmatrix}
\end{aligned}
$$

G.1.2 Parameters for the sim-statDNS model

$$
\begin{aligned}
b_0 &= \begin{pmatrix} 1.2 & -2 & 0.4 & 1.2 & -2 & 0.4 \end{pmatrix}^T \\
P_0 &= \operatorname{diag}\begin{pmatrix} \sqrt{0.02} & \dots & \sqrt{0.02} \end{pmatrix} \\
c &= \begin{pmatrix} 0.3 & -1.6 & 1.4 & 0.9 & -1.2 & 1 \end{pmatrix}^T \\
\Phi &= \begin{pmatrix} -0.3 & & & 0.5 & & \\ & -0.4 & & & 0.2 & \\ & & 0.6 & & & 0.4 \\ -0.4 & & & 0.5 & & \\ & -0.4 & & & 0.7 & \\ & & 0.2 & & & 0.4 \end{pmatrix} \\
\lambda_{swap} &= 0.1195 \\
\lambda_{bond} &= 0.24 \\
Q &= \operatorname{diag}\begin{pmatrix} \sqrt{0.04} & \sqrt{0.04} & \sqrt{0.04} & \sqrt{0.04} & \sqrt{0.04} & \sqrt{0.04} \end{pmatrix}^T \\
H &= \operatorname{diag}\begin{pmatrix} \sqrt{0.05} & \dots & \sqrt{0.05} \end{pmatrix}
\end{aligned}
$$

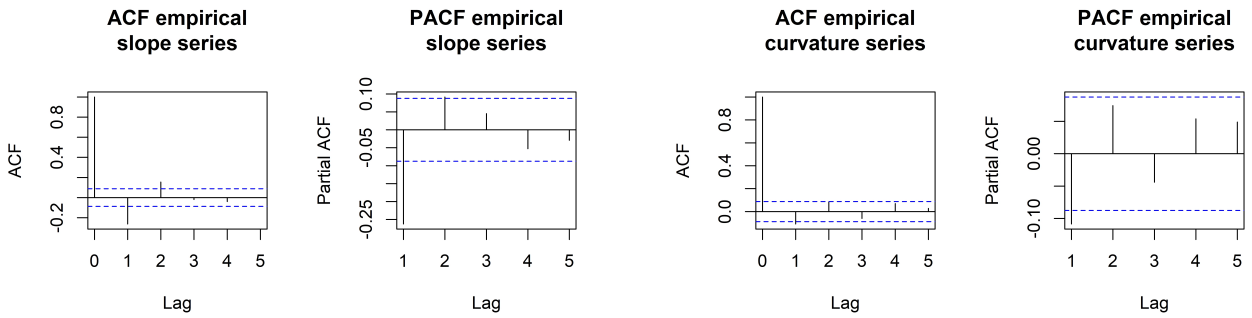
G.1.3 Parameters for the nonstatDNS model

$$
\begin{aligned}
b_0 &= \begin{pmatrix} 1 & -2 & 3.5 & 1 & -2 & 3.5 \end{pmatrix}^T \\
P_0 &= \operatorname{diag}\begin{pmatrix} \sqrt{0.02} & \sqrt{0.02} & \sqrt{0.02} \end{pmatrix} \\
c &= \begin{pmatrix} 0 & 0 & 0 \end{pmatrix}^T \\
\tilde{\Phi} &= \begin{pmatrix} -0.3 & & \\ & -0.5 & \\ & & -0.15 \end{pmatrix} \\
\lambda &= 0.1195 \\
Q &= \operatorname{diag}\begin{pmatrix} \sqrt{0.0015} & \sqrt{0.0015} & \sqrt{0.0015} \end{pmatrix}^T \\
H &= \operatorname{diag}\begin{pmatrix} \sqrt{0.0025} & \dots & \sqrt{0.0025} \end{pmatrix}
\end{aligned}
$$

G.1.4 Parameters for the sim-nonstatDNS model

$$
\begin{aligned}
b_0 &= \begin{pmatrix} 2 & -2.5 & 0.4 & 1.2 & -2 & 0.4 \end{pmatrix}^T \\
P_0 &= \operatorname{diag}\begin{pmatrix} \sqrt{0.02} & \dots & \sqrt{0.02} \end{pmatrix} \\
c &= \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}^T \\
\overline{\Phi} &= \begin{pmatrix} -0.4 & & & 0.5 & & \\ & 0.6 & & & 0.7 & \\ & & -0.4 & & & -0.7 \\ -0.3 & & & -0.5 & & \\ & -0.6 & & & 0.4 & \\ & & 0.3 & & & -0.7 \end{pmatrix} \\
\lambda_{swap} &= 0.1195 \\
\lambda_{bond} &= 0.24 \\
Q &= \operatorname{diag}\begin{pmatrix} \sqrt{0.025} & \dots & \sqrt{0.025} \end{pmatrix}^T \\
H &= \operatorname{diag}\begin{pmatrix} \sqrt{0.005} & \dots & \sqrt{0.005} \end{pmatrix}
\end{aligned}
$$

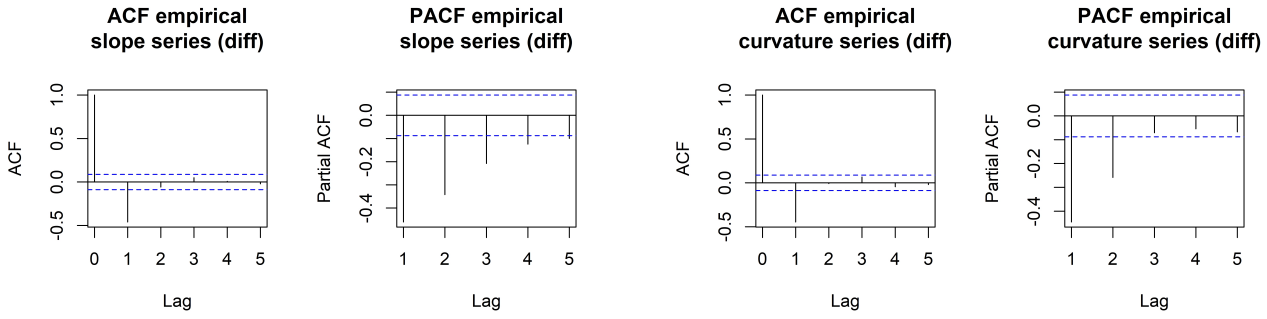
G.2 Empirical counterparts of the factors



(a) Slope factor series

(b) Curvature factor series

Figure 44: The ACF and PACF plot of the series of empirical counterparts for the slope and curvature factors. Results for data simulated form the statDNS model



(a) Slope factor series

(b) Curvature factor series

Figure 45: The ACF and PACF plot of the series of empirical counterparts for the slope and curvature factor after transformation by the first order difference. Results for data simulated form the nonstatDNS model

