Yield Curve Modeling And Estimation

Botkuljak, Juraj

JOSIP JURAJ STROSSMAYER UNIVERSITY OF OSIJEK
SCHOOL OF APPLIED MATHEMATICS AND INFORMATICS

University Graduate Study in Mathematics
Module: Financial Mathematics and Statistics

Yield Curve Modeling and Estimation

MASTER'S THESIS

Supervisor:
Dr. Danijel Grahovac

Author:
Juraj Botkuljak

Osijek, 2024

Contents

1 | Introduction

By serving as an essential input into virtually all forms of financial decision-making, interest rates have wide-reaching effects on the whole economy. Given its rate-setting mandate, understanding the dynamics of interest rates is of particular relevance for the central bank. Financial markets provide an information-rich environment needed to extract the relevant aspects of these dynamics. A particularly important aspect is the term structure of interest rates, most commonly represented as a yield curve. By describing the relationship between residual maturity and the level of interest rate, the yield curve provides valuable insight into the market participants' expectations of the future path of interest rates. Given its importance for monetary and macroprudential policy, most major central banks implement an internal system for estimating yield curves ([2], [4], [9], [31], [48]).

In practice, yield curves are derived by fitting a specific functional form to price data of fixed-income securities like bonds and interest rate swaps. Typically, two classes of approaches to fitting the yield curve are employed: spline-based and parsimonious models. Due to their simplicity and ready applicability to the broader framework of term structure modeling, parsimonious models are often preferred by academics and central banks. In particular, extensive literature has been developed over the years on the popular Nelson-Siegel (NS) model, which brought rise to a class of related models, often collectively referred to as the NS class. Although the NS model is still used in practice, the Svensson extension of the NS functional form is preferred among practitioners, including the central banks, for its added flexibility ([48]).

Deploying yield curve models for systematic large-scale applications comes with a particular set of challenges, arising both as a consequence of the models' statistical properties and the methodology applied in estimating them. Specifically, one of the often-reported findings is that the estimated parameters of the NS-class models are prone to exhibit erratic behavior, the severity of which depends on the treatment of the exponential decay parameter τ in the estimation setup ([3], [31], [56]). Given that such behavior can hardly be justified economically, it diminishes the credibility of the model required for its application in a broader context of economic inference and generalization (which is one of its main advantages relative to spline-based alternatives). Due to the specific structure of the functional form, the Svensson extension is particularly prone to this type of behavior, as demonstrated in [31], [56] and Chapter 5 of this thesis. Furthermore, the approach by which the Svensson model extends the NS model cannot be readily generalized further to accommodate more complex yield curve shapes.

This thesis considers an alternative path to extending the NS model with Laguerre polynomials, which has been hinted at in the original paper by Nelson and Siegel ([46]) and explored in more detail in [11], [35], [39], [40], and [41]. The approach is based on the observation that the NS forward rate functional form, when orthogonalized, can be viewed as a special case of the Fourier-Laguerre series expansion. This formulation provides a way to extend the model with an arbitrary number of terms while, in contrast to the Svensson extension, retaining the important theoretical properties of the NS-class that enable its applicability in the broader term structure modeling context ([42]).

Despite favorable theoretical and statistical properties, some of which have been noted in the present literature, an extensive evaluation of this approach has not been conducted. This work contributes to the existing literature with an analysis of the theoretical and empirical profile of such an approach, particularly in comparison to the widely used Svensson model.

When it comes to parameter estimation, the NS model (and especially the Svensson extension) has been known to exhibit erratic behavior of parameters if the nonlinear decay parameter is allowed to vary between consecutive periods. Moreover, certain values of the nonlinear decay parameters, although resulting in an optimal fit, may lead to unreasonable values of the remaining linear parameters, undermining their potential for economic interpretation. This gives rise to the trade-off between stability, goodness-of-fit, and the model's capacity to generalize, which is shown to be mediated by different values of the nonlinear decay parameter. The conducted empirical study aims to investigate the nature of this trade-off when different approaches are used for estimating the nonlinear decay parameter.

Finally, a novel method is tested to estimate model parameters from a sample containing coupon-bearing bonds, when the nonlinear decay parameter is fixed. The method initializes with an initial parameter estimate and proceeds as a sequence of two-step iterations:

1. In the first step, coupons are stripped from each coupon-bearing bond. The procedure involves computing the present value of the bond's coupon payments using the current yield curve estimate and subtracting it from the bond's price.

2. In the second step, the zero-coupon yields are computed for all bonds in the sample. Stripped prices obtained in the previous step are used when computing zero-coupon yields of coupon-bearing bonds. The updated yield curve estimate is obtained by performing ordinary least squares over the sample of zero-coupon yields.

The procedure terminates when the difference in parameter estimates from subsequent iterations is within a specified interval. A heuristic analysis of convergence is presented.

2 | Term structure of interest rates

Interest rates are one of the most fundamental drivers of the economy. Often described as the price of money, they act as a balancing force between the supply and demand for money. Interest rates, almost unlike any other macroeconomic variable, directly influence the decisions of economic participants across all sectors. Whether it is a business evaluating a capital investment, a household requiring a mortgage, or an investor pricing a government-issued asset, interest rates are an essential input.

In its simplest definition, interest rate refers to the agreed percentage amount that the lender is compensated for a loan granted to the borrower. Naturally, there may be as many different values of interest rates as there are potential borrowers. One particular kind is particularly prominent in economic models: the **real short-term risk-free interest rate**, which refers to the rate of interest that is

1. incurred on a hypothetical short-term[1] loan;
2. expressed in real, rather than nominal terms;
3. determined under the assumption of guaranteed loan repayment (i.e., no risk).

In practice, however, economic actors may want to determine the appropriate rate of interest for loans with properties vastly different compared to the ideal one that the real short rate is concerned with. We can consider the implications of deviating from these idealized properties.

1. The loan may have a different **maturity**. The short rate is determined dynamically, in line with the supply and demand for money. Therefore, the appropriate interest rate on a loan that matures at a future point in time needs to reflect the expected time-average value of the short rate until maturity. If the interest rate is higher than the expected short rate (plus a term spread), the borrower may prefer to take on a series of short-term loans instead, repaying the previous one with the subsequent one until the desired maturity is reached.

2. **Inflation** can diminish the real value of (nominal) interest. A rational lender would require to be compensated for this reduction. However, as the inflation rate is not known in advance, its expected value would be factored into the interest rate demanded by the lender.

[1] An infinitesimally short repayment period; it is a theoretical construct ([25]).

3. The borrower may not repay the loan. To compensate for potential losses stemming from the borrower's **credit risk**, the lender will demand an additional premium.

For a standard loan (i.e., without special contractual features), the required interest rate can be assumed to be purely comprised of the three components listed above.[2] If we consider the interest rate as a function of the remaining time to maturity, **residual maturity**, we obtain a **yield curve.**

2.1 Yield curves

The yield curve is a financial indicator depicting the term structure of interest rates. Put more simply, it shows the relation between the residual maturity and the (annual) interest rate, usually relating to some hypothetical future cash flow (e.g., a loan). Since the discounted value of expected future cash flows is directly reflected in the price of financial assets such as bonds, prices of such assets can then be used to reverse-engineer the interest rate that investors apply in their pricing process.

Although yield curves can be estimated from other instruments such as interest rate swap rates and deposit contracts, this thesis focuses on yield curves estimated from government bonds. Government bonds are usually traded by large institutional investors and are typically highly liquid, approximating the theoretical concept of perfectly competitive markets. Furthermore, given the large size of the market for government bonds, government bond yield curves are also essential for analyses related to financial stability.

When analyzing the term structure of interest rates, there are several benefits to modeling yields as continuous (and, ideally, smooth) functions of residual maturity, rather than relying purely on a discrete set of yields.

- First, the set of residual maturities for which yields could be reliably calculated from the available financial market data, without any model assumptions, is often limited. This makes it practically impossible to adequately compare the yields of different issuers, or the same issuer across time. For example, it is unlikely that an issuer, let alone multiple issuers, will have an outstanding bond with exactly 10 years of residual maturity. Having a smooth estimate of the yield curve therefore allows one to obtain an estimate of the yield for an arbitrary residual maturity.

- Second, the yield curve enables effective visualization of the term structure of interest rates. In this way, it allows one to view interest rates as smooth curves across maturities, rather than finite sets of discrete points. Aside from providing the basis for pricing financial instruments, this facilitates more complex inference that is based on analyzing the **shape of the yield curve**. By reflecting interest rate and inflation expectations as well as the term spread, the shape of the yield curve reveals the underlying dynamics of the business

[2] A finer structure of risks that determine the appropriate level of interest rate is given in [25].

cycle. This is why the yield curve assumes a pivotal role in policy modeling for central banks. Additionally, analyzing the effects of yield curve shifts is relevant in the context of interest rate risk management, particularly from the perspective of financial corporations that manage large fixed-income portfolios such as banks and insurance companies ([5]).

A simple and common way to characterize a yield curve's shape is by its **slope**. In a similar way that the yield curve can be considered as a visual analog of the term structure of interest rates, the slope can be seen as the visual analog of the **term spread**. For most practical applications, the slope is defined simply as the difference between the long-term and the short-term yield. 10-year and 3-month yields are commonly chosen to represent long-term and short-term yields, respectively. Although there are ways to define the slope without relying on a somewhat arbitrary choice of residual maturities used to represent short- and long-term yields,[3] the provided definition based on the yield spread is often considered to provide a reasonable combination of accuracy and robustness ([24]). A positive yield curve slope implies a positive term spread: long-term assets yield a higher rate of interest compared to their short-term counterparts. Correspondingly, such yield curves are described as **upward-sloping**. In the converse scenario, the yield curves are referred to as **downward-sloping** or **inverted**. Several representative examples can be seen in Figure 2.1.[4] The characterization of downward-sloping yield curves as inverted explains why an upward-sloping yield curve is often referred to as *the normal yield curve*. Indeed, yield curves are most commonly upward-sloping, as evidenced by their predominantly positive slope historically (Figure 2.1).

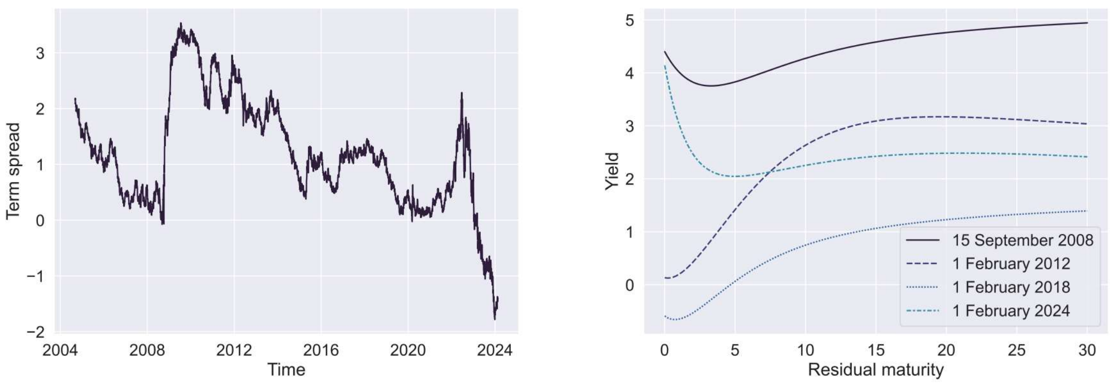

Figure 2.1: Historical 10-year and 3-month yield spread of the euro area yield curve (left); upward-sloping (1 February 2012, 1 February 2018) and inverted (15 September 2008[6], 1 February 2024) euro area yield curve examples (right)

[3]For example, when performing principal component analysis on yield curve time series, the second principal component is often referred to as the slope of the yield curve.

[4]Euro area yield curves can be plotted and compared interactively using the ECB's online tool: ecb.europa.eu/stats/financial_markets_and_interest_rates/euro_area_yield_curves

[6]The slope, calculated as the spread between the 10-year and the 3-month yield, is negative. Alternative definitions (e.g., based on 10- and 1-year yields) would imply an upward-sloping curve.

Yield curve inversion, a relatively rare but significant event characterized by a negative slope, is widely considered a leading indicator of an impending recession ([23]). However, the relationship between the degree of inversion and recession probability is nuanced and has been contested in recent literature ([29]).

2.2 Why is the normal yield curve upward-sloping?

Four hypotheses are popularly used to explain the shape of the yield curve ([48]).

1. **The pure expectations hypothesis** postulates that each interest rate level across the yield curve reflects the expectation of the average short rate in the period between now and a given (residual) maturity. The hypothesis is motivated by the assumption that an investor should be able to, under the absence of arbitrage, substitute any long-term bond investment with a series of short-term bond investments rolled in the period until the long-term bond's maturity. Although theoretically elegant, the pure expectations hypothesis has been demonstrated to fail empirically ([49], [51]). This is also apparent anecdotally: the normal (upward-sloping) shape of the yield curve would, under the pure expectations hypothesis, imply the expectation of rising short-term interest rates. However, despite the prevalence of the upward-sloping shape of the euro area yield curves in recent history, short-term interest rates have not been increasing; instead, they have been decreasing or stagnating.[7] A similar trend can be observed across other major economies.

2. **The segmented market hypothesis** relies on the assumption that debt instruments of varying maturities are not substitutes. Instead, different groups of investors (market segments) are dominant in generating the supply and demand of debt instruments with particular maturities. The dynamics of specific preferences across these segments is what generates the particular yield curve shape.

3. According to **the preferred habitat hypothesis**, investors have a preferred investment horizon and require a sufficient premium to deviate from it. Under this hypothesis, a relative surplus of demand for short-term debt (compared to long-term debt) would generate the observed upward-sloping shape.

4. **The liquidity preference hypothesis** complements the pure expectations hypothesis with an additional assumption that investors demand a premium over the expected short-term rate to compensate for tying up liquidity in a long-term asset. If the interest rates were to rise after a bond has been acquired, the bond price would drop, resulting in a capital loss if the investor liquidates the bond before maturity. To compensate for this liquidity risk, a rational investor will demand an excess return on long-duration bonds. The fact that risk is assumed to increase with the duration for which the liquidity

[7] The ECB deposit facility rate can be found on the ECB Data Portal: data.ecb.europa.eu/data/datasets/FM/FM.D.U2.EUR.4F.KR.DFR.LEV

is tied up (i.e., the liquidity premium rises as a function of duration) explains the prevalence of the positive yield curve slope.

2.3 Yield curves at the European Central Bank

In many aspects, monetary policy can be viewed as a two-way interaction: the central bank acts and reacts to input it receives from the economy, adjusting its policy to achieve optimal effectiveness in implementing its mandate. Financial markets form a key part of this interaction. Apart from acting as one of the key channels of monetary policy transmission, they contain real-time information on market participants' evaluation of the monetary policy and its expected implications. Financial markets, being a key component of the financial system, also reflect developments that are relevant for the broader financial stability. Therefore, the timely availability of high-quality financial indicators is essential for supporting informed decision-making at the European Central Bank (ECB).

The ECB estimates and publishes two euro area government bond yield curves:[8]

- The *AAA-rated yield curve* is estimated from euro area government bonds rated AAA, reflecting the most favorable credit risk assessment.
- The second is estimated from all euro area government bonds.

Since the AAA-rated yield curve reflects market expectations of nominal interest rates under negligible credit risk, it can be used as a proxy for the risk-free yield curve.[9] Generally, risk-free yield curves are more universally applicable as they don't reflect issuer-specific credit risk. Instead, they are primarily influenced by macroeconomic factors such as market expectations of the real interest rate, inflation, and the term premium. These factors play a crucial role in informing monetary policy decisions. Given that it approximates the lower bound to borrowing costs, the euro risk-free yield curve also facilitates the computation of credit risk premia for euro-denominated debt, which is vital for assessing financial stability and integration in the euro area.

The contribution of expected inflation can be eliminated from the (nominal) risk-free yield curve by subtracting the implied inflation rates ([22]). **Real risk-free yield curves** reflect solely the expected real short-term rate and the term premium. Decomposing the yield curve further generally requires a dedicated theoretical framework. Despite the usefulness of real yield curves for economic analysis, nominal yield curves are most commonly reported.

Although they are most commonly used to model nominal government bond yield curves, most standard yield curve modeling frameworks, including the one presented in the next chapter, can be readily applied to a wider array of modeling contexts (e.g., corporate bond yield curves, swap curves, and real yield curves derived from inflation-linked bonds).

[8] ecb.europa.eu/stats/financial_markets_and_interest_rates/euro_area_yield_curves/html/index.en.html

[9] Among issuers with the same credit rating, the actual credit risk as recognized by the market can vary. German Bunds have become widely accepted as a proxy for the euro risk-free rate due to their perceived high credit quality (even relative to its AAA-rated peers).

