



Stripping the Swiss discount curve using kernel ridge regression

Nicolas Camenzind¹ · Damir Filipović^{1,2}

Received: 15 November 2023 / Revised: 14 March 2024 / Accepted: 8 May 2024 /
Published online: 7 June 2024
© The Author(s) 2024

Abstract

We analyze and implement the kernel ridge regression (KR) method developed in Filipovic et al. (Stripping the discount curve—a robust machine learning approach. Swiss Finance Institute Research Paper No. 22–24. SSRN. <https://ssrn.com/abstract=4058150>, 2022) to estimate the risk-free discount curve for the Swiss government bond market. We show that the insurance industry standard Smith–Wilson method is a special case of the KR framework. We recapitulate the curve estimation methods of the Swiss Solvency Test (SST) and the Swiss National Bank (SNB). In an extensive empirical study covering the years 2010–2022 we compare the KR curves with the SST and SNB curves. The KR method proves to be robust, flexible, transparent, reproducible and easy to implement, and outperforms the benchmarks in- and out-of-sample. We show the limitations of all methods for extrapolating the yield curve and propose possible solutions for the extrapolation problem. We conclude that the KR method is the preferred method for estimating the discount curve.

Keywords Yield curve estimation · Swiss government bond market · Smith–Wilson method · Swiss Solvency Test · Swiss National Bank · Machine learning in finance · Reproducing kernel Hilbert space

JEL Classification C14 · C55 · E43 · E52 · G12 · G22

We thank Jens Christensen, Nikola Mirkov, Robert Müller, Thomas Nellen, Thomas Nitschka, Michael Schmutz, Lutz Wilhelmy, members of the SAA (Swiss Association of Actuaries) Working Group on yield curves, and conference participants of the Annual Meeting of the Swiss Association of Actuaries 2022, and CONVENTION A 2022, and two anonymous referees for helpful comments.

✉ Damir Filipović
damir.filipovic@epfl.ch

Nicolas Camenzind
nicolas.camenzind@epfl.ch

¹ EPFL, Lausanne, Switzerland

² Swiss Finance Institute, Zurich, Switzerland

1 Introduction

The risk-free discount curve, or zero-coupon yield curve, is a key variable for valuing and hedging assets and liabilities and for various other tasks. It reflects market expectations regarding the current and future states of the economy. The discount curve is not observed and must be estimated from noisy quotes of fixed income instruments priced by market participants. Any preferred method for estimating the discount curve should arguably have the following desirable characteristics: (i) simple and fast to implement, (ii) transparent and reproducible, (iii) data-driven, (iv) precise representation of the term structure taking into account all market signals, (v) robust to outliers and data selection choices, (vi) flexible for integration of external views, (vii) consistent with finance principles.

We show that the kernel ridge regression (hereafter referred to as KR) method developed in [10] satisfies these properties, and we apply it to the Swiss government bond market. Compared to major bond markets, such as the U.S. Treasury market, the Swiss market is much smaller and liquidity is lower, which may result in instruments being priced less efficiently. This poses an additional challenge for downstream applications that rely on yield information for less liquid maturity ranges for which only limited data are available. This results in the need for reliable interpolation and extrapolation of the yield curve in a suitable function space. The KR method does just that. The KR curve is given in closed form as the solution of a kernel ridge regression in a reproducing kernel Hilbert space (RKHS) [18] consisting of twice differentiable functions on the positive half line. The KR curve is obtained by trading off the fitting error and its smoothness.

Kernel methods such as KR are an integral part of machine learning, see, e.g., [24]. KR is a non-parametric estimation method for a curve in an infinite-dimensional RKHS. The so-called Representer Theorem implies that the infinite-dimensional estimation problem reduces to the determination of a finite number of coefficients. This number depends on and is implied by the prevailing data. The KR estimator is calibrated with three hyperparameters, one of which tunes the trade-off between fitting error and smoothness and the other two determine the smoothness measure. These hyperparameters are selected by cross-validation, making the KR method fully data-driven. KR thus differs fundamentally from parametric curve estimators, in which a specific functional form of the discount (or yield) curve is predefined.

Applications of the KR method are manifold. KR curves can be used, for example, as the basis for solvency capital calculations in the insurance industry or to reflect the term structure of government bond prices as published by central banks. To this end, we show how the current market standard in the European insurance industry, the Smith–Wilson method [26], is formally embedded in the KR framework. We also recapitulate the parametric curve estimation methods of the Swiss Solvency Test (SST) and the Swiss National Bank (SNB).

We conduct an extensive empirical study with daily data on Swiss government bonds for the years 2010–2022, which are publicly available from a SNB website [23]. However, matured bonds are retrospectively removed from this website. The SNB provided us with the complete daily data on request, but only from September

2018 for licensing reasons.¹ The long sample 2010–2022, which we use for the long-term analysis of KR, therefore has missing data. In turn, we use the complete but short sample 2018–2022 for benchmarking. We discuss model selection and find that the KR method is robust with respect to the choice of hyperparameters. In a comparative study, we compare the fit and shapes of KR curves with the SST and SNB curves. The KR method outperforms the benchmarks on all error metrics in- and out-of-sample. Our results hold both at the aggregate level and for all pre-defined maturity buckets and are robust over time.