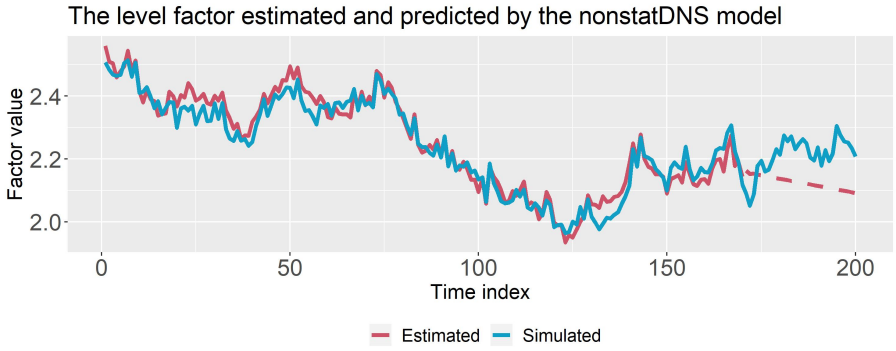
G.3 Initial value for λ

Factor	Range for $\hat{\lambda}_{MLE}$	Q1 (β)	Median (β)	Q2 (β)	Min (β)	Max (β)
level	$\lambda < 0.1189$	0.3177	1.256	1.585	-0.2727	1.991
level	$0.1189 \leq \lambda < 0.1199$	1.117	1.361	1.732	0.1458	2.56
level	$0.1199 \leq \lambda < 0.1216$	0.8591	1.307	1.801	-0.1306	2.486
level	$\lambda \geq 0.1216$	0.4539	0.8272	1.346	-0.1649	1.998
slope	$\lambda < 0.1189$	-2.067	-1.755	-1.643	-2.655	-1.456
slope	$0.1189 \leq \lambda < 0.1199$	-1.965	-1.637	-1.449	-2.76	-0.4707
slope	$0.1199 \leq \lambda < 0.1216$	-2.464	-2.244	-2.003	-2.766	-0.4697
slope	$\lambda \geq 0.1216$	-2.944	-2.531	-1.674	-3.269	-1.264
curvature	$\lambda < 0.1189$	3.357	3.592	4.082	2.841	5.569
curvature	$0.1189 \leq \lambda < 0.1199$	3.651	4.05	4.379	3.131	5.527
curvature	$0.1199 \leq \lambda < 0.1216$	3.37	3.79	4.849	2.696	5.524
curvature	$\lambda \geq 0.1216$	2.987	3.276	3.637	-1.685	4.121

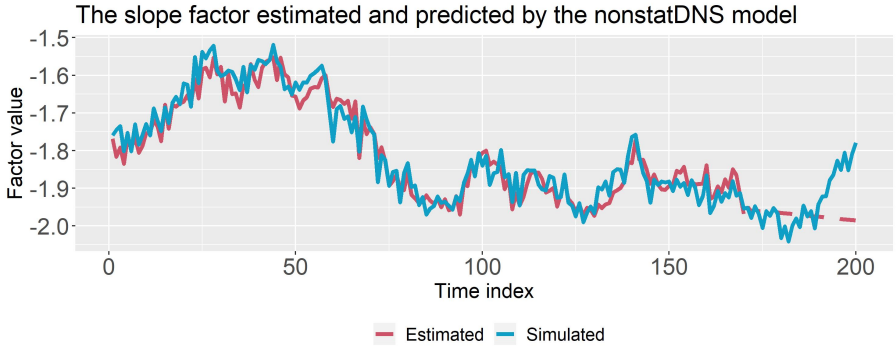
Table 45: The descriptive statistics for the estimated factors

G.4 Validation of the nonstatDNS model

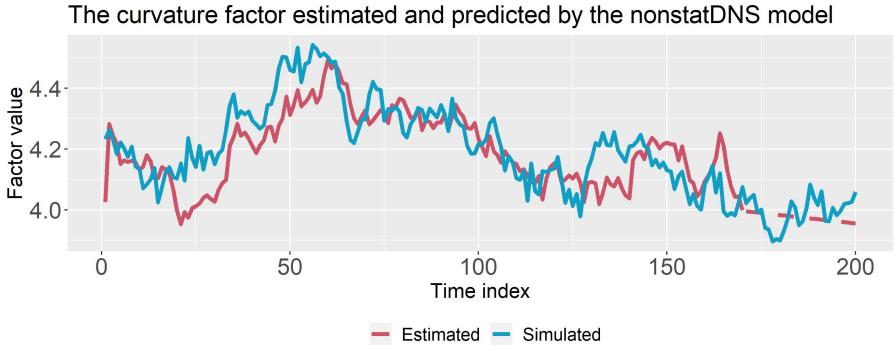
G.4.1 First simulation



(a) The level factor estimated and predicted by the nonstatDNS model

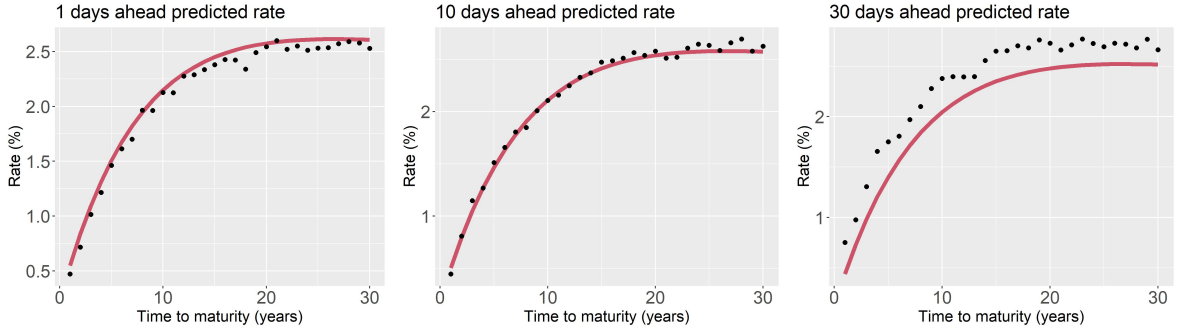


(b) The slope factor estimated and predicted by the nonstatDNS model



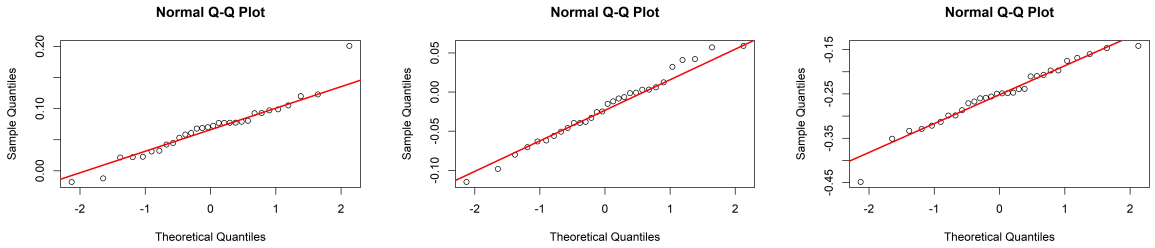
(c) The curvature factor estimated and predicted by the nonstatDNS model

Figure 46: Results of the estimated and predicted factors for one of the 36 simulations for the nonstatDNS model.



(a) 1 day ahead prediction (b) 10 days ahead prediction (c) 30 days ahead prediction

Figure 47: The interest rate curves predicted by the nonstatDNS model.