3 | Yield curve modeling

In practice, yield curves are not observed. The indirect influence of the yield curve can however be observed in the prices of financial assets. The primary goal of the yield curve model, therefore, is to infer the yield curve that is implied in these observed prices. Since prices, rather than yields, are observable directly in financial markets, the estimation problem cannot be generally reduced to simple curve fitting. This is, however, possible in the special case in which the sample contains only **zero-coupon bonds**.

The fair price of a risk-free zero-coupon bond paying a unit principal at time t is given by

$$p(t) = \exp\left(-ty(t)\right), \tag{3.1}$$

where t is residual maturity, and $y(t)$ is the continuously compounded annual yield at maturity t. Solving for yield in (3.1), we can obtain an expression for transforming a sample of prices into a sample of (zero-coupon) yields

$$y(t) = -\frac{1}{t}\ln(p(t)). \tag{3.2}$$

From the obtained sample of maturities and yields $(t_1, y_1), \ldots, (t_N, y_N)$, we can reformulate the objective to look for a curve that best fits the set of points (Figure 3.1).[1]

3.1 Common approaches to modeling yield curves

There is a multitude of criteria according to which one can categorize various yield curve models. In most applications, it is practical to consider two classes ([48]):

- spline and kernel-based models; and
- parsimonious models.

Spline and kernel-based models can have arbitrary flexibility, and consequently superior goodness-of-fit. This makes them particularly useful if the estimated

[1]Since zero-coupon bonds are scarce for longer residual maturities (more than 5 years; see Figure 5.1), Figure 3.1 also includes prices and yields of coupon bonds whose coupon has been stripped (i.e., their value subtracted to obtain an estimate of the equivalent zero-coupon price and yield) using the method described in Section 4.2.

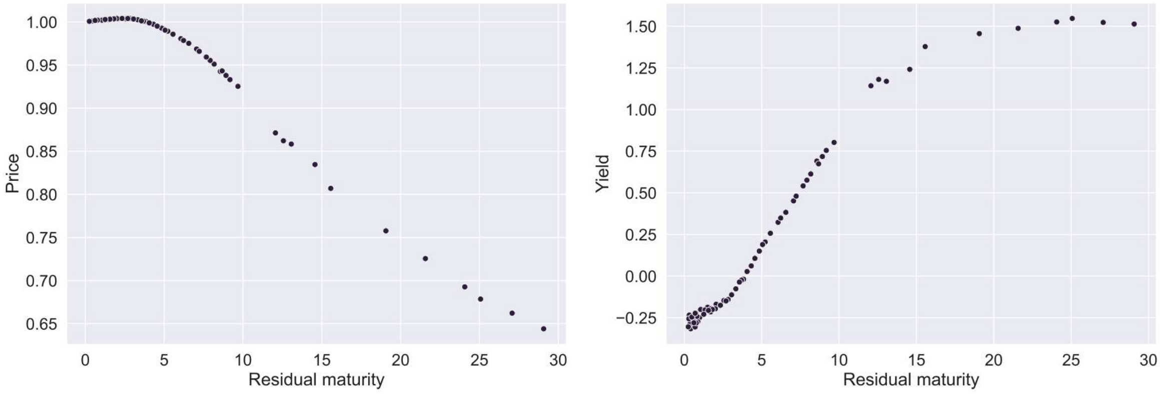

Figure 3.1: 15 June 2015 closing prices (left) and yields (right) of German zero-coupon government bonds plotted against residual maturity

curve is required to closely match the observed bond yields (e.g. when pricing financial assets).[2] Spline models, originally introduced by McCulloh ([45]), typically involve piecewise cubic polynomials that are joined at knot points, the number of which can be arbitrary. Normally, a larger number of knot points allows for more flexibility but risks overfitting the observed data, potentially resulting in unrealistic shapes. It was later identified that the roughness of the curve produced by the original McCulloh model was contributing to instability in the estimated forward curve ([55]). Although such behavior is theoretically possible, frequent and significant changes in the term structure are rather unlikely. Further developments incorporated shape constraints to counteract such behavior, often in the form of a roughness penalty. The model put forward by Waggoner ([55]) involved applying a variable roughness penalty in the form of a monotonically increasing three-tiered step function that increases at maturities 1 and 10. Applying a higher roughness penalty for longer maturities had a dampening effect on long-end oscillations. The Waggoner model is used among several major central banks ([48]), including the Bank of England ([2]).

Spline and kernel-based models are often considered examples of non-parametric regression, not necessarily because they do not involve parameters, but primarily because parameters are not of substantive interest. In contrast, the defining characteristic of well-specified parsimonious models is that the estimated parameters reflect efficient representations of the term structure that can be subjected to further analysis (using, for example, time series methods). This feature makes such models appealing to researchers and policymakers whose objective goes beyond achieving maximum fitting accuracy. For this reason, parsimonious models are predominantly used by the central banks (although exceptions exist). Generally, parsimonious models are based on a functional form that attempts to capture the stylized features of the yield curve, with the ultimate goal of achieving parsimony. Parsimony, however, often comes at the expense of in-sample goodness-of-fit but often performs similarly out-of-sample ([7]). Moreover, the

[2]Cboe Volatility Index® uses a cubic spline to interpolate yields used in the calculation of implied volatility ([13]).

lack of flexibility may serve to prevent overfitting, which can disrupt inference about variables such as interest rate expectations. Ultimately, the key challenge when defining such models is to identify a functional form that closely matches various yield curve shapes with a minimum number of parameters.

3.1.1 Nelson-Siegel class

The relation (3.2) establishes the link between the zero-coupon bond's price and its (zero-coupon) yield. The implicit assumption in this definition is that the accumulation period starts now (at $s = 0$) and extends until maturity ($s = t$). If the start of the accumulation period is instead deferred to some $t_1 > 0$ and lasts until t_2, the applicable interest for the period $[t_1, t_2]$ is referred to as the **forward rate** ([25])

$$f(t_1, t_2) = \frac{t_2 y(t_2) - t_1 y(t_1)}{t_2 - t_1} \overset{(3.1)}{=} \frac{1}{t_2 - t_1}(\ln p(t_1) - \ln p(t_2)). \tag{3.3}$$

If we fix $h > 0$ and set $t_2 = t_1 + h$, we can define the h-year[3] forward rate

$$f_{(h)}(t) = \frac{1}{h}(\ln p(t) - \ln p(t+h)). \tag{3.4}$$

The instantaneous forward rate (also referred to as the **force of interest**) reflects the annualized interest rate associated with an infinitesimally short period $h \to 0$ starting at some t

$$f(t) = \lim_{h \to 0} \frac{1}{h}(\ln p(t) - \ln p(t+h)) = -\frac{d}{dt} \ln p(t). \tag{3.5}$$

The instantaneous forward rate can be viewed as the forward analog of the short-term rate described in the previous chapter.

In most practical cases, *the yield* of a bond refers to its **spot rate**, which is the constant annual interest rate applicable to the period between now and some time t. Spot rate can be represented as the time average of the instantaneous forward rate in the period $[0, t]$

$$y(t) \overset{(3.2)}{=} -\frac{1}{t}\ln(p(t)) = -\frac{1}{t}\int_0^t \frac{d}{ds} \ln p(s) ds \overset{(3.5)}{=} \frac{1}{t}\int_0^t f(s) ds. \tag{3.6}$$

Nelson-Siegel class includes models based on the original Nelson-Siegel model, which proposes a functional form for the instantaneous forward curve ([46])

$$f(t; \boldsymbol{\beta}, \tau) = \beta_0 + \beta_1 \exp\left(-\frac{t}{\tau}\right) + \beta_2 \frac{t}{\tau} \exp\left(-\frac{t}{\tau}\right). \tag{3.7}$$

Using (3.6) we can derive the spot rate

$$\begin{aligned} y(t; \boldsymbol{\beta}, \tau) &= \frac{1}{t}\int_0^t f(s; \boldsymbol{\beta}, \tau) ds \\ &= \beta_0 + \beta_1 \left[\frac{1 - \exp(-\frac{t}{\tau})}{t/\tau}\right] + \beta_2 \left[\frac{1 - \exp(-\frac{t}{\tau})}{t/\tau} - \exp\left(-\frac{t}{\tau}\right)\right]. \end{aligned} \tag{3.8}$$

[3] The time domain is typically assumed to be expressed in years.

Owing to its favorable empirical and theoretical properties, the *NS* model is popular and widely used in academia and practice, particularly among central banks. The model can successfully fit a wide range of common yield curve shapes, contributing to an overall favorable in and out-of-sample performance profile ([18], [3]). The model has also been shown to be useful in yield curve prediction ([26]). The theoretical appeal of the model stems mostly from the structure of its functional form, which consists of three basis curves that can be conveniently described as the *level*, *shape*, and *curvature* ([44]). This representation is also consistent with loadings obtained via principal component analysis applied to yield curve data ([8]).[4] To better illustrate this intuition, the functional form (3.7) can be rewritten as a linear combination of its basis curves

$$f(t;\boldsymbol{\beta},\tau) = \beta_0 f_0(t;\tau) + \beta_1 f_1(t;\tau) + \beta_2 f_2(t;\tau). \tag{3.9}$$

- The first $f_0(t;\tau) = 1$ is constant and corresponds to the **level** of the yield curve. Since the remaining two basis curves approach zero in infinite time, the corresponding parameter, β_0, can then be described as the long-term (asymptotic) yield. When treated as time series, changes in β_0 reflect vertical shifts of the yield curve. It can be easily checked algebraically that the same interpretation applies to both forward and spot curves. The fact that the *NS* model effectively constrains the long end of the curve is an appealing feature of the model. It is supported empirically and consistent with the concept of the (long-term) nominal neutral rate. This stylized feature is not directly enforced by a spline model, which can potentially diverge to large (positive or negative) values when extrapolated to long maturities. This unnatural behavior suggests potentially poor out-of-sample performance, which is why most practical spline models incorporate additional model elements to constrain such behavior ([46]).

- The second basis curve $f_1(t;\tau) = \exp\left(-\frac{t}{\tau}\right)$ is a decreasing function, corresponding to the **slope** of the yield curve. Since $f_0(0;\tau) = f_1(0;\tau) = 1$ and $f_2(0;\tau) = 0$, it follows that $f(0) = \beta_0 + \beta_1$. This provides a natural interpretation of β_1 as the spread between the short (at $t = 0$) and long-term yield (at $t \to \infty$).

- The third basis curve is a concave function, corresponding to the **curvature** of the yield curve. It affects the shape of the yield curve primarily in the medium-term maturity range. Due to its concave nature, the third basis curve has been visually associated with a *U-shaped hump* in the yield curve. The magnitude and direction of this hump are determined by the parameter β_3 ([48]).

Parameter τ is an exponential decay parameter. Typically, lower values imply more curvature in the short end, balanced by more smoothness in the long end. The converse holds for larger values of τ ([46]). The parameter can be also thought of as determining the residual maturity at which the third (curvature)

[4] [17] shows that a similar pattern of principal curves emerges in other, seemingly unrelated, applications.

component achieves its maximum ([20]). Since τ enters the functional form in the exponent, it has a nonlinear effect. In fact, τ is the only nonlinear parameter; if fixed to a constant, it would render the functional form (3.7) linear in the remaining parameters. Fixing τ to a prescribed value is a commonly used approach to fitting the NS model, as detailed in Chapter 4.

The NS model has attracted significant research interest, which is partly because it performs well empirically, yet has few parameters with an intuitive interpretation. This makes the NS model readily applicable in a broader context of term structure modeling. Diebold and Li ([20]) used the NS functional form to derive the dynamic Nelson-Siegel (DNS) model in which $\beta_t = (\beta_{1t}, \beta_{2t}, \beta_{3t})$ is treated as a latent dynamic factor. Despite favorable theoretical and empirical properties, the (dynamic) NS model, however, is lacking in one important theoretical aspect: it does not comply with the **no-arbitrage assumption**, which is foundational in finance. The assumption asserts that in perfectly competitive markets,[5] all potential opportunities for riskless profit (i.e., arbitrage) are instantly eliminated through supply and demand forces. The departure from the no-arbitrage assumption is manifested in the finding that it is not possible to define a nontrivial interest rate model that would produce the forward curves from the NS family ([6], [28]). The appropriate corrections to the functional form need to be incorporated to render the model compliant with the no-arbitrage assumption ([15], [41]). The obtained class of **arbitrage-free Nelson-Siegel (AFNS)** models is argued to be more theoretically rigorous, providing the basis for its use in a time series context. However, the finding by Coroneo, Nyholm, and Vivada-Koleva that the original DNS model is compatible with the no-arbitrage assumption in the sense that the estimated parameters are not statistically different from its arbitrage-free counterpart ([16]) brings into question the practical usefulness of such a modification.

Although most yield curve shapes are well-captured by the NS curve family, certain more complex term structures may require more flexibility to reach a satisfactory goodness of fit. Below, we explore two approaches to increasing the flexibility of the NS model.

3.1.2 Nelson-Siegel-Svensson model

The NS model extended by Svensson ([53]) is the most widely used model by major central banks ([4]), including the Eurosystem ([48]) and the Federal Reserve ([31]). The Nelson-Siegel-Svensson (NSS) model extends the functional form (3.7) by including the fourth term

$$f(t;\boldsymbol{\beta},\tau) = \beta_0 + \beta_1 \exp\left(-\frac{t}{\tau_1}\right) + \beta_2 \frac{t}{\tau_1} \exp\left(-\frac{t}{\tau_1}\right) + \beta_3 \frac{t}{\tau_2} \exp\left(-\frac{t}{\tau_2}\right). \quad (3.10)$$

Similarly, the functional form can be rewritten in terms of basis curves to explore the underlying intuition

$$f(t;\boldsymbol{\beta},\tau) = \beta_0 f_0(t) + \beta_1 f_1(t;\tau_1) + \beta_2 f_2(t;\tau_1) + \beta_3 f_3(t;\tau_2). \quad (3.11)$$

[5] Bonds typically trade in deep and well-organized markets, matching closely the characteristics of a perfectly competitive market.

The functional form of f_3 is identical to f_2, but it is parametrized by a different nonlinear parameter τ_2. The nonlinear decay parameter is therefore a two-dimensional vector $\tau = (\tau_1, \tau_2)$. Similar to how f_2 can be associated with a U-shaped hump, an analogous association can be made for f_3. However, since the shape of f_3 is determined by a different decay parameter τ_2, the characteristics of this hump can be different. For this reason, the Svensson extension of the *NS* model can be described as adding a secondary U-shaped hump to the functional form.

As in the *NS* model, the functional form of the spot yield curve can be derived from (3.10) by applying (3.6)

$$
\begin{aligned}
y(t;\boldsymbol{\beta},\tau) = \beta_0 + \beta_1 \left[\frac{1-\exp(-\frac{t}{\tau_1})}{t/\tau_1} \right] + \beta_2 \left[\frac{1-\exp(-\frac{t}{\tau_1})}{t/\tau_1} - \exp\left(-\frac{t}{\tau_1}\right) \right] \\
+ \beta_3 \left[\frac{1-\exp(-\frac{t}{\tau_2})}{t/\tau_2} - \exp\left(-\frac{t}{\tau_2}\right) \right].
\end{aligned}
\tag{3.12}
$$

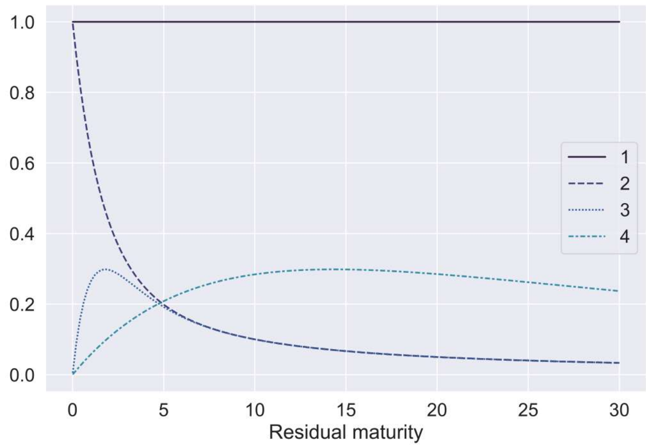

Figure 3.2: *NSS* model, spot rate basis curves for $\tau = (1, 10)$

The popularity of the *NSS* model, particularly among central banks, can be in part attributed to its ability to strike a good balance between flexibility and simplicity. However, the model is deficient in certain important areas.