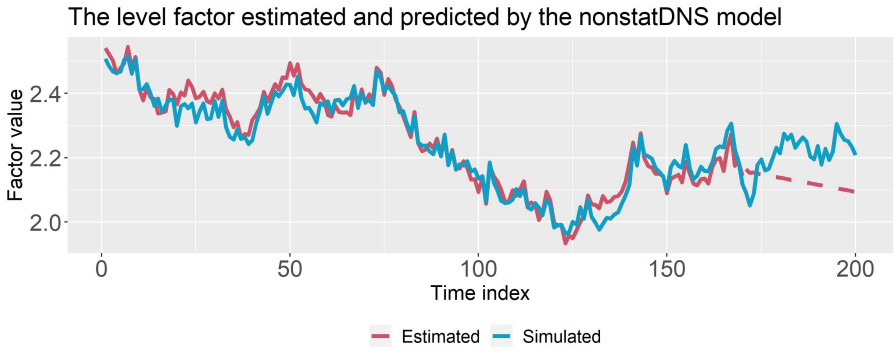
(a) Errors of the 1 day ahead prediction

(b) Errors of the 10 day ahead prediction

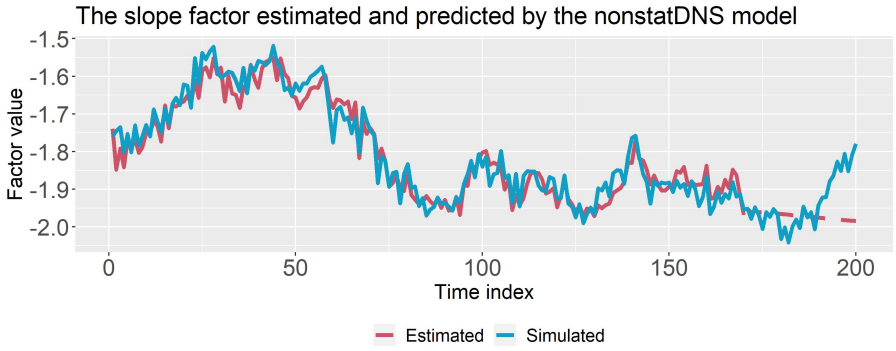
(c) Errors of the 30 day ahead prediction

Figure 48: QQ plots of the error terms corresponding to the predicted rates.

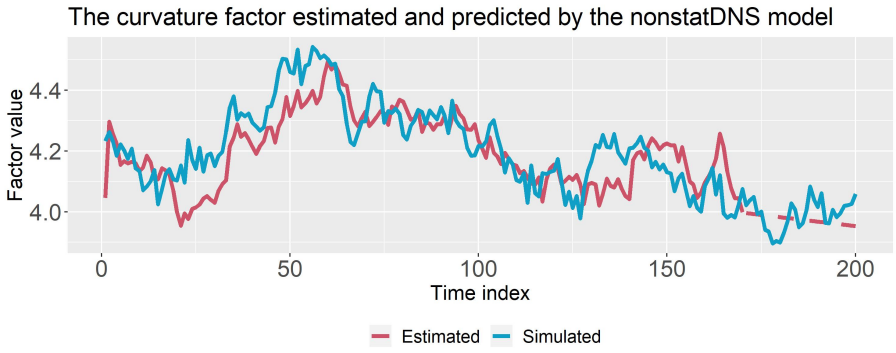
G.4.2 Second simulation



(a) The level factor estimated and predicted by the nonstatDNS model

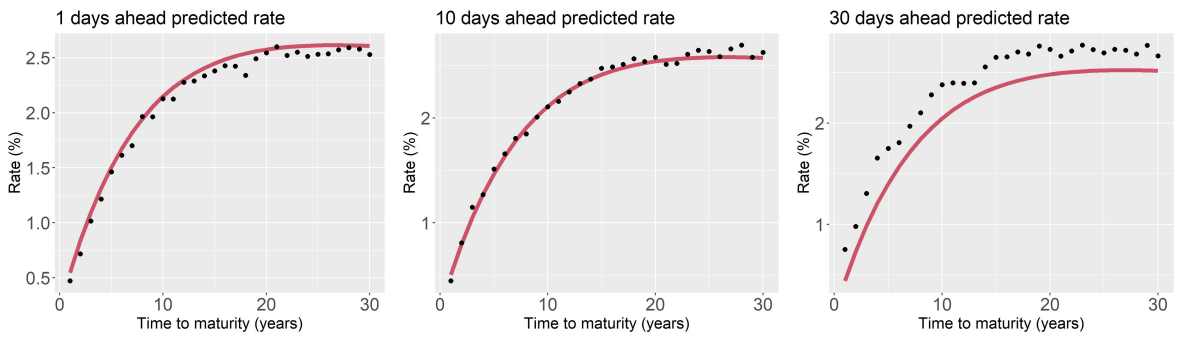


(b) The slope factor estimated and predicted by the nonstatDNS model



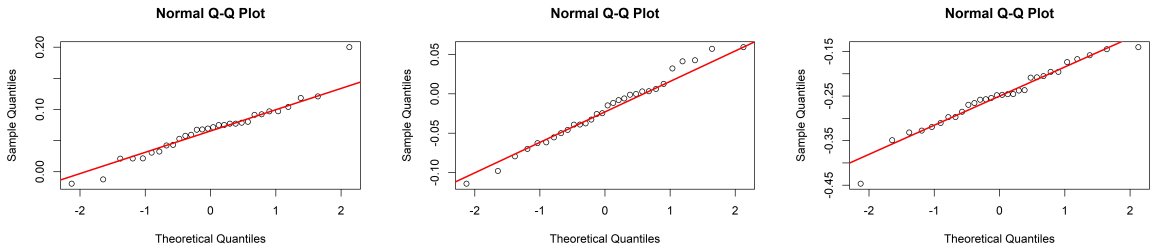
(c) The curvature factor estimated and predicted by the nonstatDNS model

Figure 49: Results of the estimated and predicted factors for one of the 36 simulations for the nonstatDNS model.



(a) 1 day ahead prediction (b) 10 days ahead prediction (c) 30 days ahead prediction

Figure 50: The interest rate curves predicted by the nonstatDNS model.

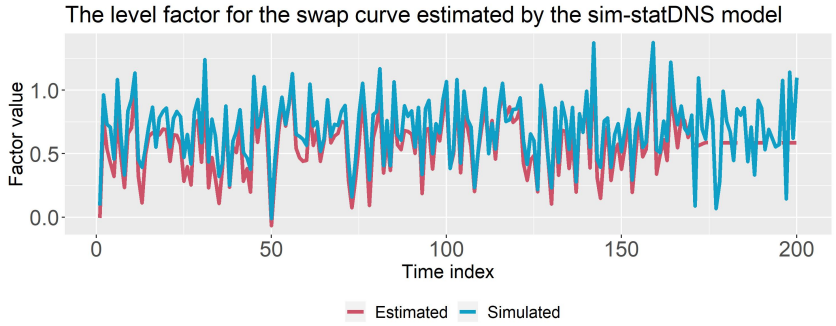


(a) Errors of the 1 day ahead prediction (b) Errors of the 10 day ahead prediction (c) Errors of the 30 day ahead prediction

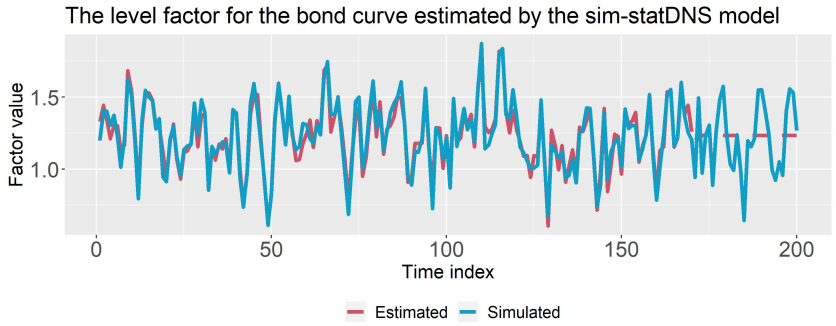
Figure 51: QQ plots of the error terms corresponding to the predicted rates.

G.5 Validation of the sim-statDNS model

G.5.1 First simulation



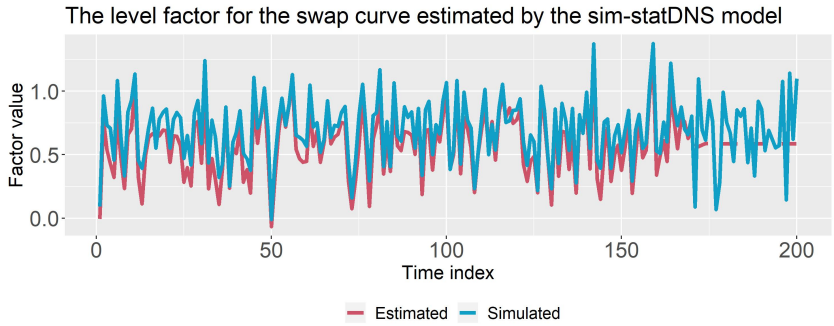
(a) The level factor estimated and predicted by the nonstatDNS model



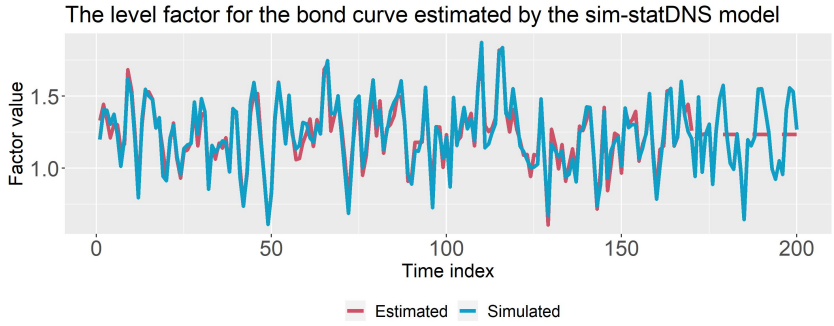
(b) The slope factor estimated and predicted by the nonstatDNS model

Figure 52: Results of the estimated and predicted factors for one of the 36 simulations for the nonstatDNS model.

G.5.2 Second simulation



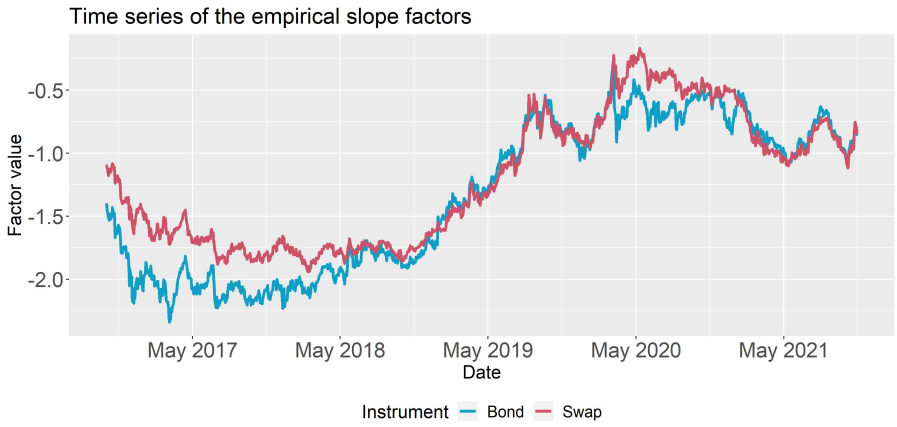
(a) The level factor estimated and predicted by the nonstatDNS model



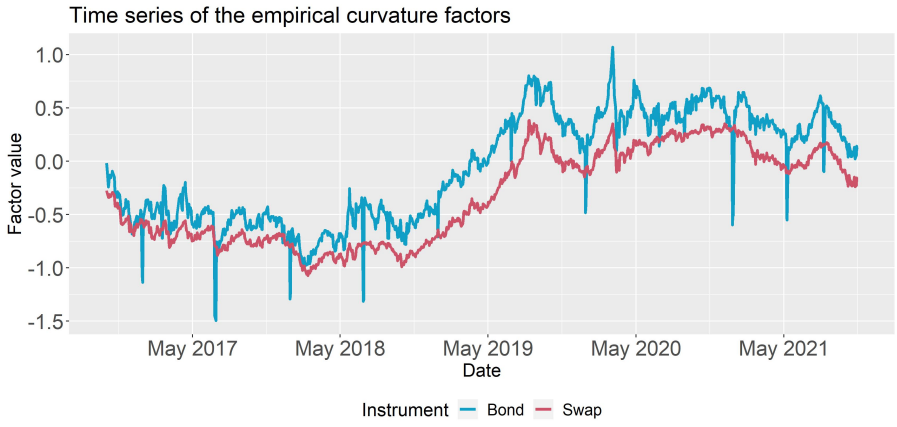
(b) The slope factor estimated and predicted by the nonstatDNS model

Figure 53: Results of the estimated and predicted factors for one of the 36 simulations for the nonstatDNS model.

H Case study results

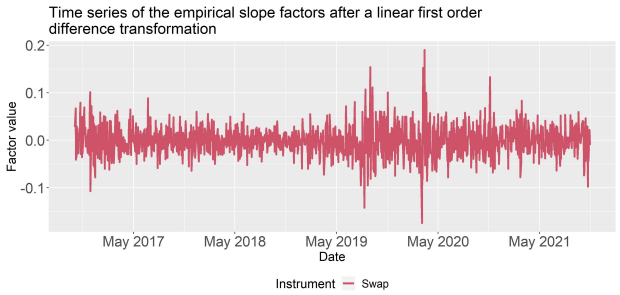


(a) Time series of the empirical slope factors computed from the real data

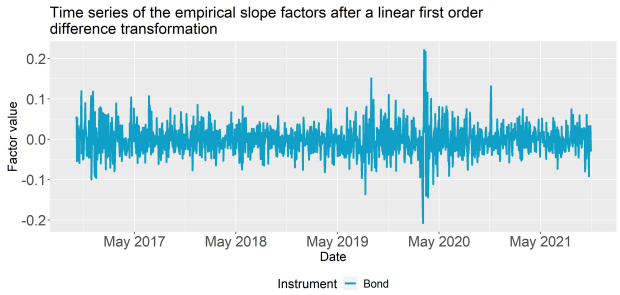


(b) Time series of the empirical curvature factors computed from the real data

Figure 54: Empirical factors for the case study

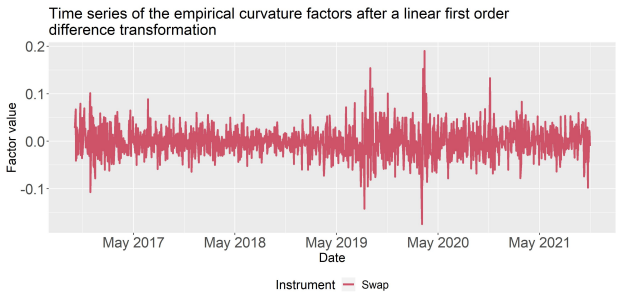


(a) Time series of the transformed empirical slope factors computed from the real swap data

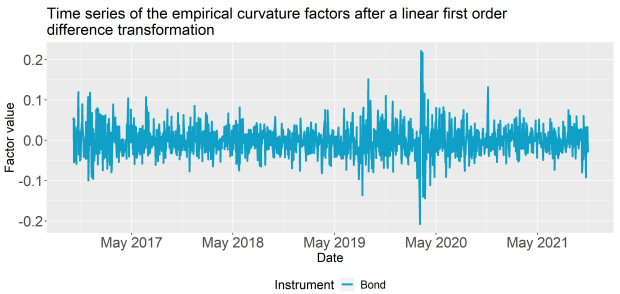


(b) Time series of the transformed empirical slope factors computed from the real bond data

Figure 55: Transformed empirical factors for the case study



(a) Time series of the transformed empirical curvature factors computed from the real swap data



(b) Time series of the transformed empirical curvature factors computed from the real bond data