- First, it has two nonlinear decay parameters that complicate model calibration, often involving a non-convex optimization problem. The model may also be badly conditioned for certain regions of $\tau = (\tau_1, \tau_2)$, leading to a situation in which small perturbations in data lead to disproportionate changes in parameter estimates. As an example, if $\tau_1 = \tau_2$, the corresponding two U-shaped humps will coincide and the model degenerates to a standard *NS* model with the magnitude of the curvature equal to the sum $\beta_2 + \beta_3$, resulting in perfect collinearity. It has been demonstrated in [31] that having

$\tau_1 \approx \tau_2$ can lead to β_2, and β_3 parameters assuming large absolute values, up to 10^5. In fact, all parameters of the NSS model are susceptible to instability, even though estimated yield curves may exhibit a relatively stable shape and favorable goodness-of-fit scores ([31], [56]). Without certain adjustments, this property of the model makes it difficult to argue that estimated parameters reflect *true* structural quantities that can be used for meaningful economic reasoning and inference.

- Second, some of the favorable theoretical properties of the NS model are not preserved with the Svensson extension. Christensen, Diebold, and Rudebusch show that it is not possible to incorporate corrections that would render the model compliant with the no-arbitrage assumption ([14]). This is because the NSS model includes an extra *curvature* without the corresponding *slope* component. This fact negatively affects the NSS model's potential for use in a more general time series and term structure modeling context ([35], [42]).

- Finally, the Svensson functional form does not provide a general way to extend the model with additional terms, possibly to fit more complex yield curve shapes.

3.2 A model based on Laguerre polynomials

Laguerre polynomials are one of three classical orthogonal polynomials. The closed form for the k-th Laguerre polynomial is given by

$$L_k(x) = \sum_{j=0}^{k} \binom{k}{j} \frac{(-1)^j}{j!} x^j. \tag{3.13}$$

The first few Laguerre polynomials are given in Table 3.1.

k	$L_k(x)$
0	1
1	$-x+1$
2	$\frac{1}{2}(x^2-4x+2)$
3	$\frac{1}{6}(-x^3+9x^2-18x+6)$
4	$\frac{1}{24}(x^4-16x^3+72x^2-96x+24)$

Table 3.1: First few Laguerre polynomials

Laguerre polynomials are orthogonal on the positive semi-infinite interval $(0, \infty)$ with respect to the measure $w(x) = e^{-x}$. Formally,

$$\int_0^\infty e^{-s} L_i(s) L_j(s) ds = 0, \ i \neq j. \tag{3.14}$$

From the properties of generalized Fourier series, any function $a(x) : \mathbb{R}_+ \to \mathbb{R}$ that is square-integrable with respect to the measure $w(x) = e^{-x}$

$$\int_0^\infty e^{-s} a(s)^2 ds < \infty \tag{3.15}$$

can be expanded into **the Fourier-Laguerre series**

$$a(x) = \exp\left(-\frac{x}{2}\right) \sum_{k=0}^{\infty} c_k L_k(x), \tag{3.16}$$

where $c_k \in \mathbb{R}$ are coefficients. The condition (3.15) holds for all functions of practical interest, including all conceivable yield curve shapes. Representation (3.16) provides the starting point for using Laguerre polynomials to approximate a function with arbitrary accuracy, which is done by choosing the desired number of polynomials as the basis.[6] The coefficients $c_0, \ldots, c_K \in \mathbb{R}$ can be estimated using **polynomial regression** ([33]).

Motivated by (3.16), the **Orthogonal Laguerre Polynomial (OLP) model** of degree K defines the instantaneous forward curve as a constant plus $K - 1$ Fourier-Laguerre terms

$$f(t; \boldsymbol{\beta}, \tau) = \beta_0 + \exp\left(-\frac{t}{\tau}\right) \sum_{k=1}^{K-1} \beta_k L_{k-1}\left(\frac{2t}{\tau}\right). \tag{3.17}$$

Remark 1. *Although not a part of the standard Fourier-Laguerre expansion, the inclusion of a constant β_0 in (3.17) is convenient for modeling the asymptotic interest rate (the level component) implied by the instantaneous forward rate curve. This formulation can be derived from (3.16) if we set $a(x) = f(x) - \beta_0$.*

In representation (3.17), the polynomial basis depends on the decay parameter τ which scales the independent variable t, producing nonlinear effects. Similar to other NS-class models, the role of the decay parameter can be described as controlling how curvature is distributed along the residual maturity axis. Lower values imply faster decay, concentrating curvature at lower residual maturities.

Like the NS-class models, the spot yield functional form is derived from (3.6)

$$y(t; \boldsymbol{\beta}, \tau) = \beta_0 + \sum_{k=1}^{K-1} \beta_k y_k(t) = \beta_0 + \sum_{k=1}^{K-1} \beta_k \frac{1}{t} \int_0^t f_k(s) ds, \tag{3.18}$$

with f_k and y_k being the instantaneous forward rate and the spot rate curve basis curves, respectively, several of which are given in Table 3.2.

[6] A more comprehensive overview of the theory behind orthogonal polynomials is given in [54].

k	$f_k(t)$	$y_k(t)$
1	$\exp\left(-\frac{t}{\tau}\right)$	$\frac{\tau}{t} - \frac{\tau}{t}\exp\left(-\frac{t}{\tau}\right)$
2	$\exp\left(-\frac{t}{\tau}\right)\left(\frac{-2t}{\tau}+1\right)$	$\frac{\tau}{t}\left(\frac{2t}{\tau}+1\right)\exp\left(-\frac{t}{\tau}\right) - \frac{\tau}{t}$
3	$\exp\left(-\frac{t}{\tau}\right)\left(\frac{2t^2}{\tau^2} - \frac{4t}{\tau} + 1\right)$	$\frac{\tau}{t} - \frac{\tau}{t}\left(\frac{2t^2}{\tau^2}+1\right)\exp\left(-\frac{t}{\tau}\right)$
4	$\exp\left(-\frac{t}{\tau}\right)\left(-\frac{4t^3}{\tau^3} + \frac{18t^2}{\tau^2} - \frac{18t}{\tau} + 3\right)$	$\frac{\tau}{t}\left(\frac{4t^3}{\tau^3} - \frac{6t^2}{\tau^2} + \frac{6t}{\tau} + 3\right)\exp\left(-\frac{t}{\tau}\right) - \frac{3\tau}{t}$
5	$\exp\left(-\frac{t}{\tau}\right)\left(\frac{2t^4}{\tau^4} - \frac{16t^3}{\tau^3} + \frac{36t^2}{\tau^2} - \frac{24t}{\tau} + 3\right)$	$\frac{3\tau}{t} - \frac{\tau}{t}\left(\frac{2t^4}{\tau^4} - \frac{8t^3}{\tau^3} + \frac{12t^2}{\tau^2} + 3\right)\exp\left(-\frac{t}{\tau}\right)$

Table 3.2: First few elements of the *OLP* basis; instantaneous forward rate (f_k) and spot rate (y_k)

Remark 2. *If an alternative formulation of the instantaneous forward rate is used (one in which the parameter to* L_{k-1} *is not scaled by* 2)

$$\tilde{f}(t;\boldsymbol{\beta},\tau) = \beta_0 + \exp\left(-\frac{t}{\tau}\right)\sum_{k=1}^{K-1}\beta_k L_{k-1}\left(\frac{t}{\tau}\right), \tag{3.19}$$

the functional form of the spot rate assumes a general form ([35])

$$\tilde{y}(t;\boldsymbol{\beta},\tau) = \beta_0 + \beta_1\frac{\tau}{t}\left(1-\exp\left(-\frac{t}{\tau}\right)\right) + \exp\left(-\frac{t}{\tau}\right)\sum_{k=1}^{K-2}\frac{\beta_{k+1}}{k}\sum_{j=0}^{k-1}L_j\left(-\frac{t}{\tau}\right). \tag{3.20}$$

However, scaling by 2 *ensures the orthogonality of the polynomial basis, which can be easily checked from the orthogonality property of Laguerre polynomials* (3.14).

Generally, several main reasons motivate the use of the Laguerre polynomial basis to model yield curves.

- First, it is a natural generalization of the *NS* model.[7] Linear transformation of f_2 in the *NS* functional form (3.7) to obtain an orthogonal basis on $(0,\infty)$ would yield

$$\begin{aligned} f(t;\boldsymbol{\beta},\tau) &= \beta_0 + \beta_1\exp\left(-\frac{t}{\tau}\right) + \beta_2\exp\left(-\frac{t}{\tau}\right)\left(1-\frac{2t}{\tau}\right) \\ &= \beta_0 + \beta_1\exp\left(-\frac{t}{\tau}\right)L_0\left(\frac{2t}{\tau}\right) + \beta_2\exp\left(-\frac{t}{\tau}\right)L_1\left(\frac{2t}{\tau}\right), \end{aligned} \tag{3.21}$$

  which is equivalent to *OLP*(3). The theoretical consistency of the *OLP* generalization ensures that many of the important theoretical properties of the *NS* model are readily extensible to the *OLP* model. Most importantly, Krippner showed that the *OLP* functional form can be adjusted to obtain an arbitrage-free and intertemporally consistent specification ([40]).

[7] Extension using Laguerre polynomials has been implicated by Nelson and Siegel in [46].

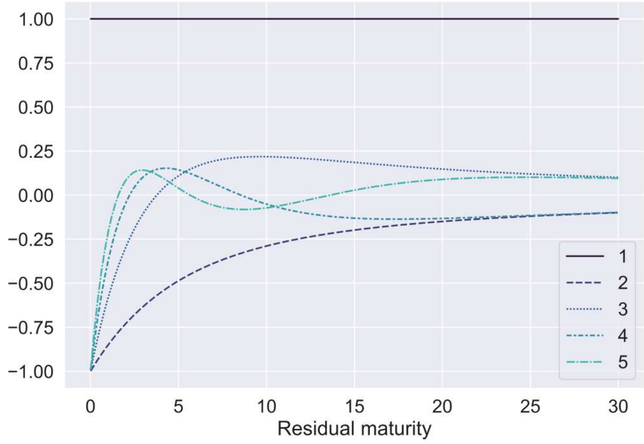

Figure 3.3: *OLP* model, first four spot rate basis curves for $\tau = 3$

- Second, the model doesn't prescribe the number of parameters, allowing one to readily adjust the degrees of freedom to optimize the parsimony-accuracy trade-off for a given use case. Given the varying complexity of yield curve shapes through time, the possibility to dynamically fine-tune the optimal degree of model flexibility can enhance its overall performance: during periods of relatively *simple* yield curve shapes, excess flexibility can lead to overfitting which compromises out-of-sample forecasting performance. Conversely, periods of relatively *complex* yield curve shapes warrant more degrees of freedom to result in adequate goodness-of-fit. Dynamic adjustment of model flexibility need not be manually performed by a practitioner; rather, model selection methods (e.g. information criteria, L_1 regularization) can be applied to automate the decision process.

- Finally, the orthogonality of the polynomial basis may lead to better empirical properties of parameter estimates. The degree to which the orthogonality of the polynomial basis leads to a reduction in regressor multicollinearity will depend on two factors: the range of residual maturities included in the sample (Laguerre polynomials are orthogonal on $(0, \infty)$; not on a finite interval) and the value of τ (which scales the residual maturity parameter, thus controlling the distribution of curvature across the residual maturity axis). Moreover, the orthogonality of the *OLP* basis used to represent the instantaneous forward rate, which is not observable in practice for individual bonds, does not imply the orthogonality of the basis used to represent the spot rate (3.18). Nevertheless, the *OLP* specification appears to be empirically robust, as noted in [39] and exemplified in Chapter 5.

Although a forward rate functional form based on Laguerre polynomials has been explored in the literature ([11], [35], [39]), it hasn't garnered much attention from academics and practitioners. Correspondingly, extensive empirical studies involving this approach have not been conducted.

Misspeficiation of yield curve models. Most yield curve models are simplifications and are likely misspecified as statistical models in a strict sense ([21]). It is difficult to argue that residual maturity is the sole factor on which the term structure of interest rates depends. Although models that incorporate various macroeconomic variables in addition to maturity do exist, it is still likely that the actual dynamics of the term structure are very complicated. Most yield curve models can be seen as curve fitting problems, rather than statistical models in a strict sense. That does not mean that such models are not useful. It does mean, however, that care must be taken when statistical inference is based on these models. In the context of a misspecified model, there is no *true value* for which $f_{\boldsymbol{\beta}}$ represents the data-generating distribution. Instead, the **pseudo-true value** is defined to be $\boldsymbol{\beta}^*$, which, within the assumed model, best explains the data distribution in terms of the chosen likelihood function. In the context of OLS estimation, that likelihood function is implicitly assumed to be Gaussian.

4 | Model fitting and deployment

While prices of individual bonds are observable in practice, yields are generally not. Because of this, and the fact that yield curve models typically model yields instead of prices, one cannot simply apply classical curve fitting methods to tackle the parameter estimation problem. Such methods are applicable in the (fortunate but rare) case when a sufficiently sized sample of zero-coupon bond prices is available. As seen previously, these prices can be transformed into yields with a simple closed-form expression (3.2).

Given a sample of zero-coupon yields with corresponding maturities $\{(t_1, y_1), \ldots, (t_N, y_N)\}$, model parameters can then be estimated using the least squares approach by solving

$$\min_{\beta,\tau} \sum_{i=1}^{N} (y(t_i; \boldsymbol{\beta}, \tau) - y_i)^2. \tag{4.1}$$

The least squares estimator is equivalent to the one obtained with maximum likelihood estimation under the assumption of independent normally distributed yield errors. However, the error normality assumption may not be fully justified, especially for longer residual maturities ([21]). Nevertheless, in most practical situations this assumption is deemed a satisfactory approximation and is commonly used ([48]).

The functional form of the NS model (and the NSS model by extension) is linear in $\boldsymbol{\beta}$, but nonlinear in τ. Therefore, (4.1) corresponds to the least squares parameter estimate of a nonlinear regression model. The solution can only be obtained using numerical optimization. In practice, variants of quasi-Newton methods are used widely across Eurosystem central banks ([48]). The primary downside of such methods is that they, in their standard implementation, fail to control for the possibility of ill-conditioning that occurs for certain parameter ranges ([30]). This can result in pronounced parameter sensitivity, by which small perturbations in the data lead to disproportionate effects on parameter estimates. Instead of jointly optimizing over both τ and $\boldsymbol{\beta}$, parameters can be obtained through nested optimization

$$\min_{\tau} \min_{\beta} \sum_{i=1}^{N} (y(t_i; \boldsymbol{\beta}, \tau) - y_i)^2. \tag{4.2}$$

In the formulation, the inner minimization problem can be solved via OLS regression, while τ can be found numerically. **The grid search** approach involves evaluating the objective function for manually chosen values of τ, ultimately selecting the one with the most favorable value. The values of τ are typically evaluated for equally spaced points that form a finite grid.[1] The grid search can effectively find a good optimum, but often at a high computational cost: as an example, performing grid search on the NSS model to find the optimal $\hat{\tau} = (\hat{\tau}_1, \hat{\tau}_2)$ would require solving m^2 ordinary least squares optimization problems, where m is the number of points to be evaluated for each component of τ parameter.[2] In addition to being computationally demanding, this approach may often lead to erratic changes in estimated parameters ([3]).

The second approach involves linearizing the objective function by fixing the nonlinear shape parameter τ to a prescribed value. This reduces (4.1) to OLS optimization

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{N} (\boldsymbol{\beta}' \boldsymbol{x}(t_i) - y_i)^2, \tag{4.3}$$

where $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_k)'$ is the vector of parameters and $\boldsymbol{x}(t_i) = (x_0(t_i), \ldots, x_k(t_i))'$ is the vector of model basis functions evaluated at t_i.[3] While K and the form of individual basis functions $x_k(t)$ will depend on the underlying model, the form of the optimization objective (4.3) remains the same. This formulation implies a linear projection model

$$\begin{aligned} Y &= \boldsymbol{\beta}' \boldsymbol{x}(T) + e, \\ \mathbb{E}[\boldsymbol{x}(T) e] &= \mathbf{0}, \\ \boldsymbol{\beta} &= (\mathbb{E}[\boldsymbol{x}(T)\boldsymbol{x}(T)'])^{-1} (\mathbb{E}[\boldsymbol{x}(T)\boldsymbol{x}(T)']), \end{aligned} \tag{4.4}$$

where T, e are random variables reflecting residual maturity and error, respectively.

Although it inevitably comes at the expense of in-sample fit, the fixed τ approach has several benefits. First, eliminating nonlinearity from the functional form enables obtaining a simple closed-form expression for the globally optimal $\boldsymbol{\beta}$ (for a given τ). It also enables one to justify the use of linear methods to compute statistics such as confidence intervals and leverage (which can be used in outlier detection). Second, constant τ ensures that basis curves remain the same, which ensures both inter-temporal and cross-sectional comparability of parameters. Inter-temporal comparability is particularly relevant in case yield curves are estimated across multiple periods and analyzed in a time series context. Cross-sectional comparability is required if yield curve parameters associated with different issuers are to be meaningfully compared against each other or, perhaps, aggregated.[4] Third, allowing τ to vary may lead to increased parameter instability

[1] If τ is a scalar, equally spaced points in a finite interval are evaluated; if it is a vector, a Cartesian product of such intervals for each of the components can be used.