Figure 56: Transformed empirical factors for the case study

H.1 Estimated parameters

H.1.1 MLE parameters of the nonstatDNS model

MLE parameters for swap data

$$\hat{b}_0 = \begin{pmatrix} 1.52 & -1.83 & -1.83 & 1.52 & -1.83 & -1.83 \end{pmatrix}^T$$
$$\hat{P}_0 = diag\begin{pmatrix} 0.54 & 0.60 & 0.23 & 0.54 & 0.60 & 0.23 \end{pmatrix}$$
$$\hat{c} = \begin{pmatrix} 0.006 & 1.02 & -0.007 & 0 & 0 & 0 \end{pmatrix}^T$$
$$\hat{\tilde{\Phi}} = diag\begin{pmatrix} 0.074 & 0.043 & 0.083 \end{pmatrix}$$
$$\hat{\lambda}_{swap} = 0.3462123$$
$$\hat{Q} = diag\begin{pmatrix} 282.85 & 74.69 & 1.64 & 0 & 0 & 0 \end{pmatrix}^T$$
$$\hat{H} = diag\begin{pmatrix} 0.0571 & 1.2 & 0.00109 & 1.41 \times 10^{-06} & 0.000198 & 0.000287 & 0.313 \\ 2.76 & 1.63 & 2.07 \times 10^{-05} & 4.2 \times 10^{-08} & 9.87 \times 10^{-06} & 1.68 \times 10^{-05} & 1.02 \times 10^{-05} \\ 4.12 \times 10^{-08} & 3.15 \times 10^{-05} & 0.000161 & 0.615 & 4.56 & 23.7 & 31.7 \\ 0.0175 & 0.0197 & 0.0133 & 0.215 & 0.848 & 1.67 & 3.43 \\ 5.46 & 5.63 & & & & & \end{pmatrix}^T$$

MLE parameters for bond data

$$\hat{b}_0 = \begin{pmatrix} 1.49 & -2.17 & -2.56 & 1.49 & -2.17 & -2.56 \end{pmatrix}^T$$
$$\hat{P}_0 = diag\begin{pmatrix} 0.587 & 0.749 & 0.233 & 0.587 & 0.749 & 0.233 \end{pmatrix}$$
$$\hat{c} = \begin{pmatrix} 0.0248 & 0.235 & 0.114 & 0 & 0 & 0 \end{pmatrix}^T$$
$$\hat{\tilde{\Phi}} = diag\begin{pmatrix} 0.1034 & 0.027 & 0.638 \end{pmatrix}$$
$$\hat{\lambda}_{bond} = 0.009870768$$
$$\hat{Q} = diag\begin{pmatrix} 1289.932 & 201.6421 & 15.20545 & 0 & 0 & 0 \end{pmatrix}$$
$$\hat{H} = diag\begin{pmatrix} 35.3 & 5.29 & 0.184 & 7.7 & 0.036 & 53.8 & 38.6 \\ 44.5 & 20.3 & 0.0162 & 0.00127 & 0.0189 & 0.000992 & 0.00156 \\ 0.00598 & 27.5 & 153 & 862 & 203 & 83.5 & 53.1 \\ 0.0683 & 0.000515 & 0.239 & 0.151 & 15.7 & 185 & \\ 882 & 42.5 & 3.55 & & & & \end{pmatrix}^T$$

H.1.2 Parameters of the sim-nonstatDNS model

$$\hat{b}_0 = \begin{pmatrix} 1.52 & -1.83 & -1.89 & 1.49 & -2.17 & -2.58 & 1.52 & -1.83 & -1.89 & 1.49 \\ -2.17 & -2.58 & & & & & & & & \end{pmatrix}^T$$

$$\hat{P}_0 = diag \begin{pmatrix} 0.537 & 0.602 & 0.369 & 0.527 & 0.586 & 0.215 & 0.537 & 0.602 & 0.369 & 0.527 \\ 0.586 & 0.215 & & & & & & & & \end{pmatrix}$$

$$\hat{c} = \begin{pmatrix} 0.004920 & -0.000574 & 0.000771 & 0.000121 & -0.000334 & -0.000393 & 0 & 0 & 0 & 0 \\ 0 & 0 & & & & & & & & \end{pmatrix}$$

$$\hat{\tilde{\Phi}} = \begin{pmatrix} 0.00135 & & & 0.044 & & \\ & 0.0862 & & & -0.369 & \\ & & 0.290 & & & -0.504 \\ -0.00432 & & & -0.100 & & \\ & -0.00089 & & & 0.649 & \\ & & 0.364 & & & 0.320 \end{pmatrix}$$

$$\hat{\lambda}_{swap} = 0.4410770$$

$$\hat{\lambda}_{bond} = 0.3474915$$

$$\hat{Q} = diag \begin{pmatrix} 4.53 \times 10^{1} & 4.82 \times 10^{-4} & 9.93 \times 10^{-4} \\ 2.08 \times 10^{-7} & 1.60 \times 10^{-4} & 4.53 \times 10^{-4} \\ 4.53 \times 10^{1} & 4.82 \times 10^{-4} & 9.93 \times 10^{-4} \\ 2.08 \times 10^{-7} & 1.60 \times 10^{-4} & 4.53 \times 10^{-4} \end{pmatrix}$$

$$\hat{H} = diag \begin{pmatrix} -0.00524 & -0.000592 & 0.00551 & 0.00533 & -0.0062 \\ 0.00144 & -0.0011 & -0.00336 & 0.0191 & 0.00177 \\ 0.0027 & 0.00291 & 0.004 & -0.00252 & 0.0126 \\ -0.00194 & 0.00464 & 0.00246 & 0.00675 & -0.0119 \\ -0.011 & 0.00151 & 0.00165 & 0.0128 & -0.000319 \\ -0.000459 & 0.00634 & -0.000538 & 0.00418 & -0.00219 \\ -0.027 & 0.00417 & 2.26 \times 10^{-5} & -0.0092 & 0.00387 \\ 0.0063 & 0.00462 & -0.00161 & 0.00856 & 0.0131 \\ 0.00562 & 0.00209 & 0.0111 & 0.0213 & 0.00707 \\ -0.000984 & 0.00429 & 0.00626 & 0.0191 & 0.00346 \\ 0.0193 & 0.00574 & -0.019 & 0.018 & 0.0023 \\ 0.0169 & 0.0132 & 0.00653 & 0.000139 & 0.0274 \end{pmatrix}$$

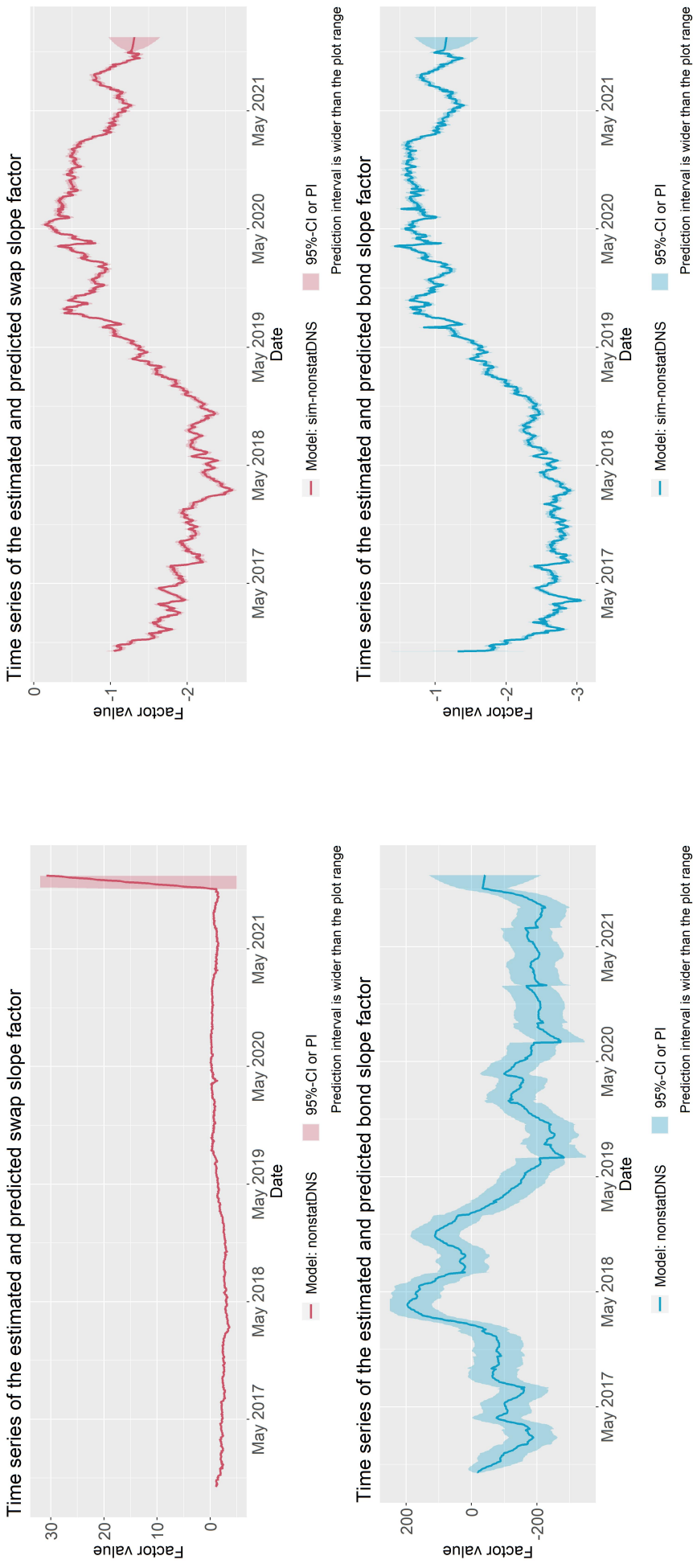


Figure 57: Time series of the estimated and predicted slope factors in the case study

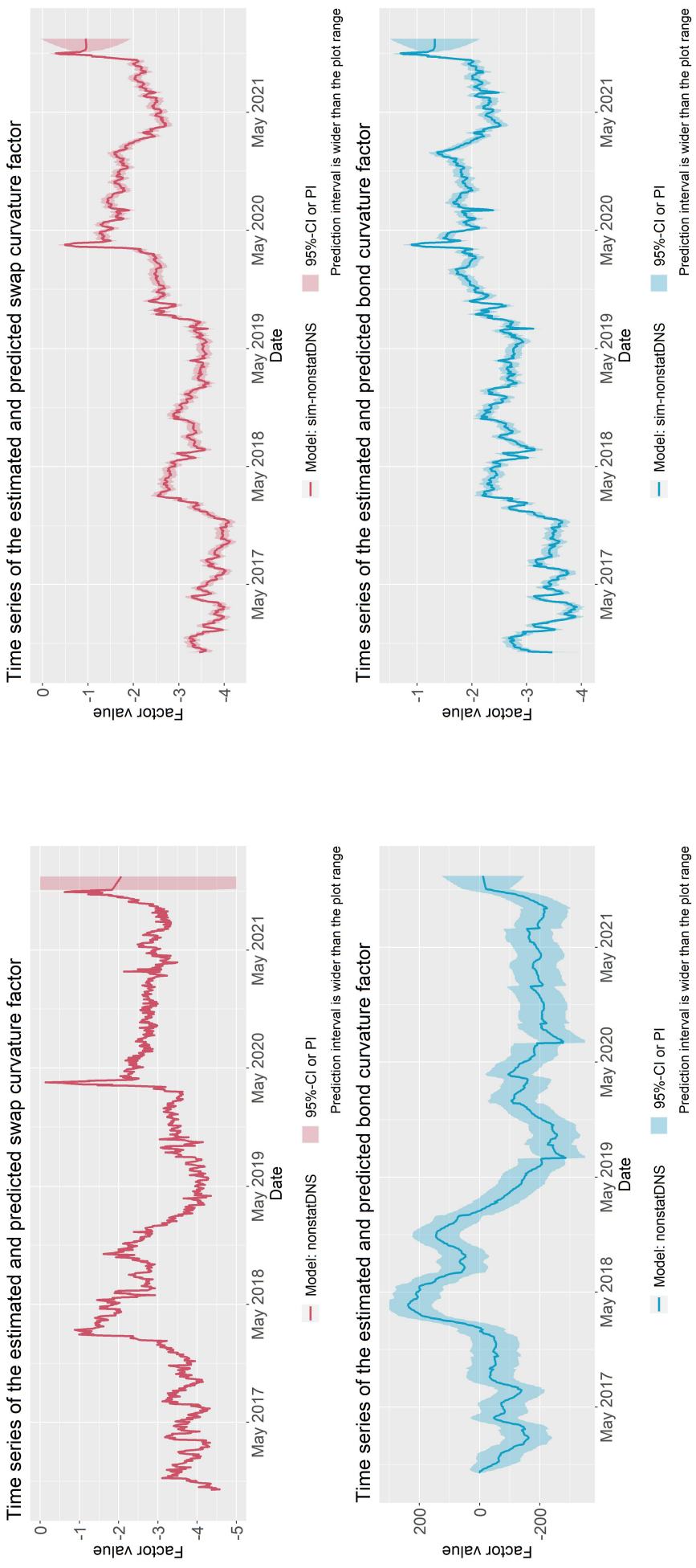


Figure 58: Time series of the estimated and predicted curvature factors in the case study

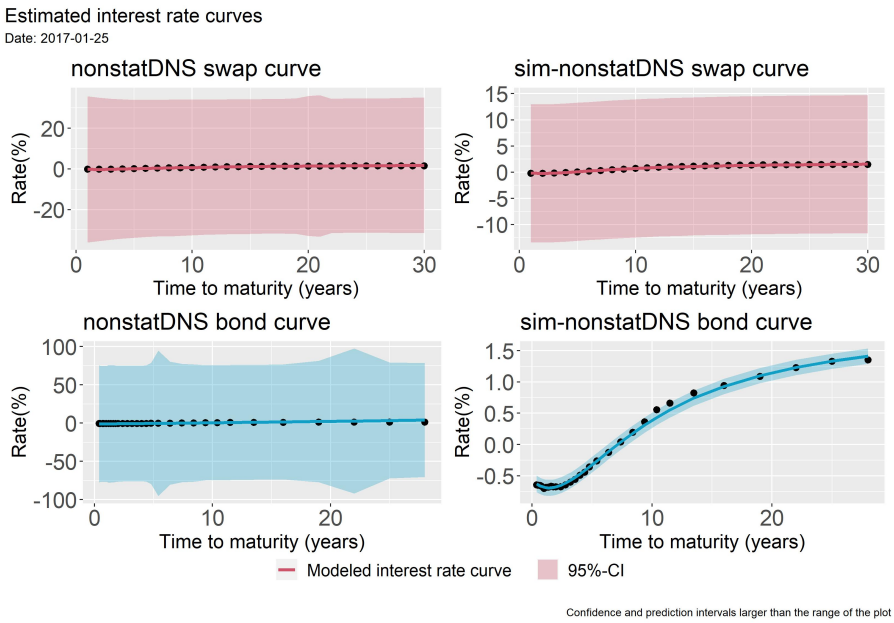


Figure 59: Example of the estimated interest rate curves for the nonstatDNS and the sim-nonstatDNS models with the full 95%-confidence interval.