[2] Evaluating all multiples of 0.5 in the interval $[0, 30]$ would require $(30/0.5)^2 = 3600$ such evaluations.

[3] The model basis function $\boldsymbol{x} : \mathbb{R} \to \mathbb{R}^{K \times 1}$ is a (deterministic) vector function of time T, which is assumed to be a random variable.

[4] Aggregation of multiple yield curves may be relevant, for example, in the problem of obtaining a representative yield curve of the euro area.

([3]), reinforcing the issue of reduced comparability. Finally, the reduction of in-sample fit brought on by fixing τ may not be significant in practice ([46]), which is also demonstrated in Chapter 5. However, when fixing τ, it is hardly possible to conclusively justify any particular choice, so determining its value remains as the primary challenge. Along the lines of the previously discussed interpretation of τ, higher values typically result in more curve flexibility in the longer end of the curve, while the converse holds for lower values. Annaert, De Ceuster, and Zhang observed that when performing grid search over $[0, 10]$ for the NS model, in the majority of days between 1999 and 2009, the optimal value of τ was in the interval $[0, 4]$. They also observed increased ill-conditioning coupled with parameter instability for higher values of τ ([3]). Diebold and Li fixed $\tau = 1.37$ ([20])[5], while Fabozzi, Martellini and Priaulet used $\tau = 3$ ([26]).

4.1 Estimation in the presence of fixed-coupon bonds

As mentioned previously, in many practical situations, a sufficiently sized sample of zero-coupon bonds may not be available. Even in cases when it is, one may still prefer to use a larger sample (one that is not restricted only to zero-coupon bonds) to obtain a more reliable estimate.

In addition to the principal paid at maturity, fixed-coupon bonds pay a predefined (fixed) cash amount, **coupon**, at predefined points in time before maturity. The fair price of a risk-free bond at the time t with a principal value of 1 and a fixed coupon payable annually, is given by

$$p(t, c) = \exp\left(-ty(t)\right) + c \sum_{s=1}^{t} \exp\left(-sy(s)\right), \tag{4.5}$$

where t is the residual maturity, c is the coupon rate, and $y(t)$ is the continuously compounded annual interest rate at maturity t. In the special case of a zero-coupon bond (when $c = 0$), the second term vanishes and the expression reduces to (3.1). If the bond is coupon-bearing, it is generally impossible to solve for yield $y(t)$ in (4.5). Intuitively, this is because the bond's fair price also depends on unknown zero-coupon yields $y(1), \ldots, y(t-1)$.

Yield-to-maturity (or **redemption yield**), which can also be computed for fixed-coupon bonds, is the bond's **internal rate of return**. More precisely, it is defined as the discount rate y_{TM} at which the present value of the bond's future cash flows (i.e., principal and coupon payments) is equal to the bond's price p. Formally, it solves the equation

$$p = \exp\left(-ty_{TM}\right) + c \sum_{s=1}^{t} \exp\left(-sy_{TM}\right), \tag{4.6}$$

for a given price p, residual maturity t, and a coupon rate c. Since a closed-form expression for y_{TM} does not exist, the equation can only be solved numerically. If

[5] The original paper uses $\lambda = 0.0609$, which is used to multiply (rather than divide) the residual maturity expressed in months (rather than years). Therefore, the equivalent τ can be obtained with $\tau = (12\lambda)^{-1} = 1.37$.

the coupon rate is equal to zero, the expression reduces to the standard closed-form expression for the zero-coupon bond price.

The implicit simplifying assumption in (4.6) is that the yield curve is flat, i.e. that $y(1) = \ldots = y(t) = y_{TM}$. Yield-to-maturity of a bond with a price p can, therefore, be described as the level y_{TM} of a hypothetical flat yield curve under which the bond's fair price is equal to its actual price p. This impairs the comparability of yield-to-maturity with zero-coupon yields: if the true zero-coupon yield curve used to price the bond is not flat, its yield-to-maturity will not coincide with the zero-coupon yield at maturity t. Intuitively, this is because yield-to-maturity also depends on zero-coupon yields before maturity t.

To illustrate the lack of comparability, assume that the true zero-coupon yield curve is known and that it is strictly increasing. When pricing a fixed-coupon bond under such a yield curve, coupon payments maturing earlier will be discounted with a lower rate of interest compared to their counterparts maturing later. Since yield-to-maturity is the single discount rate at which all coupons are discounted, it would need to account for the lower discount rate on coupons received before maturity. Consequently, its yield-to-maturity will be lower than the zero-coupon yield associated with the bond's residual maturity t. The more the bond's coupons contribute to its value, the more skewed the yield-to-maturity will be toward the zero-coupon yields before maturity.[6] Conversely, if the yield curve is strictly decreasing, yield-to-maturity will be higher than the zero-coupon yield for bonds with identical maturity. With simple algebraic manipulations, it can be shown that if the investor acquires a fixed-coupon bond and reinvests all received coupons at a rate equal to y_{TM}, the average annual yield, calculated from the accumulated amount at maturity t, would equal y_{TM}. This property of yield-to-maturity (which can also be used as its definition) better illustrates why yields-to-maturity are not directly comparable to equivalent zero-coupon yields. Nevertheless, when the sample contains fixed-coupon bonds, comparing observed yields-to-maturity to those implied by the estimated zero-coupon yield curve provides a useful goodness-of-fit measure.

Zero-coupon yield curve can be estimated from a sample containing coupon-bearing bonds by minimizing squared yield-to-maturity errors. Assuming fixed τ, obtaining optimal $\hat{\boldsymbol{\beta}}$ would require solving a modified version of (4.1)

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{N} (\hat{y}_{TM_i}(\boldsymbol{\beta}) - y_{TM_i})^2, \tag{4.7}$$

where $\hat{y}_{TM_i}(\boldsymbol{\beta})$ is the yield-to-maturity of the i-th bond in the sample, each computed by discounting with a zero-coupon yield curve $y(t;\boldsymbol{\beta})$. One should note that, unlike in (4.1), the sample contains fixed-coupon bonds whose yield-to-maturity needs to be computed numerically by solving (4.6).[7] Solving (4.7) is, therefore, computationally demanding, as the computation needs to be repeated for all bonds in each step of the iterative optimization procedure.

[6] This is known as the **coupon effect** ([12]).

[7] In practice, the computation of yield-to-maturity function is more nuanced. Factors such as accrued interest, the duration of settlement, and day count conventions (which may differ between bonds) need to be accounted for.

A more simple formulation is to minimize price error instead of yield error

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{N} (\widehat{p}_i(\boldsymbol{\beta}) - p_i)^2. \tag{4.8}$$

This approach eliminates the need to compute the yields-to-maturity numerically but does not render the problem linear since the price of a fixed-coupon bond (4.5) depends nonlinearly on the zero-coupon yield curve and, therefore, $\boldsymbol{\beta}$. The model implied by (4.8) is

$$\begin{aligned} P &= P(T; \boldsymbol{\beta}) + e, \\ \mathbb{E}[e|T] &= 0. \end{aligned} \tag{4.9}$$

The main issue with this model formulation is that price errors are heteroscedastic, as the error variance increases for bonds maturing farther in the future ([9]). This is why price error minimization has been shown to result in suboptimal estimates of yield-to-maturity for short-term bonds ([48]). Weighted least squares can be used to offset the increasing variance and reduce the bias to longer-term bonds. Most commonly, **inverse bond durations** are used as weights ([48]). In practice, (4.7) and (4.8) are typically solved with an iterative procedure based on the gradient and (quasi-) Newton methods ([48]).

As an alternative to minimizing yield-to-maturity or price error, one may look for ways to transform the sample to obtain the underlying zero-coupon bond prices. This would reduce the fitting problem to OLS. The standard approach for doing so is **the bootstrapping method** ([25]). The technique constructs zero-coupon yields iteratively, starting with the shortest and progressing toward bonds with longer maturities. Typically, it is assumed that the zero-coupon yields up to a certain (short) maturity $m^{(i)}$ are known. These zero-coupon yields can be derived from short-term debt securities, which are usually zero-coupon. The zero-coupon yield $y^{(i+1)}$ in the next iteration at longer maturities is then obtained by equating the $(i+1)$-st bond's price $p^{(i+1)}$ with its present value and solving for its (zero-coupon) yield $y^{(i+1)}$

$$p^{(i+1)} = \exp\left(-y^{(i+1)} m^{(i+1)}\right) + \sum_{j=1}^{i} c \exp\left(-y^{(j)} m^{(j)}\right). \tag{4.10}$$

In case of sufficient availability of quality data, the bootstrapping technique produces internally consistent prices, in the sense that there are no opportunities for arbitrage within the sample ([25]). However, in many practical cases, the bond sample may not include bonds at the precise maturities required to accurately compute rates for discounting coupon payments. Furthermore, the prices of individual bonds can be affected by multiple nuisance effects, such as bond liquidity differences, measurement errors, and tax effects. These effects are propagated to subsequent iterations, resulting in increasingly inaccurate zero-coupon yield estimates. Alternatives based on linear programming have been proposed to address these issues ([1]).

4.2 Iterated OLS approach

In this chapter, we outline an alternative approach to solving (4.3) given a heterogeneous sample consisting of both zero- and fixed-coupon bonds. The detailed analysis is presented in [10].

As noted previously, unlike prices, bond yields are not observed directly. For this reason, the yield curve needs to be estimated from a sample of bonds, where for each bond b_i we know

- its market price, p_i
- its residual maturity t_i
- its coupon rate expressed as a percentage of the principal amount, c_i.

For simplicity, this configuration assumes that all coupons are paid at times $1, \ldots, t_i$. The analysis can be easily generalized to account for different coupon frequencies and accrued interest (if the bond is valued in between coupon payments). The final coupon is usually paid with the unit principal and therefore has residual maturity equal to t_i. Formally, the yield curve is estimated from a sample of bonds $\{b_1, \ldots, b_N\}$, with $b_i = (p_i, t_i, c_i)$. In the case of a yield of a zero-coupon bond, y_i is uniquely determined by its price p_i and residual maturity t_i via (3.2). This is not the case for fixed-coupon bonds, for which *yield* has no natural definition as discussed in Section 4.1.

In the iterated OLS method, this issue is addressed with **coupon stripping**.[8] Coupon stripping is a method to transform the prices of coupon-bearing bonds to obtain their equivalent zero-coupon prices, i.e. their hypothetical prices if they weren't paying a coupon. The **coupon-stripped price** can be obtained by rewriting (4.5)

$$p(t,c) = \exp\left(-ty(t)\right) + c\sum_{s=1}^{t} \exp\left(-sy(s)\right) = p(t,0) + c\sum_{s=1}^{t} \exp\left(-sy(s)\right), \quad (4.11)$$

and solving for the zero-coupon component $p(t,0)$. This gives the expression for the bond's coupon-stripped price as price less the present value of its coupon payments

$$p(t,0) = p(t,c) - c\sum_{s=1}^{t} \exp(-sy(s)). \quad (4.12)$$

Since the values of coupon-stripped prices depend on the yield curve $y(t;\boldsymbol{\beta})$ used to discount the coupon payments, coupon-stripped prices can be written as functions of $\boldsymbol{\beta}$

$$p(t,0;\boldsymbol{\beta}) = p(t,c) - c\sum_{s=1}^{t} \exp(-sy(s;\boldsymbol{\beta})). \quad (4.13)$$

[8] While in this context, "coupon stripping" refers to the estimation of the present value of the principal payment (i.e., the zero-coupon component of the bond), the term can also refer to the act of splitting the bond into two separately tradeable securities, each representing its coupon and principal. These securities are called STRIPS (Separate Trading of Registered Principal and Interest Securities).

Since the true $\boldsymbol{\beta}^*$ is not known, an approximation $\boldsymbol{\beta}^{(0)}$ can be used to estimate the coupon-stripped price. After applying (4.13) to replace all coupon-bearing bonds with their coupon-stripped equivalents, the resulting sample of coupon-stripped prices can be transformed via (3.1) to consist exclusively of zero-coupon yields. The updated estimate $\boldsymbol{\beta}^{(1)}$ can then be obtained by solving (4.3). Provided that convergence criteria are fulfilled, repeating this process results in successively improving estimates of zero-coupon prices (yields) and, by extension, yield curves $y(t;\boldsymbol{\beta})$ that are estimated from these samples. In summary, the procedure can be viewed as a series of two-step iterations:

1. first, generate a sample of zero-coupon prices (yields) by stripping coupons from fixed-coupon bonds using the current yield curve estimate $y(t;\boldsymbol{\beta}^{(i)})$;
2. second, estimate the subsequent zero-coupon yield curve estimate $y(t;\boldsymbol{\beta}^{(i+1)})$ from the generated zero-coupon bond sample.

The first step requires an initial approximation $\boldsymbol{\beta}^{(0)}$. If estimations are performed sequentially, the previous period's estimated yield curve can be used as the initial $\boldsymbol{\beta}^{(0)}$. However, as discussed in Section 4.2.1 and exemplified by the empirical study, the convergence of the method does not appear to be very sensitive to the choice of the initial approximation.

A similar approach that combines bootstrapping with estimation in a single procedure was presented and applied to an interpolation model in [32].

To present a more formal derivation and establish convergence criteria, we show that the described procedure can be implemented as the **fixed point iteration method** for solving a system of equations.

4.2.1 Convergence

The fixed point iteration method attempts to solve a system of equations of the form

$$\boldsymbol{a} = \boldsymbol{G}(\boldsymbol{a}), \tag{4.14}$$

where $\boldsymbol{a} \in \mathbb{R}^n$ and $\boldsymbol{G} : \mathbb{R}^n \to \mathbb{R}^n$ is a nonlinear function. Starting from the initial estimate $\boldsymbol{a}^{(0)}$, the method produces a sequence of iterations $\boldsymbol{a}^{(i+1)} = \boldsymbol{G}(\boldsymbol{a}^{(i)})$ which is hoped to converge to a **fixed point**. A fixed point $\hat{\boldsymbol{a}}$ is defined as the solution to (4.14)

$$\hat{\boldsymbol{a}} = \boldsymbol{G}(\hat{\boldsymbol{a}}). \tag{4.15}$$

The existence of a (unique) solution is therefore equivalent to the existence of a (unique) fixed point. The fixed point iteration method converges under the conditions of the Banach fixed-point theorem, which requires that $G : D \to D$ is a **contraction**, i.e. that there exists a $q \in [0,1)$ such that

$$d(\boldsymbol{G}(\boldsymbol{a}), \boldsymbol{G}(\boldsymbol{b})) \leq q d(\boldsymbol{a}, \boldsymbol{b}), \tag{4.16}$$

for some metric $d : D \times D \to [0, \infty)$. When the domain is $D \subseteq \mathbb{R}^n$, an equivalent condition on the L_2 matrix norm of the Jacobian $\dot{\boldsymbol{G}}$ can be used to verify the contraction property

$$\|\dot{\boldsymbol{G}}(\boldsymbol{\beta})\|_2 \leq q. \tag{4.17}$$

For $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, the L_2 matrix norm $\|\cdot\|_2 : \mathbb{R}^{n\times n} \to \mathbb{R}$ is defined as

$$\|\boldsymbol{A}\|_2 = \sup_{\|\boldsymbol{y}\|_2=1} \|\boldsymbol{A}\boldsymbol{y}\|_2. \tag{4.18}$$

If (4.17) holds, $\boldsymbol{G}$ is a contraction with respect to the L_2 metric $d(\boldsymbol{a}, \boldsymbol{b}) = \|\boldsymbol{a} - \boldsymbol{b}\|_2$.