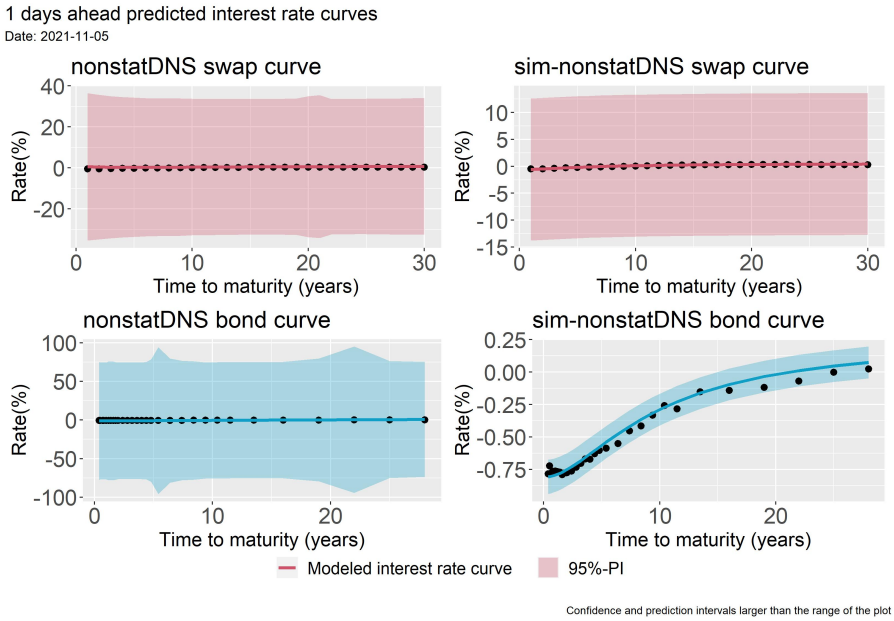


Figure 60: Example of the predicted interest rate curves for the nonstatDNS and the sim-nonstatDNS models with the full 95%-prediction interval.

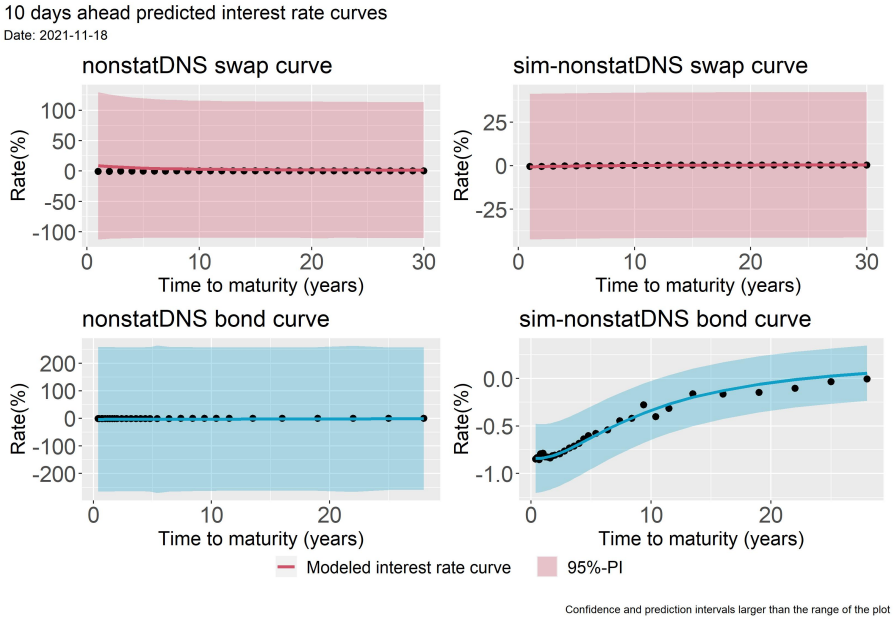
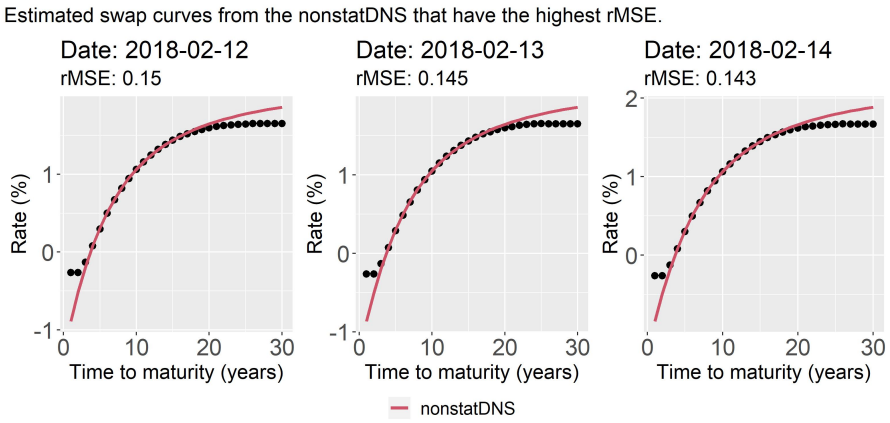
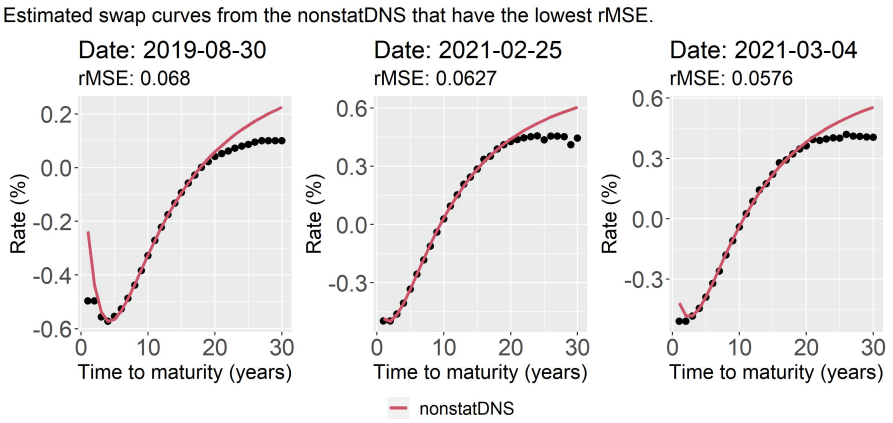


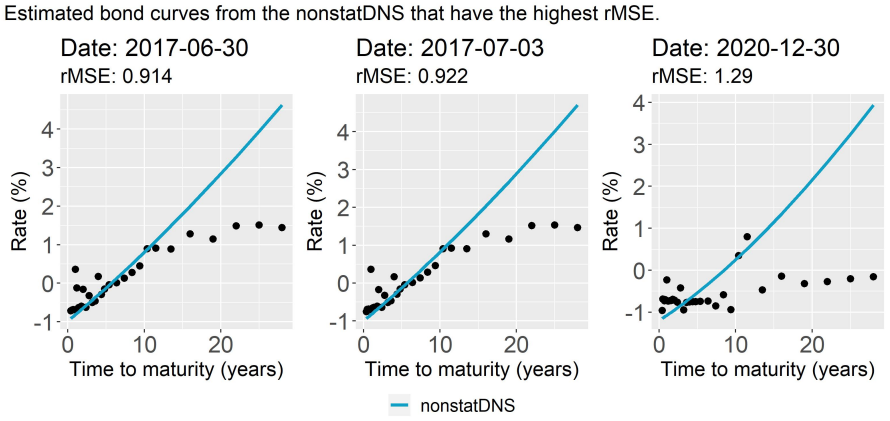
Figure 61: Example of the predicted interest rate curves for the nonstatDNS and the sim-nonstatDNS models with the full 95%-prediction interval.