For a sample of zero-coupon bonds $b_i = (p_i, t_i)$, the sum of squared errors from (4.3) can be equivalently rewritten using zero-coupon bond prices instead of yields

$$\begin{aligned} F(\boldsymbol{\beta}) &= \sum_{i=1}^{N} \left(\boldsymbol{\beta}'\boldsymbol{x}(t_i) - \frac{1}{t_i} \ln p_i^{-1} \right)^2 \\ &= \sum_{i=1}^{N} \left(\boldsymbol{\beta}'\boldsymbol{x}(t_i) + \frac{1}{t_i} \ln p_i \right)^2, \end{aligned} \tag{4.19}$$

where $x : \mathbb{R}_0^+ \to \mathbb{R}^{K\times 1}$, $\boldsymbol{x}(t) = (x_0(t), \ldots, x_K(t))'$ maps residual maturity to model factors. If bonds in the sample are assumed to pay a fixed coupon c_i,[9] the zero-coupon yield curve can be estimated from coupon-stripped prices.

$$F(\boldsymbol{\beta}) = \sum_{i=1}^{N} \left(\boldsymbol{\beta}'\boldsymbol{x}(t_i) + \frac{1}{t_i} \ln \left(p_i - c_i \sum_{s=1}^{t_i} \exp\left(-s\boldsymbol{\beta}'\boldsymbol{x}(s)\right)\right) \right)^2. \tag{4.20}$$

Intuitively, minimizing $F(\boldsymbol{\beta})$ corresponds to finding a curve of the form $y(t; \hat{\boldsymbol{\beta}}) = \hat{\boldsymbol{\beta}}'\boldsymbol{x}(t)$ which, if used to strip the coupons from fixed-coupon bonds, best fits (in the least-squares sense) the zero-coupon yields derived from the coupon-stripped prices. The minimum of $F(\boldsymbol{\beta})$ does not have a closed-form expression.

We can allow the coupons to be stripped by a possibly different zero-coupon yield curve $y(t; \boldsymbol{\beta}^{(0)})$

$$F(\boldsymbol{\beta}, \boldsymbol{\beta}^{(0)}) = \sum_{i=1}^{N} \left(\boldsymbol{\beta}'\boldsymbol{x}(t_i) + \frac{1}{t_i} \ln \left(p_i - c_i \sum_{s=1}^{t_i} \exp\left(-s\boldsymbol{\beta}^{(0)\prime}\boldsymbol{x}(s)\right)\right) \right)^2. \tag{4.21}$$

If we fix $\boldsymbol{\beta}^{(0)}$, the expression $D_{\boldsymbol{\beta}^{(0)}}(\boldsymbol{\beta}) = F(\boldsymbol{\beta}, \boldsymbol{\beta}^{(0)})$ becomes a sum of squared residuals of a linear model, allowing us to obtain a closed-form expression for its global minimum using normal equations

$$0 = 2\sum_{i=1}^{N} \boldsymbol{x}(t_i) \left(\boldsymbol{x}(t_i)'\hat{\boldsymbol{\beta}} + \frac{1}{t_i} \ln \left(p_i - c_i \sum_{s=1}^{t_i} \exp\left(-s\boldsymbol{\beta}^{(0)\prime}\boldsymbol{x}(s)\right)\right)\right) \tag{4.22}$$

$$\hat{\boldsymbol{\beta}} = -\left(\sum_{i=1}^{N} \boldsymbol{x}(t_i)\boldsymbol{x}(t_i)'\right)^{-1} \sum_{i=1}^{N} \boldsymbol{x}(t_i) \frac{1}{t_i} \ln \left(p_i - c_i \sum_{s=1}^{t_i} \exp\left(-s\boldsymbol{\beta}^{(0)\prime}\boldsymbol{x}(s)\right)\right) \tag{4.23}$$

[9] In this case, bonds are assumed to be represented by triplets $b_i = (p_i, t_i, c_i)$

By setting $\boldsymbol{\beta}^{(1)} = \hat{\boldsymbol{\beta}}$ and generalizing for arbitrary k, an iterative procedure can be defined

$$\boldsymbol{\beta}^{(k+1)} = \boldsymbol{G_N}(\boldsymbol{\beta}^{(k)}), \tag{4.24}$$

where

$$\boldsymbol{G_N}(\boldsymbol{\beta}) = -\left(\sum_{i=1}^{N} \boldsymbol{x}(t_i)\boldsymbol{x}(t_i)'\right)^{-1} \sum_{i=1}^{N} \boldsymbol{x}(t_i)\frac{1}{t_i}\ln\left(p_i - c_i\sum_{s=1}^{t_i}\exp\left(-s\boldsymbol{\beta}'\boldsymbol{x}(s)\right)\right). \tag{4.25}$$

Provided that the necessary conditions hold for $\boldsymbol{G_N}$, Banach fixed-point theorem implies that the series $\left(\boldsymbol{\beta}^{(k)}\right)$ converges to a unique fixed point

$$\hat{\boldsymbol{\beta}} = \lim_{k\to\infty} \boldsymbol{\beta}^{(k)} = \lim_{k\to\infty} \boldsymbol{G_N}\left(\boldsymbol{\beta}^{(k-1)}\right) = \boldsymbol{G_N}\left(\lim_{k\to\infty} \boldsymbol{\beta}^{(k-1)}\right) = \boldsymbol{G_N}(\hat{\boldsymbol{\beta}}). \tag{4.26}$$

The prices (and yields) of zero-coupon bonds are unaffected by stripping. Therefore, if the sample consists of purely zero-coupon bonds, the fixed point is found trivially and is equivalent to OLS. When the sample includes fixed-coupon bonds, it is reasonable to assume that in almost all practical scenarios, (4.25) will be a contraction and $\left(\boldsymbol{\beta}^{(k)}\right)$ will converge to a unique fixed point. A heuristic argument of why this is the case is based on the observation that coupons typically contribute to a fraction of the value of fixed-coupon bonds *on average*. Since parameter $\boldsymbol{\beta}$ to the function $\boldsymbol{G_N}$ determines the yield curve used to strip the coupons, a change in its value can induce only a comparatively smaller change in the optimal parameters $\boldsymbol{G_N}(\boldsymbol{\beta})$ that are estimated from the coupon-stripped sample. The presence of zero-coupon bonds, whose yields are unaffected by the parameter $\boldsymbol{\beta}$, would further diminish the impact of the parameter on the value of $\boldsymbol{G_N}(\boldsymbol{\beta})$.

To illustrate the argument, we examine a simple case requiring three assumptions:

1. First, the yield curve is modeled with an intercept-only model $y(t;\beta) = \beta$, $\beta \in \mathbb{R}$, i.e. a horizontal line. This implies that the polynomial basis is a scalar $x(t) = 1$, from which it follows that

$$\left(\sum_{i=1}^{N} x(t_i)x(t_i)'\right)^{-1} = \frac{1}{N}. \tag{4.27}$$

2. Second, we assume that no bond pays the final coupon at maturity (without loss of generality).[10] Formally, each bond in the sample with maturity t_i pays its coupons at times $1, \ldots, t_i - 1$.

[10] Although it is typical for bonds to pay the final coupon together with the principal payment at maturity, this simplifying assumption does not affect the calculations below, as discussed in Remark 3.

3. Finally, we assume that there exists an interval $I = (a, b)$ such that, for each bond in the sample, the present value of its coupons valued under $y(t; \beta) = \beta, \beta \in I$ is no more than $k_i \geq 0$ times the bond's price

$$c_i \sum_{s=1}^{t_i-1} \exp(-s\beta) \leq k_i p_i \implies \frac{1}{k_i} c_i \sum_{s=1}^{t_i-1} \exp(-s\beta) \leq p_i, \;\; k_i \in [0, 1) \tag{4.28}$$

Applying the triangle inequality, from (4.27) and (4.28), we have

$$\begin{aligned} \left|\frac{d}{d\beta} G_N(\beta)\right| &= \frac{1}{N} \left| \sum_{i=1}^{N} \frac{c_i \sum_{s=1}^{t_i-1} s \exp(-s\beta)}{t_i \left(p_i - c_i \sum_{s=1}^{t_i-1} \exp(-s\beta)\right)} \right| \\ &\leq \frac{1}{N} \sum_{i=1}^{N} \left| \frac{c_i \sum_{s=1}^{t_i-1} s \exp(-s\beta)}{(k^{-1} - 1) t_i c_i \sum_{s=1}^{t_i-1} \exp(-s\beta)} \right| \\ &= \frac{1}{N} \sum_{i=1}^{N} \left| \frac{k_i}{(1 - k_i) t_i} \frac{\sum_{s=1}^{t_i-1} s \exp(-s\beta)}{\sum_{s=1}^{t_i-1} \exp(-s\beta)} \right|. \end{aligned} \tag{4.29}$$

Expressions

$$\frac{\sum_{s=1}^{t_i-1} s \exp(-s\beta)}{\sum_{s=1}^{t_i-1} \exp(-s\beta)} \tag{4.30}$$

can be interpreted as weighted averages of coupon maturities, with weights $w(s) = \exp(-s\beta)$. (4.30) is also equivalent to the definition of bond duration, with the only difference being that, in this case, it is calculated exclusively from coupon payments. If the yield β is non-negative, $w(s)$ is a non-increasing function of maturity. Consequently, since lower maturities are weighted higher, the weighted average cannot exceed the simple arithmetic average

$$\frac{\sum_{s=1}^{t_i-1} s \exp(-s\beta)}{\sum_{s=1}^{t_i-1} \exp(-s\beta)} \leq \frac{1}{t_i - 1} \sum_{s=1}^{t_i-1} s = \frac{t_i(t_i - 1)}{2(t_i - 1)} = \frac{t_i}{2}. \tag{4.31}$$

Since all values are positive, the same inequality holds for absolute values. Combining (4.29) and (4.31), we have

$$\left|\frac{d}{d\beta} G_N(\beta)\right| \leq \frac{1}{N} \sum_{i=1}^{N} \left| \frac{k_i}{2(1 - k_i)} \right| = \frac{1}{N} \sum_{i=1}^{N} \frac{k_i}{2(1 - k_i)} = q \tag{4.32}$$

The right-hand side expression in some sense reflects the *average* contribution of coupons to bonds' prices in the sample. Lower average contributions translate to the lower sensitivity of the estimated interest rate $G_N(\beta)$ to the interest rate β under which coupons are stripped. For example, if we set an upper bound $k_i \leq q < \frac{2}{3}$, we have

$$\left|\frac{d}{d\beta} G_N(\beta)\right| \leq \frac{1}{N} \sum_{i=1}^{N} \frac{k_i}{2(1 - k_i)} \leq q < 1 \tag{4.33}$$

because $f(x) = \frac{x}{2(1-x)}$ is strictly increasing. The assumption $k_i \leq q < \frac{2}{3}$ would imply that the present value of coupons contributes to less than two-thirds of the total value of a bond. While this will almost certainly be true for bonds with short maturities (and trivially true for zero-coupon bonds), it may not necessarily hold for longer-dated bonds with high coupon rates. However, all bonds don't need to fulfill this requirement, but rather *on average*. It is reasonable to expect that this is indeed fulfilled in practice.

Remark 3. *If the final coupon payment is paid together with the principal (which is the general norm in practice), this can be treated as a single rather than two separate payments. Consequently, only coupons paid before maturity ought to be stripped. In that case, the final payment is assumed to equal* $1 + c$*, and coupons are assumed to be paid at times excluding the bond maturity* $1, \dots, t-1$*. Incorporating this into the expression for* G_N *(4.25), we get*

$$\boldsymbol{G_N}(\boldsymbol{\beta}) = -\left(\sum_{i=1}^{N} \boldsymbol{x}(t_i)\boldsymbol{x}(t_i)'\right)^{-1} \sum_{i=1}^{N} \boldsymbol{x}(t_i)\frac{1}{t_i}\ln\left(\frac{p_i - c_i \sum_{s=1}^{t_i-1} \exp\left(-s\boldsymbol{\beta}'\boldsymbol{x}(s)\right)}{1+c_i}\right) \tag{4.34}$$

$$= -\left(\sum_{i=1}^{N} \boldsymbol{x}(t_i)\boldsymbol{x}(t_i)'\right)^{-1} \sum_{i=1}^{N} \boldsymbol{x}(t_i)\frac{1}{t_i}\ln\left(p_i - c_i \sum_{s=1}^{t_i-1} \exp\left(-s\boldsymbol{\beta}'\boldsymbol{x}(s)\right)\right) \tag{4.35}$$

$$+\left(\sum_{i=1}^{N} \boldsymbol{x}(t_i)\boldsymbol{x}(t_i)'\right)^{-1} \sum_{i=1}^{N} \boldsymbol{x}(t_i)\frac{1}{t_i}\ln(1+c_i) \tag{4.36}$$

Since the second term (4.36) does not depend on $\boldsymbol{\beta}$*, the exclusion of the final coupon payment does not impact the computation of the Jacobian (4.29).*

The presented argument is heuristic and is not theoretically rigorous. Moreover, the bounds used are not the tightest possible and represent conservative approximations. The convergence is empirically demonstrated in Chapter 5, where the method is applied to estimate the NSS and OLP models on historical data.

The yield curve estimate obtained via this iterative procedure is internally consistent, in the no-arbitrage sense. By construction, the yield curve estimated via iterated OLS is the curve that best fits the zero-coupon yields of the coupon-stripped sample, with the coupon-stripping performed using the same optimal yield curve. It can be verified empirically that this property does not generally hold for yield curves obtained using (4.8) and (4.7). In the case of the bootstrapping procedure, the zero-coupon yield estimates of bonds within the sample are consistent, but not necessarily when out-of-sample zero-coupon yields are included. To illustrate a particular kind of inconsistency that the yield curve estimated via iterated OLS doesn't suffer from, consider a simple example where a yield curve $\hat{y}(t)$ is estimated from a sample of bonds observed without errors, using (4.7). If $\hat{y}(t)$ is used to strip the sample and the yield curve $\hat{y}_{stripped}(t)$ obtained from that coupon-stripped sample differs from $\hat{y}(t)$ at some residual maturity t^*, an arbitrage opportunity will emerge. An arbitrage portfolio would consist of a coupon bond with residual maturity t^* and short-sold coupon payments.[11] This portfolio would replicate a zero-coupon bond with residual maturity t^* but with a different yield than $\hat{y}(t^*)$. As a result, risk-free profit can be realized. In short, the arbitrage emerges as a consequence of the inconsistency between the rate at which coupons are discounted and the zero-coupon yields. The iterated OLS approach achieves this consistency by construction.

Finally, we note that this method is substantially less computationally complex to execute compared to (4.7), which requires the gradient and yields-to-maturity for the whole sample to be computed numerically at each step.

Covariance matrix and confidence intervals Without explicitly computing it, we can demonstrate that the k-th fixed point iteration $\boldsymbol{\beta}^{(k)}$ appears to have a reduced variance

$$\begin{aligned}\mathrm{Var}_{\boldsymbol{\beta}}[\boldsymbol{\beta}^{(k)} - \boldsymbol{\beta}] &= \mathrm{Var}_{\boldsymbol{\beta}}[\boldsymbol{G}_N(\boldsymbol{\beta}^{(k-1)}) - \boldsymbol{\beta}] \\ &\approx \mathrm{Var}_{\boldsymbol{\beta}}[\dot{\boldsymbol{G}}_N(\boldsymbol{\beta})(\boldsymbol{\beta}^{(k-1)} - \boldsymbol{\beta})] \\ &= \dot{\boldsymbol{G}}_N(\boldsymbol{\beta})(\boldsymbol{\beta}^{(k-1)} - \boldsymbol{\beta})\dot{\boldsymbol{G}}_N(\boldsymbol{\beta})'.\end{aligned} \tag{4.37}$$

Due to the contraction property of $\boldsymbol{G}_N(\boldsymbol{\beta})$, applying matrix norm submultiplicativity, we have the following bound for the norm of the covariance matrix

$$\left\|\mathrm{Var}_{\boldsymbol{\beta}}[\boldsymbol{\beta}^{(k)} - \boldsymbol{\beta}]\right\|_2 \leq \|\dot{\boldsymbol{G}}_N(\boldsymbol{\beta})\|_2 \left\|\boldsymbol{\beta}^{(k-1)} - \boldsymbol{\beta}\right\|_2 \|\dot{\boldsymbol{G}}_N(\boldsymbol{\beta})'\|_2 \leq q^2 \left\|\boldsymbol{\beta}^{(k-1)} - \boldsymbol{\beta}\right\|_2. \tag{4.38}$$

The L_2 matrix norm is equivalent to the spectral norm, defined as the largest eigenvalue $\|\boldsymbol{A}\|_2 = \sigma_{\max}(\boldsymbol{A})$. Based on this, the squared norm of a covariance matrix can be interpreted as the maximum variance explained by a single principal component. The relation (4.38) therefore implies that iterations have a variance-reducing effect in the L_2 sense. Estimates of the covariance matrix and the related statistics (e.g. regression intervals) may be considered conservative.

4.2.2 Estimating equations

The described iterated OLS approach may be analyzed using the concept of **estimating equations**. Estimating equation is defined as

$$g_N(\boldsymbol{\theta}) = \boldsymbol{0}, \tag{4.39}$$

where g_N is the **estimating function** that depends on the sample of size N and some statistical parameter $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p$. The dependence on the sample is suppressed from the notation for convenience. We can define an estimator $\hat{\boldsymbol{\theta}}_N$ as the solution to the equation (4.39). It is usual to assume that the estimating equation (4.39) is **unbiased** ([52]) in the sense that

$$\mathbb{E}_{\boldsymbol{\theta}}[g_N(\boldsymbol{\theta})] = \boldsymbol{0},\ \forall \boldsymbol{\theta} \in \Theta. \tag{4.40}$$

[11] Since each coupon payment can be treated as a zero-coupon bond, it can be short-sold for the interest rate implied by the zero-coupon yield curve $\hat{y}(t)$.

To reformulate the iterated OLS method in terms of estimating equations, (4.14) can be rewritten as

$$g_N(\boldsymbol{\beta}) \stackrel{\text{def}}{=} \boldsymbol{G}_N(\boldsymbol{\beta}) - \boldsymbol{\beta} = \boldsymbol{0}. \tag{4.41}$$

We now proceed to specify the model under which the estimating equation (4.41) is unbiased. Let the yield of a zero-coupon bond be generated by the model

$$\begin{aligned} Y &= \boldsymbol{\beta}'\boldsymbol{x}(T) + e, \\ \mathbb{E}[e|T] &= 0. \end{aligned} \tag{4.42}$$

By applying (3.1), the price of a zero-coupon bond is generated by

$$\begin{aligned} P &= \exp\left(-T(\boldsymbol{\beta}'\boldsymbol{x}(T) + e)\right), \\ \mathbb{E}[e|T] &= 0. \end{aligned} \tag{4.43}$$

If the bond has a fixed coupon with a (deterministic) coupon rate c, from (4.5), the corresponding model is

$$\begin{aligned} P &= \exp\left(-T(\boldsymbol{\beta}'\boldsymbol{x}(T) + e)\right) + c\sum_{s=1}^{T} \exp\left(-s\boldsymbol{\beta}'\boldsymbol{x}(s)\right), \\ \mathbb{E}[e|T] &= 0. \end{aligned} \tag{4.44}$$

Remark 4. *Given that fixed-coupon bonds are theoretically equivalent to portfolios of zero-coupon bonds, the price of a fixed-coupon bond in (4.44) could have included an error term in the yield at which coupon payments are valued*

$$P = \exp\left(-T(\boldsymbol{\beta}'\boldsymbol{x}(T) + e)\right) + c\sum_{s=1}^{T} \exp\left(-s(\boldsymbol{\beta}'\boldsymbol{x}(s) + e_s)\right). \tag{4.45}$$

This formulation would imply that the bond's principal payment and individual coupon payments are all stochastic, as they would carry a specific (idiosyncratic) error term. In contrast, the model (4.44) treats the present value of the principal (the zero-coupon component) as stochastic, and the present value of coupon payments (the coupon component) as deterministic.

If the bond prices are generated by (4.44), from the unbiasedness of the OLS estimator, we have

$$\begin{aligned} \mathbb{E}_{\boldsymbol{\beta}}[\boldsymbol{G}_N(\boldsymbol{\beta})] &= \mathbb{E}_{\boldsymbol{\beta}}\left[-\left(\sum_{i=1}^{N} \boldsymbol{x}(T_i)\boldsymbol{x}(T_i)'\right)^{-1} \sum_{i=1}^{N} \boldsymbol{x}(T_i)\frac{1}{T_i}\ln\left(P_i - c_i\sum_{s=1}^{T_i}\exp(-s\boldsymbol{\beta}'\boldsymbol{x}(s))\right)\right] \\ &= \mathbb{E}_{\boldsymbol{\beta}}\left[-\left(\sum_{i=1}^{N} \boldsymbol{x}(T_i)\boldsymbol{x}(T_i)'\right)^{-1} \sum_{i=1}^{N} \boldsymbol{x}(T_i)\frac{1}{T_i}\ln\left(\exp(-T_i(\boldsymbol{\beta}'\boldsymbol{x}(T_i) + e_i))\right)\right] \\ &= \mathbb{E}_{\boldsymbol{\beta}}\left[\left(\sum_{i=1}^{N} \boldsymbol{x}(T_i)\boldsymbol{x}(T_i)'\right)^{-1} \sum_{i=1}^{N} \boldsymbol{x}(T_i)(\boldsymbol{\beta}'\boldsymbol{x}(T_i) + e_i)\right] = \boldsymbol{\beta}. \end{aligned} \tag{4.46}$$

The second equality holds due to the coupon-stripping term canceling out the actual contribution of coupons to the bond's value. Equation (4.46) implies that the estimating equation (4.41) is unbiased

$$\mathbb{E}_{\boldsymbol{\beta}}[\boldsymbol{g}_N(\boldsymbol{\beta})] = \mathbb{E}_{\boldsymbol{\beta}}[\boldsymbol{G}_N(\boldsymbol{\beta})] - \boldsymbol{\beta} = \mathbf{0}, \ \forall \boldsymbol{\beta}. \tag{4.47}$$

Being a more general concept, the unbiasedness of the estimating equation does not imply the unbiasedness of the associated estimator $\hat{\boldsymbol{\theta}}_N$. However, under additional assumptions, consistency and asymptotic normality follow ([19], [38], [43]). Jacod and Sørensen ([36]) provide the following general result for the consistency of $\hat{\boldsymbol{\theta}}_N$.

Theorem 1. *Assume that the following conditions hold for a parameter value $\boldsymbol{\theta}^* \in \operatorname{int} \Theta$ (the interior of Θ), a neighborhood M of $\boldsymbol{\theta}^*$, and a (possibly random) $\mathbb{R}^p$-valued function $\boldsymbol{g}$ on M:*

1. *$\boldsymbol{g}_N(\boldsymbol{\theta}^*) \xrightarrow{\mathbb{P}} \mathbf{0}$, as $N \to \infty$ (convergence in probability under the true probability measure $\mathbb{P}$) and $\boldsymbol{g}(\boldsymbol{\theta}^*) = \mathbf{0}$.*

2. *$\boldsymbol{g}_N$ and $\boldsymbol{g}$ are $\mathbb{P}$-almost surely continuously differentiable on M, and as $N \to \infty$*

$$\sup_{\boldsymbol{\theta} \in M} \|\dot{\boldsymbol{g}}_N(\boldsymbol{\theta}) - \dot{\boldsymbol{g}}(\boldsymbol{\theta})\|_2 \xrightarrow{\mathbb{P}} 0. \tag{4.48}$$

3. *The Jacobian $\dot{\boldsymbol{g}}_N(\boldsymbol{\theta}^*)$ is $\mathbb{P}$-almost surely nonsingular.*

Then, the sequence of estimators $(\widehat{\boldsymbol{\theta}}_N)$, defined as roots to the estimating equation $\boldsymbol{g}_N(\boldsymbol{\theta}_N) = \mathbf{0}$, converges in probability to $\boldsymbol{\theta}^*$ as $N \to \infty$. Moreover, if $(\hat{\boldsymbol{\theta}}'_N)$ is any other sequence of $\boldsymbol{g}_N$-estimators that converges in probability to $\boldsymbol{\theta}^*$, then $\mathbb{P}(\hat{\boldsymbol{\theta}}'_N \neq \hat{\boldsymbol{\theta}}_N) \to 0$ as $N \to \infty$.

The Condition (2) can be verified under stricter but more easily verifiable assumptions using the result from [47] (Corollary 2.2 and the accompanying discussion).

If $\boldsymbol{g}_N$ satisfies the conditions from Theorem 1, the iterated OLS estimator $\hat{\boldsymbol{\beta}}_N$ is consistent. Under additional assumptions on $\boldsymbol{g}_N$, the associated estimator is asymptotically normal ([36]).

5 | Empirical study

This chapter aims to empirically complement the theoretical discussion of yield curve calibration provided in Chapter 3 and Chapter 4. While the empirical characteristics of NS-class yield curve models are mostly known, an empirical evaluation of the OLP model presented in Section 3.2 has, to my knowledge, not been conducted, despite the model's appealing theoretical properties. In light of this, the primary objective of this empirical study is to compare the performance of the OLP model to the widely used NSS model. The analysis focuses purely on comparing the adequacy of the respective models' functional forms in a raw setting. Due to its relevance for practical implementations of NS-class models, special attention is given to exploring the role of the exponential decay parameter τ in mediating the trade-off between flexibility (i.e., goodness of fit) and the stability of model parameters.

5.1 Data selection

The bond dataset used in the empirical study includes daily closing prices of selected German government bonds in the period between December 2013 and December 2023, provided by MTS Markets.[1] The model evaluation is conducted on a set of 130 trading days, evenly spaced throughout the 10-year time period under consideration. The choice of the German bond population is motivated by the general perception of German bonds as proxies for the euro area risk-free rate. A stable level of low credit risk also contributes to sample homogeneity across time. Two additional selection criteria are applied to the initial bond population:

- Only zero-coupon and fixed-coupon bonds are selected. Bonds with special features (e.g., inflation-linked bonds, variable-coupon bonds) are excluded.

- Given the standard practice of reporting yields up to 30 years, bonds with longer residual maturity are excluded. Bonds with residual maturity shorter than three months are likewise excluded due to their generally higher volatility ([48]).

After selection, the number of bonds on each of the 130 days in the dataset averages 65.78 and ranges between 48 and 78.

[1] mtsmarkets.com

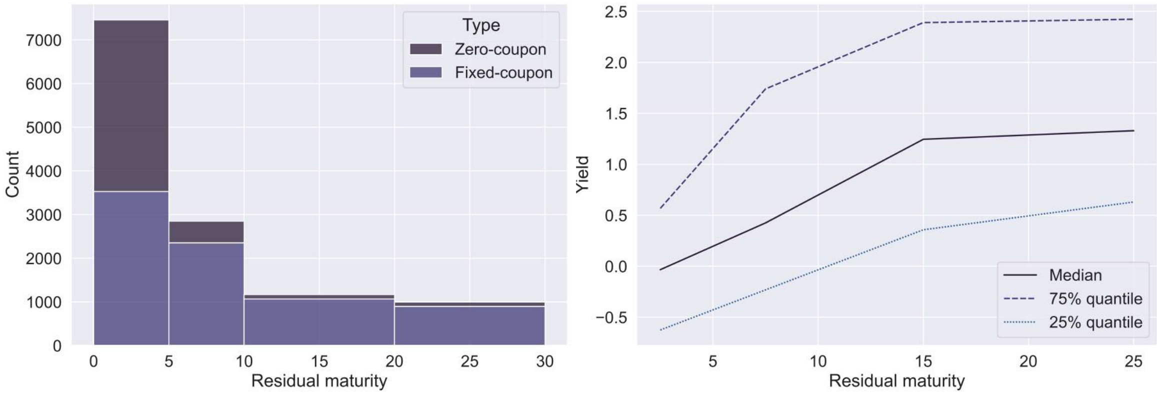

Figure 5.1: Descriptive statistics of the dataset: sample structure by residual maturity and coupon type (left); pointwise (data-based) estimates of the median yield curve with 25% and 75% quantiles (right)

5.2 Results

The performance of $OLP(4)$, $OLP(5)$, and the NSS model was analyzed. All models are estimated using the iterated OLS procedure presented in Section 4.2. The estimations are performed sequentially, where the parameters estimated in one period serve as the starting values for the iterated OLS in the next period.

The criteria against which models are evaluated are stability and flexibility.

- Stability relates to the reliability of parameter estimates and their dynamic properties. The possibility of erratic behavior of $NS(S)$ model parameters is a well-known property of these models and is generally deemed undesirable. Such behavior, particularly when exhibited by β_0, diminishes parameters' economic relevance and interpretability.

- Flexibility relates to models' ability to fit a wide range of possible yield curve shapes. The ideal level of flexibility achieves the right balance between goodness-of-fit and the capacity to generalize out of the sample.

5.2.1 Stability

Being the exponential decay parameter, τ controls how the curvature of basis curves is spread along the maturity axis: lower values of τ imply faster decay, resulting in curvature being concentrated at shorter maturities (Figure 5.2).

Remark 5. *In the case of the NSS model where $\tau = (\tau_1, \tau_2)$ is not scalar, relational qualifications such as "lower" and "higher" generally refer to either of the two individual components τ_1, τ_2.*

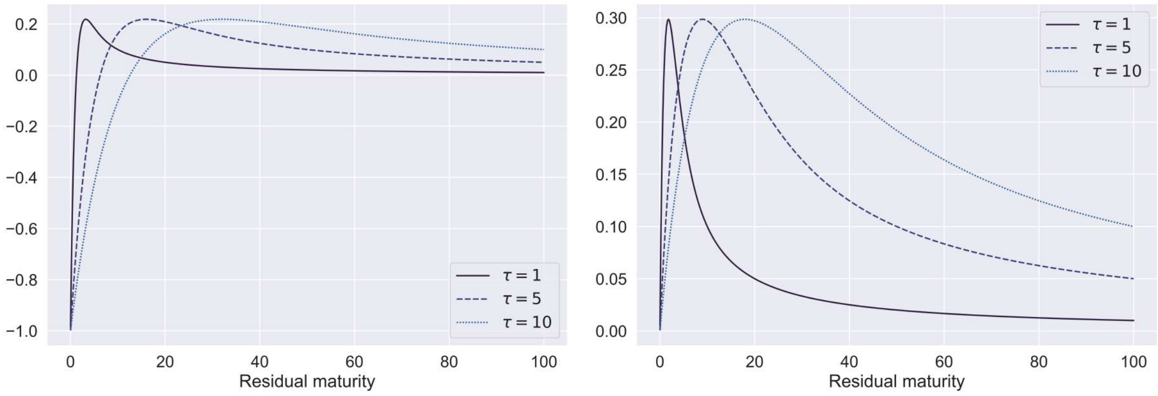

Figure 5.2: OLP fourth basis curve (left); NSS third/fourth basis curve (right)

Consequently, the role of τ can be understood as determining the trade-off between the flexibility to fit the short and the long end of the maturity spectrum (Figure 5.3). This can be directly observed by its effect on goodness of fit in dif-

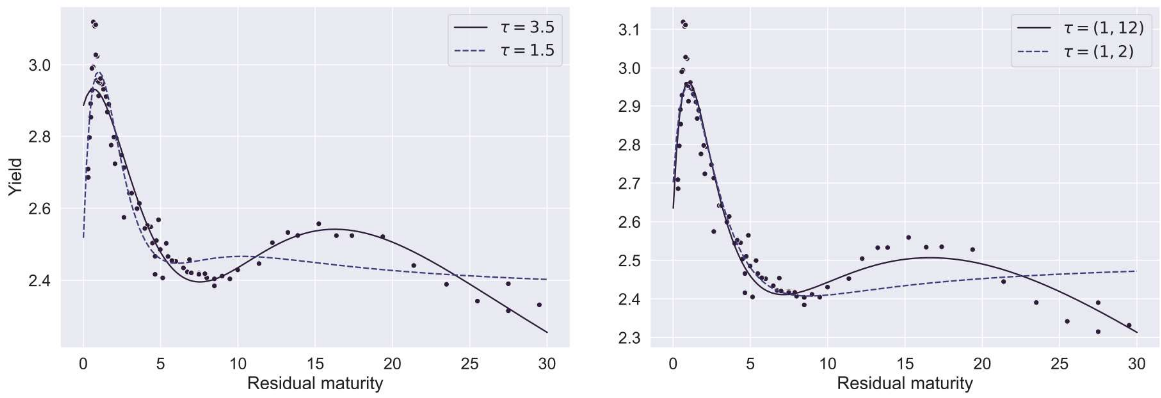

Figure 5.3: Optimal curve for different τ, $OLP(5)$ (left) and NSS (right)

ferent maturity ranges (Figure 5.4, Figure 5.5).