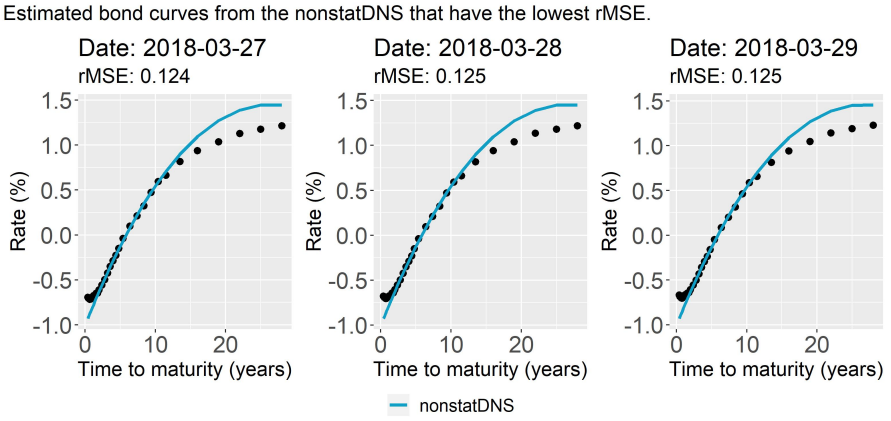
(a) Swap curves estimated by the nonstatDNS model on the real data that have the lowest rMSE



(b) Swap curves estimated by the nonstatDNS model on the real data that have the highest rMSE

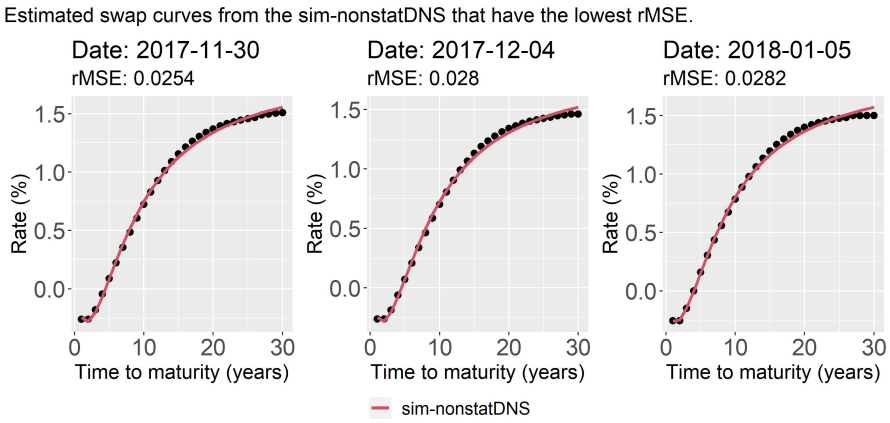


(c) Bond curves estimated by the nonstatDNS model on the real data that have the lowest rMSE

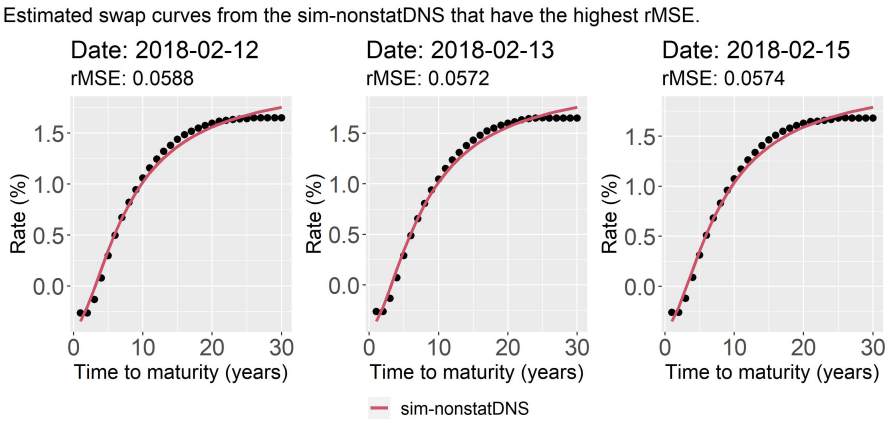


(d) Bond curves estimated by the nonstatDNS model on the real data that have the highest rMSE

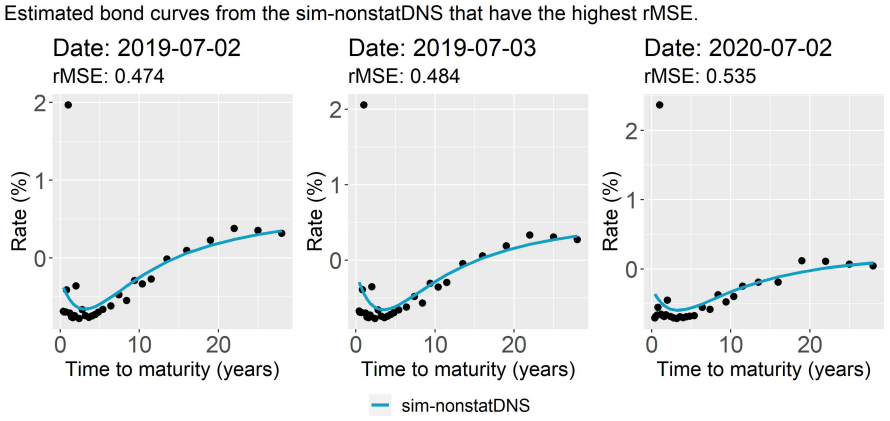
Figure 62: Estimated curves with the lowest and highest rMSE values for the nonstatDNS model



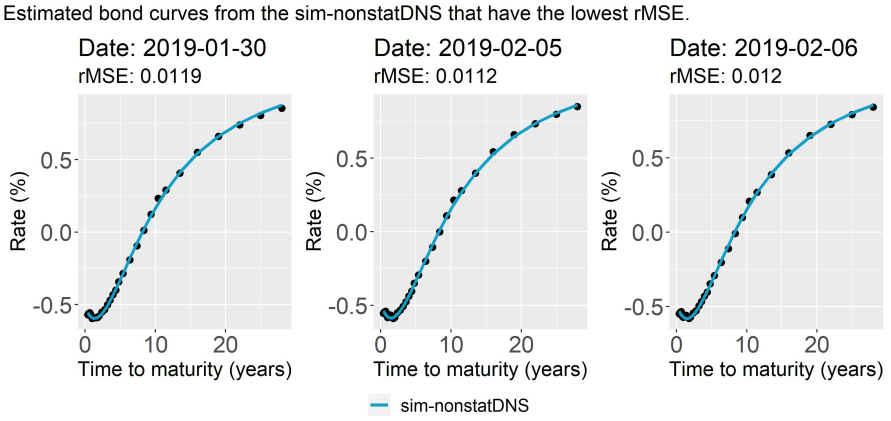
(a) Swap curves estimated by the sim-nonstatDNS model on the real data that have the lowest rMSE



(b) Swap curves estimated by the sim-nonstatDNS model on the real data that have the highest rMSE



(c) Bond curves estimated by the sim-nonstatDNS model on the real data that have the lowest rMSE



(d) Bond curves estimated by the sim-nonstatDNS model on the real data that have the highest rMSE

Figure 63: Estimated curves with the lowest and highest rMSE values for the sim-nonstatDNS model