Perhaps unsurprisingly, τ also affects the inter-temporal stability of estimated $\boldsymbol{\beta}$ parameters. Stability is often conveniently gauged from the time series dynamics of β_0, which is the implied long-term (asymptotic) interest rate. For high values of τ, β_0 exhibits erratic behavior (Figure 5.6; left figure) that is unlikely, from an economic standpoint, to reflect the actual dynamics of long-term interest rates. How τ affects stability is closely related to its effect on the curvature of basis curves, which can extend beyond 30 (years) for high values of τ. However, no bonds with residual maturity longer than 30 years are included in the sample. This results in possible overfitting of the long end of the curve, leading to poor generalization capacity, particularly when it comes to extrapolating the yield curve to longer residual maturities. This is evidenced by unreasonable values of β_0, which represents the implied long-term yield prediction. The reasoning also generalizes to a setup that doesn't explicitly set a residual maturity limit, as the number of bonds with maturities longer than 30 years is zero or very low. The observed property can

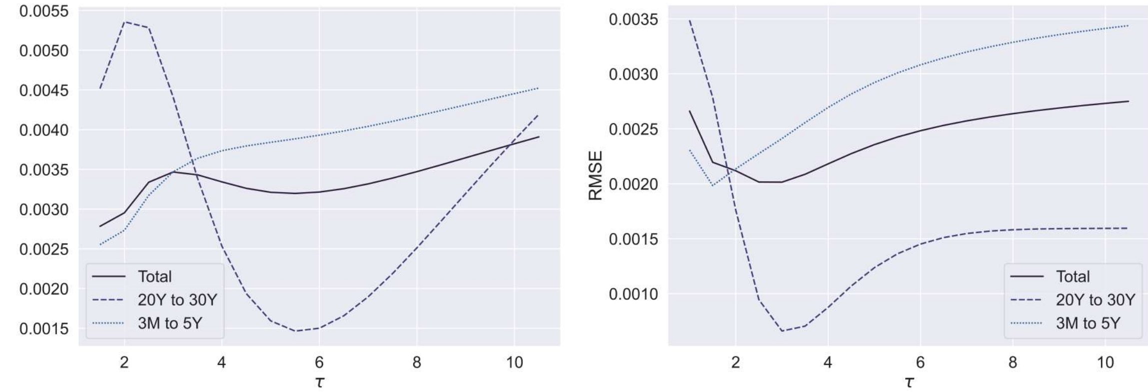

Figure 5.4: Average RMSE (whole dataset) for different τ, per maturity range, $OLP(4)$ (left) and $OLP(5)$ (right)

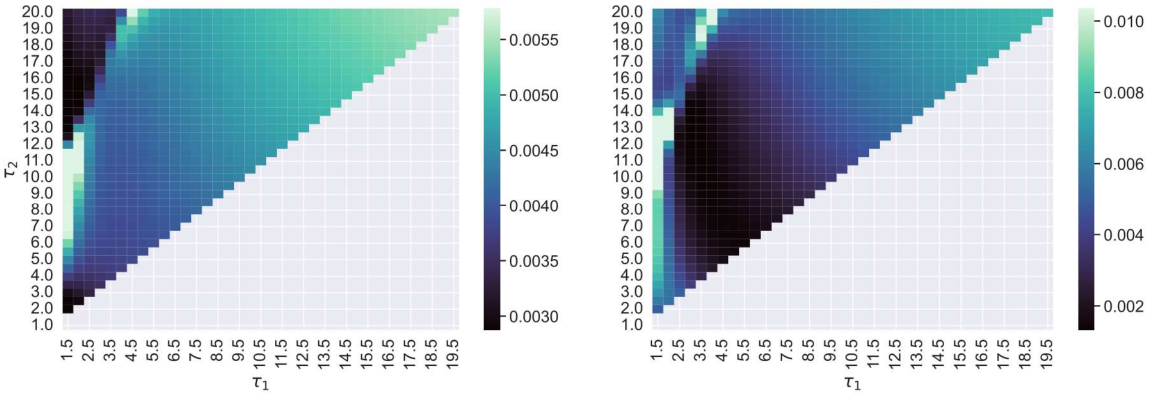

Figure 5.5: Average NSS RMSE, 3M to 5Y (left) and 20Y to 30Y (right) maturity ranges

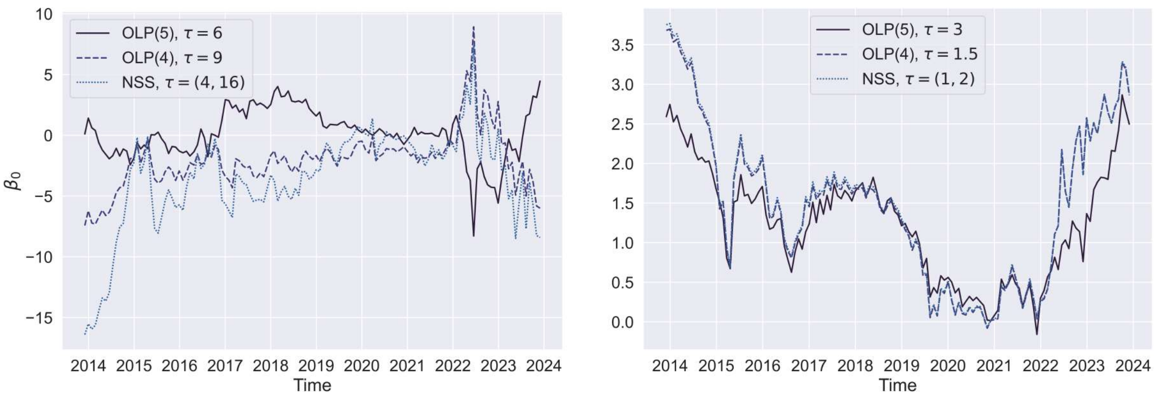

Figure 5.6: β_0 time series under different models, higher τ (left) and lower τ (right)

equivalently be seen as a consequence of ill-conditioning of the covariance matrix at higher values of τ (Figure 5.7).

This behavior can be, and often is, addressed by constraining model parameters or reducing the likelihood of unrealistic changes in their values. In the dynamic variants of $NS(S)$ models, this is achieved by assuming that parameters follow a vector autoregressive process under which high-magnitude jumps are assumed to be unlikely ([20]). Such an approach can be thought of as a form of regularization. However, since the primary purpose of this analysis is to evaluate the adequacy of different functional forms in fitting the yield curve, approaches that rely on any form of prior information are not considered. A final production model may well incorporate such, or similar features that further improve its empirical properties.

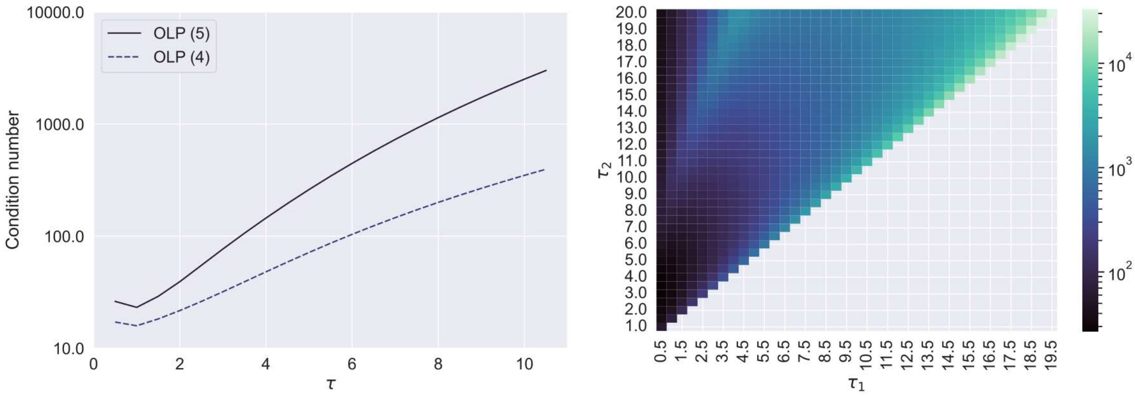

Figure 5.7: Average covariance matrix condition number for different τ, $OLP(4)$, $OLP(5)$ (left) and NSS (right)

Given the impact of τ, the methodology used to estimate it has a direct effect on the performance of the model. The dramatically different behavior of optimal $\boldsymbol{\beta}$ parameters implied by different values of τ underscores the nuance and care required in estimating it. A common approach is to use grid search to find a value that results in optimal goodness-of-fit. However, optimal values of τ may vary substantially between days, giving rise to intermittent jumps in β_0 (Figure 5.8; left figure).

Aside from directly constraining the dynamics of $\boldsymbol{\beta}$ with prior information, the stability issue can be mitigated by constraining τ not to assume values that lead to instability. If τ is allowed to vary but is constrained by $0 < \tau \leq 5$ ($0 < \tau_1, \tau_2 \leq 5$ in the case of NSS), the variability of β_0 is markedly decreased compared to when τ is searched within the full grid.[2] However, varying τ inevitably results in temporal inconsistency as $\boldsymbol{\beta}$ parameters estimated at different periods may no longer be associated with the same basis curves. This questions the applicability of time series methods.

Fixed τ configuration

If τ is fixed to a value in $0 < \tau \leq 5$ ($0 < \tau_1, \tau_2 \leq 5$ in the case of NSS) that results in the lowest average RMSE over the whole period, the variability of β_0 is

[2] The full grid includes values of τ (or both τ_1 and τ_2 on the case of NSS) that are within the range that is likely to contain the global optimum.

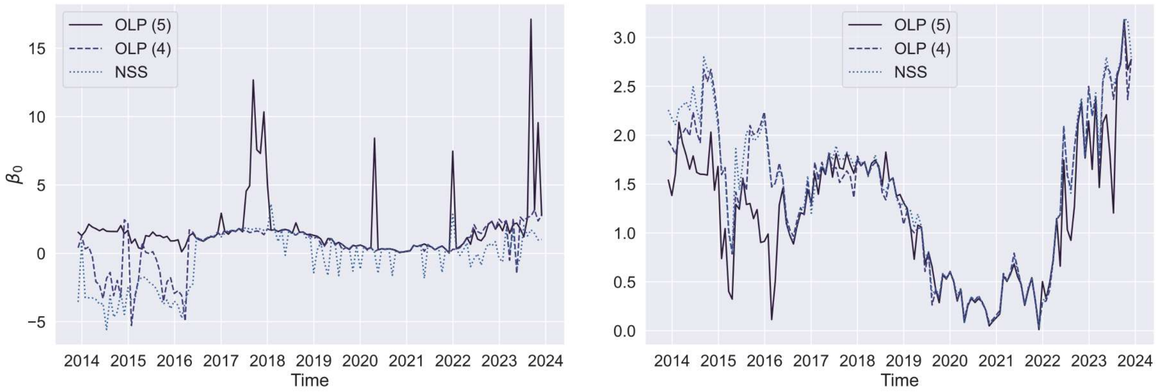

Figure 5.8: β_0 time series if τ is estimated using grid search, full grid search (left), and reduced grid search (right)

further decreased in comparison to when τ is constrained but allowed to vary between periods. In addition, β_0 assumes similar values across models (Figure 5.8; right figure), and the time series do not exhibit any apparent visual features that would contradict the expected temporal characteristics of long-term interest rates. The improved stability and generalization capacity can also be seen for the very short rates, with the fixed-τ models achieving closer alignment to the ECB deposit facility rate in the rate increase period between March 2022 and December 2023 (Figure 5.9). Since only bonds with residual maturity of no less than 3 months are included in the estimation sample, the short-term rate estimate can be seen as an out-of-sample prediction.

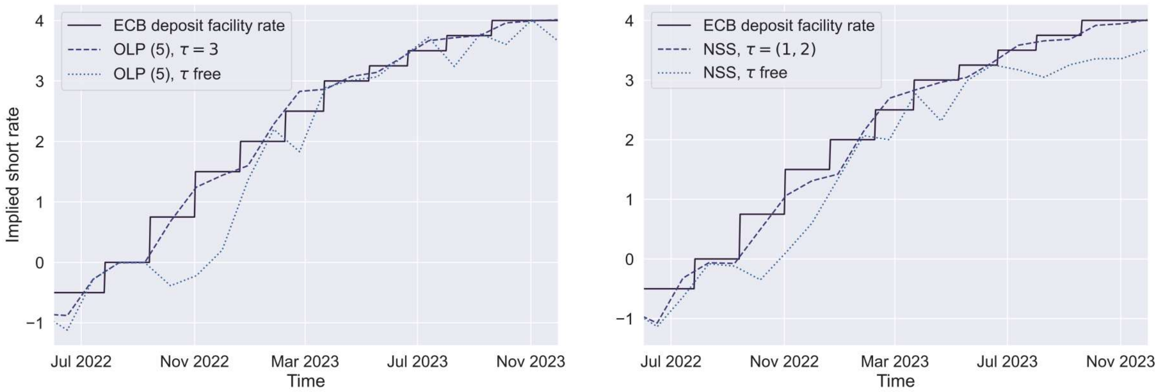

Figure 5.9: Model-implied short rate versus ECB deposit facility rate, $OLP(5)$ (left) and NSS (right)

5.2.2 Flexibility

Goodness-of-fit statistics numerically summarize the discrepancy between observed zero-coupon yields $y_1, \ldots, y_N$, and the corresponding model estimates

$\hat{y}(t_1), \ldots, \hat{y}(t_N)$. As part of the model flexibility evaluation, three statistics are considered:

- Root mean squared error

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}(t_i))^2}, \tag{5.1}$$

- Mean absolute error

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}(t_i)|, \tag{5.2}$$

- Hit rate, defined as the fraction of observed yields that deviate no more than 5 basis points (0.05 percentage points) from the corresponding estimates

$$Hit\ rate = \frac{1}{N}\sum_{i=1}^{N} I\left(y_i - \hat{y}(t_i)\right), \tag{5.3}$$

  where

$$I(x) = \begin{cases} 1, & -0.05 \leq x \leq 0.05, \\ 0, & \text{otherwise.} \end{cases} \tag{5.4}$$

Two grids in which τ is searched are considered: the full and the reduced grid.

- The full grid considers values that are likely to contain the globally optimal value of τ.
- The reduced grid considers values of τ that lead to a reasonable level of stability, as shown previously.

Model	Reduced grid	Full grid
$OLP(4)$	$0.5 \leq \tau \leq 5$	$0.5 \leq \tau \leq 20$
$OLP(5)$	$0.5 \leq \tau \leq 5$	$0.5 \leq \tau \leq 10$
NSS	$0.5 \leq \tau_1 < \tau_2 \leq 5$	$0.5 \leq \tau_1 < \tau_2 \leq 20$

Table 5.1: Grid search configuration

The additional constraint in the NSS model $\tau_1 < \tau_2$ is used to reduce the parameter search space, which would otherwise be very large. The constraint is natural given that the motivation to extend the NS model with an additional curvature

Model	RMSE	MAE	Hit rate
$OLP(5)$	0.033193	0.023835	88.8762%
$OLP(4)$	0.035853	0.026334	86.8876%
NSS	0.032863	0.023677	88.4831%

Table 5.2: Summary of error statistics, if τ is estimated with full grid search

term (with τ_2 as its exponential decay) was precisely to increase the flexibility of the long end of the curve. This constraint is also suggested in [50].

Maximal model flexibility is achieved if τ is estimated using full grid search, in which case the NSS model outperforms both OLP models in RMSE and MAE, while the $OLP(5)$ model has a superior hit rate (Table 5.2). However, the differences between $OLP(5)$ and NSS are not considerable. If the values of τ that contribute to instability are excluded (in the reduced grid), the $OLP(5)$ model achieves superior performance (Table 5.3) across all goodness-of-fit statistics. Moreover, the overall flexibility of $OLP(5)$ dropped only slightly under the reduced compared to the full grid search.

Model	RMSE	MAE	Hit rate
$OLP(5)$	0.033229	0.023935	88.7978%
$OLP(4)$	0.037833	0.027746	85.0411%
NSS	0.037409	0.027441	84.7082%

Table 5.3: Summary of error statistics, if τ is estimated with reduced grid search

Fixed τ configuration

Determining the value to which to fix τ is far from straightforward. The approach used in this empirical study involves finding a value from the reduced grid for which, if τ is fixed to that value, the average RMSE across the whole period is minimal. In this setup, the $OLP(5)$ outperforms by a wide margin (Table 5.4). While the performance is similar for *simpler* shapes, more *complex* shapes are not well captured by either $OLP(4)$ or NSS (Figure 5.10). The NSS model could fit the more complex shape if τ is free and unconstrained, but that would inevitably come at a cost of inter-temporal consistency and stability, as discussed previously. $OLP(5)$, on the other hand, can fit the complex shape reasonably well without sacrificing theoretical properties and introducing nonlinearity and nonconvexity. Moreover, fixing τ to a prescribed value constrains the set of possible yield curve

Model	Optimal τ fixing	RMSE	MAE	Hit rate
$OLP(5)$	$\tau = 3$	0.035213	0.025038	88.5952%
$OLP(4)$	$\tau = 1.5$	0.045322	0.034311	81.3514%
NSS	$\tau = (1, 2)$	0.044291	0.032983	82.2092%

Table 5.4: Summary of error statistics, if τ is fixed through the whole period

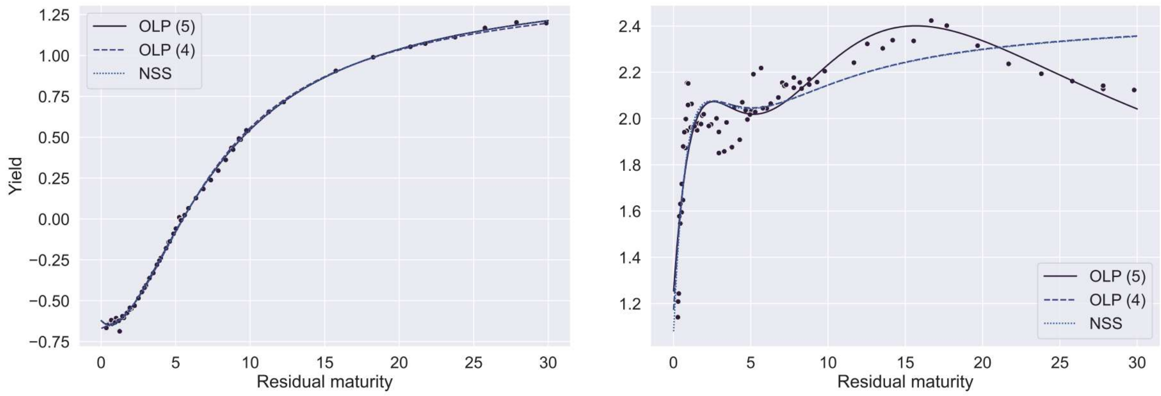

Figure 5.10: Fitting a simple (left) and a complex yield curve shape (right)

shapes. If the value is chosen carefully, the resulting model will be more robust and less prone to overfitting. Figure 5.12 illustrates that the lowest total RMSE may be achieved for curves that do not appear to capture the trend in the data evenly across the full maturity spectrum, compared to the fixed-τ setup.

Finally, fixing τ to a prescribed value has the unique benefit of allowing one to easily and more reliably generate confidence regions (Figure 5.11).

5.3 Conclusions

The analysis demonstrated the τ-mediated trade-off between the flexibility to fit the short versus the long end of the yield curve. Higher values of τ, which were shown to be associated with relatively better long-end fitting, were also shown to contribute to higher instability, particularly the intercept term β_0 or the implied long-term interest rate. Different approaches to estimating τ were explored to evaluate the nature of the trade-off between flexibility and stability. If τ is estimated with a full grid search, NSS and $OLP(5)$ demonstrated similar performance. However, erratic behavior of β_0 was observed in all models, the degree of which is not desirable in practice regardless of the goodness of fit. If τ is searched in a reduced grid with values resulting in reasonable levels of stability, $OLP(5)$ demonstrated superior performance across all statistics. Under the fixed-τ configuration, the $OLP(5)$ model was superior, while the NSS model experienced

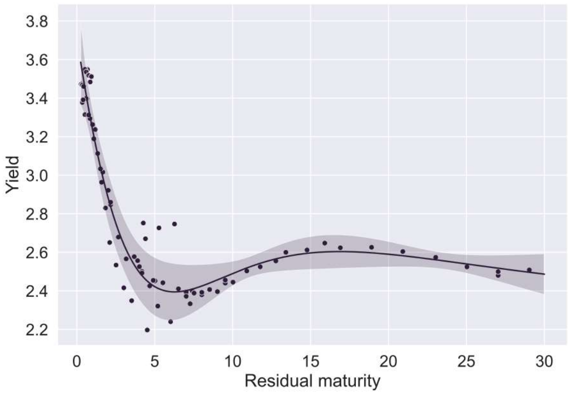

Figure 5.11: $OLP(5)$ yield curve with 95% regression interval, computed using the finite-sample Chebyshev inequality

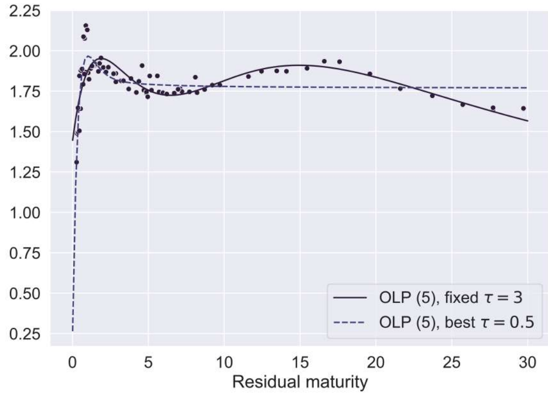

Figure 5.12: $OLP(5)$ fitting of fixed versus best (lowest-RMSE) τ

a substantial drop in flexibility. Fixed-τ configuration carries a set of particular benefits, including

- higher level of stability;
- inter-temporal consistency of basis curves, which enables one to use the parameter time series in a broader modeling context;
- linearity of the estimation problem, which implies the existence of a unique solution and the possibility to apply linear methods (e.g., to compute parameter confidence regions, in outlier detection, etc.).

Although fixing τ to a prescribed value inevitably leads to a reduction in model flexibility, the loss is not as prominent in $OLP(5)$, particularly in comparison to NSS. However, the fact that the RMSE (or any other statistic) under fixed τ could be improved under a different τ may not necessarily mean that the estimated yield curve would be more *accurate*. Idiosyncratic and stochastic factors that may be unrelated to actual interest rate dynamics can result in a different value of τ being optimal. Fixing the value of τ, therefore, constrains the ability of the yield curve to assume atypical shapes and shift in unlikely ways. In this sense, fixed τ configuration acts as a form of implicit regularization.

Bibliography

[1] D. E. Allen, L. C. Thomas, H. Zheng, *Stripping Coupons with Linear Programming*, Working Paper (2000), University of Edinburgh

[2] N. Anderson, J. Sleath, *New estimates of the UK real and nominal yield curves*, Working Paper (2001), Bank of England

[3] J. Annaert, A. G. P. Claes, M. J. K. De Ceuster, H. Zhang, *Estimating the Yield Curve Using the Nelson-Siegel Model: A Ridge Regression Approach*, International Review of Economics & Finance (2012)

[4] Bank for International Settlements, *Zero-Coupon Yield Curves – Technical Documentation*, BIS Paper No. 25 (2005)

[5] Bank for International Settlements, *Interest rate risk in the banking book*, BIS (2016)

[6] T. Björk, J. Christensen, *Interest rate dynamics and consistent forward rate curves*, Mathematical Finance (1999)

[7] R. Bliss, *Testing term structure estimation methods*, Working Paper No. 96-12a (1996), Federal Reserve Bank of Atlanta

[8] R. Bliss, *Movements in the term structure of interest rates*, Economic Review (1997), Federal Reserve Bank of Atlanta

[9] D. Bolder, D. Stréliski, *Yield Curve Modelling at the Bank of Canada*, Technical Report No. 84 (1990), Bank of Canada

[10] J. Botkuljak, D. Grahovac, *Estimating the Yield Curve: An Iterated OLS Approach*, manuscript in preparation

[11] H. J. Büttler, *An Orthogonal Polynomial Approach to Estimate the Term Structure of Interest Rates*, Swiss National Bank Working Papers (2007-8)

[12] J. Caks, *The Coupon Effect on Yield to Maturity*, The Journal of Finance (1977)

[13] *Volatility Index® Methodology: Cboe Volatility Index®*, Cboe (2023)

[14] J. Christensen, F. Diebold, G. Rudebusch *An arbitrage-free generalized Nelson-Siegel term structure model*, The Econometrics Journal (2009)

[15] J. Christensen, F. Diebold, G. Rudebusch, *The Affine Arbitrage-Free Class of Nelson-Siegel Term Structure Models*, Working Paper Series (2010), Federal Reserve Bank of San Francisco

[16] L. Coroneo, K. Nyholm, and R. Vivada-Koleva, *How Arbitrage-Free Is the Nelson-Siegel Model?*, Working Paper Series (2008), European Central Bank

[17] R. K. Crump, N. Gospodinov, *Deconstructing the Yield Curve*, Federal Reserve Bank of New York Staff Reports, no. 884 (2019)

[18] M. De Pooter, F. Ravazzolo, D. van Dijk, *De Pooter, Michiel and Ravazzolo, Francesco and van Dijk, Dick J.C., Predicting the Term Structure of Interest Rates: Incorporating Parameter Uncertainty, Model Uncertainty and Macroeconomic Information*, Tinbergen Institute Discussion Paper (2007)

[19] A. F. Desmond, *Optimal estimating functions, quasi-likelihood and statistical modelling*, Journal of Statistical Planning and Inference (1997)

[20] F. X. Diebold, C. Li, *Forecasting the term structure of government bond yields*, Journal of Econometrics (2006)

[21] P. Donati, F. Donati, *Modelling and forecasting the yield curve under model uncertainty*, Working Paper Series (2008), European Central Bank

[22] *Tools for calculating and interpreting real interest rate developments in the euro area*, Monthly Bulletin (2005), European Central Bank

[23] A. Estrella, F. S. Mishkin, *The yield curve as a predictor of recessions in the United States and Europe*, Current Issues in Economics and Finance (1996), Federal Reserve Bank of New York

[24] A. Estrella, M. R. Trubin, *The Yield Curve as a Leading Indicator: Some Practical Issues*, Current Issues in Economics and Finance (2006), Federal Reserve Bank of New York

[25] F. J. Fabozzi, *The Handbook of Fixed Income Securities*, McGraw Hill (2012)

[26] F. J. Fabozzi, L. Marterllini, P. Priaulet, *Predictability in the Shape of the Term Structure of Interest Rates*, Journal of Fixed Income (2005)

[27] T. S. Ferguson, *A Course in Large Sample Theory*, Routledge (1996)

[28] D. Filipović, *A Note on the Nelson-Siegel Family*, Mathematical Finance (1999)

[29] L. Fonseca, P. McQuade, I. Van Robays and A. L. Vladu, *The inversion of the yield curve and its information content in the euro area and the United States*, Economic Bulletin (2023), European Central Bank

[30] M. Gilli, S. Grosse, E. Schumann, *Calibrating the Nelson–Siegel–Svensson model*, COMISEF Working Papers Series (2010)

[31] R. S. Gurkaynak, B. Sack, and J. H. Wright, *The U.S. Treasury Yield Curve: 1961 to the Present*, Finance and Economics Discussion Series Divisions of Research & Statistics and Monetary Affairs (2006), Federal Reserve Board

[32] P. S. Hagan, G. West, *Interpolation Methods for Curve Construction*, Applied Mathematical Finance (2006)

[33] B. E. Hansen, *Econometrics*, Princeton University Press (2021)

[34] S. D. Horn, R. A. Horn, D.B. Duncan, *Estimating Heteroscedastic Variances in Linear Models*, Journal of the American Statistical Association (1975)

[35] A.S. Hurn, K.A. Lindsay, V. Pavlov, *Smooth estimation of yield curves by Laguerre functions*, International Congress On Modeling And Simulation Advances And Applications For Management And Decision Making (2005)

[36] J. Jacod, M. Sørensen *A review of asymptotic theory of estimating functions*, Stat Inference Stoch Process 21 (2018)

[37] R. I. Jennrich, *Asymptotic properties of nonlinear least square estimators*, The Annals of Mathematical Statistics (1969)

[38] J. Jesus, R. E. Chandler, *Estimating functions and the generalized method of moments*, Interface Focus (2011)

[39] L. Krippner, *Modelling the Yield Curve with Orthonormalised Laguerre Polynomials: A Consistent Cross-Sectional and Inter-Temporal Approach*, Department of Economics Working Paper Series (2003). Hamilton, New Zealand: University of Waikato

[40] L. Krippner, Intertemporally-Consistent and Arbitrage-Free Version of the Nelson and Siegel Class of Yield Curve Models, Department of Economics Working Paper Series (2005). Hamilton, New Zealand: University of Waikato

[41] L. Krippner, *A Theoretically Consistent Version of the Nelson and Siegel Class of Yield Curve Models*, Applied Mathematical Finance (2006)

[42] L. Krippner, *A Theoretical Foundation for the Nelson–Siegel Class of Yield Curve Models*, Journal of Applied Econometrics (2015)

[43] E. L. Lehmann, G. Casella, Theory of point estimation, Springer (1998)

[44] R. Litterman, J. Scheinkman, *Common factors affecting bond returns*, Journal of Fixed Income (1991)

[45] J. H McCulloch, *Measuring the Term Structure of Interest Rates*, The Journal of Business (1971)

[46] C. R. Nelson, A. F. Siegel, *Parsimonious Modeling of Yield Curves*, The Journal of Business (1987)

[47] W. K. Newey, *Uniform Convergence in Probability and Stochastic Equicontinuity* Econometrica (1991)

[48] P. Nymand-Andersen, *Yield curve modelling and a conceptual framework for estimating yield curves: evidence from the European Central Bank's yield curves*, Statistics Paper Series (2018), European Central Bank

[49] L. Sarno, D. L. Thornton, G. Valente, *The Empirical Failure of the Expectations Hypothesis of the Term Structure of Bond Yields*, Financial and Quantitative Analysis (2007)

[50] A. Sasongko, C. A. Utama, B. Wibowo, Z. A. Husodo, *Modifying hybrid optimisation algorithms to construct spot term structure of interest rates and proposing a standardised assessment*, Computational Economics (2019)

[51] R. J. Shiller, J. Y. Campbell, K. L. Schoenholtz, *Forward Rates and Future Policy: Interpreting the Term Structure of Interest Rates*, Brookings Papers on Economic Activity (1983)

[52] C. G. Small, J. Wang, *Numerical Methods for Nonlinear Estimating Equations*, Oxford Statistical Science Series (2003)

[53] L. E. Svensson, *Estimating and Interpreting Forward Interest Rates: Sweden 1992-1994*, Discussion Paper 1051 (1994), Centre for Economic Policy Research

[54] G. Szegő, *Orthogonal polynomials, 4th edition*, American Mathematical Society Colloquium Publications (1975), American Mathematical Society

[55] D.F. Waggoner, *Spline Methods for Extracting Interest Rate Curves from Coupon Bond Prices*, Working Paper (1997), No. 97-10, Federal Reserve Bank of Atlanta

[56] R.R. Wahlstrøm, F. Paraschiv, M. Schürle, *A Comparative Analysis of Parsimonious Yield Curve Models with Focus on the Nelson-Siegel, Svensson and Bliss Versions*, Computational Economics (2022)

[57] K. Yuan, R. I. Jennrich, *Asymptotics of Estimating Equations under Natural Conditions*, Journal of Multivariate Analysis (1998)

Abstract

The European Central Bank (ECB) relies on the timely availability of high-quality financial market indicators to facilitate data-driven decision-making. Gauges of interest rate dynamics are of particular interest to the central bank, given their relevance for monetary policy and financial stability. By capturing the term structure of interest rates, zero-coupon yield curves serve as the key barometer of market expectations of monetary policy, economic growth, and inflation. Correspondingly, the prevailing practice among major central banks, including the ECB, is to maintain an internal system for estimating yield curves. This thesis aims to contribute to three areas relevant to implementing such systems. First, a natural generalization of the Nelson-Siegel model based on Laguerre polynomials is evaluated as an alternative to the popular Svensson extension. The evidence from the conducted evaluation study points to several theoretical and empirical arguments in favor of such an approach, particularly in the areas of model stability, flexibility, and out-of-sample forecasting performance. Second, the empirical study evaluates the nature of the trade-off between goodness-of-fit and model stability resulting from applying different methodologies to estimating the nonlinear decay parameter. Finally, a computationally efficient method based on iterated least-squares is considered for fitting the zero-coupon yield curve to a sample containing fixed-coupon bonds, under a fixed nonlinear decay parameter. The method provides an alternative to the common practice of minimizing either yield-to-maturity error or (weighted) price error.

Keywords

yield curve, term structure, Nelson and Siegel model, Svensson model, Laguerre polynomials, curve fitting, coupon stripping, empirical study

Acknowledgements

I extend my gratitude to Professor Danijel Grahovac, my Thesis supervisor, for his invaluable support and direction throughout the process of preparing this Thesis.

I thank Dr. Leo Krippner for his comments on the empirical and theoretical aspects of yield curve modeling based on Laguerre polynomials.

I thank my former colleagues from the European Central Bank, Mr. Antonio Moreno and Mr. Sébastien Pérez-Duarte, for their comments on the Thesis structure and scope, their help with obtaining the data, and the opportunity to acquire practical knowledge of the yield curve production process.

About the author

In 2021, I completed my undergraduate studies in Mathematics and Computer Science at the Department of Mathematics (now School of Applied Mathematics and Informatics), University of Osijek, having graduated with *cum laude* honors. I continued to pursue my graduate studies in Mathematics (Financial Mathematics and Statistics module) at the Department. During this time, I served as a member of the Student Council from 2021 to 2023, assuming the role of the Council president during the academic year of 2021. Throughout my years of study, I participated in various student competitions. Notably, I was a part of the team that qualified for the EMEA Regional Final of the CFA Research Challenge 2022/2023, representing CFA Society Croatia. As part of my student practice, I worked at the Croatian Financial Services Supervisory Agency and the Croatian National Bank. I began my professional career as a Trainee at the European Central Bank in 2021 and progressed to the role of Research Analyst in 2022. In this capacity, I contributed to the development of the new euro area yield curve production system.