The KR method is related to Gaussian process regression and therefore allows for a Bayesian interpretation, based on which we derive confidence bands around the KR curve estimates. We show with examples that the confidence bands accurately indicate the ranges of sparse or missing data.

We show how external views can be easily incorporated into the KR curve in the form of constraints to match exogenously given yields. We also discuss the extrapolation of the discount curve beyond the quoted maturity range. We find that none of the methods in scope can credibly provide robust and accurate yield curves up to 100 years. Any estimation method in this range requires external information about long term yields. On the other hand, we find that the stability of the KR method is already significantly improved by adding longer dated bonds. We propose possible solutions to the extrapolation problem. A detailed analysis is left for further research.

Given the fundamental problem of estimating the discount curve and its wide application it is not surprising that there exists an extensive literature on the topic. The most well known methods include Nelson–Siegel–Svensson [12, 17, 25], Smith–Wilson [26], Fama–Bliss [7], Liu–Wu [15], and KR [10], arguably. Nelson–Siegel–Svensson is parametric. Here a parsimonious smooth parametric form is specified and corresponding parameters are estimated by minimizing pricing errors, which leads to a non-convex optimization problem. Smith–Wilson, Fama–Bliss, Liu–Wu, and KR fall within the category of non-parametric methods. In contrast to Nelson–Siegel–Svensson, these methods exhibit a larger flexibility. We show in the empirical part that our non-parametric method captures global as well as local nuances of the discount curve. In addition, there are many other frequently used methods such as spline based methods. The underlying assumption is to model the discount curve or yield curve locally by polynomials [1, 16, 27, 28].

The prevailing benchmark in the Swiss market is a form of the Nelson–Siegel–Svensson method [17, 25] with additional constraints estimated by the SNB. The underlying optimization problem is formulated in a BIS working paper [2]. A dynamic Nelson–Siegel model for the Swiss discount curve is estimated in [4]. The European Insurance and Occupational Pensions Authority (EIOPA) defines the technicalities used in the regulatory Solvency II framework, in particular the Smith–Wilson method [26], see [5, 14, 29]. For the Swiss market, the Financial Market Supervisory Authority (FINMA) [9] provides risk-free discount curves for the SST curves that are based on Smith–Wilson.

The paper is structured in the following way. In Sect. 2, we introduce the formal setup of the KR method. In Sect. 3, we establish the relationship between KR and

¹ A special thanks to Robert Müller from the SNB for his insights and help.

Smith–Wilson and outline the technicalities of the SST curves. In Sect. 4, we discuss the data and model selection. This includes the choice of the error metrics and hyper-parameters. In Sect. 5, we conduct a comparison study between the KR, SNB, and SST estimates. In Sect. 6, we discuss the extrapolation problem. In Sect. 7, we elaborate on the correspondence between kernel ridge regression and Bayesian interpretation stemming from Gaussian processes. In Sect. 8, we conclude and summarize the key findings. The Appendix contains an analysis of KR curves with constraints at the short end. An online appendix contains additional figures.

2 Formal setup

At a given business day, we observe prices of M fixed income securities with time to cash flow dates $0 < x_1 < \dots < x_N$. The prices are given by the vector $P = (P_1, \dots, P_M)^\top$. Cash flows are captured by the matrix C whose entries C_{ij} denote the cash flow of security i at x_j . The unobserved discount curve is represented by a function $g : [0, \infty) \rightarrow \mathbb{R}$ where $g(x)$ denotes the present value of a zero-coupon bond with time to maturity x . This relates to the zero-coupon yield $y(x)$ with maturity x via $g(x) = e^{-y(x) \cdot x}$. We simply call $y(x)$ the yield in what follows. The law of one price implies that the vector of fundamental values P^g of all securities with underlying discount curve g is equal to

$$P^g = Cg(x),$$

where we use the notation $\mathbf{x} := (x_1, \dots, x_N)^\top$ and write $f(\mathbf{x}) := (f(x_1), \dots, f(x_N))^\top$ for the corresponding array of values for any function f . The observed prices may differ from the fundamental values due to the lack of a deep, liquid, and transparent market, or data errors. Formally,

$$P = P^g + \epsilon, \tag{1}$$

where $\epsilon \in \mathbb{R}^M$ denotes pricing errors.

In [10], the function g is estimated using the theory of reproducing kernel Hilbert spaces (RKHS). This boils down to the kernel ridge regression (KR) problem

$$\min_{g \in \mathcal{G}_{\alpha, \delta}} \left\{ \underbrace{\sum_{i=1}^M \omega_i (P_i - P_i^g)^2}_{\text{pricing error}} + \lambda \underbrace{\|g\|_{\alpha, \delta}^2}_{\text{smoothness}} \right\}, \tag{2}$$

for a regularisation parameter $\lambda > 0$ and exogenous weights $\omega_i > 0$. The hypothesis space $\mathcal{G}_{\alpha, \delta}$ consists of all twice differentiable functions $g : [0, \infty) \rightarrow \mathbb{R}$ with $g(0) = 1$ and finite norm given by

$$\|g\|_{\alpha, \delta}^2 := \int_0^\infty \left(\delta g'(x)^2 + (1 - \delta)g''(x)^2 \right) e^{\alpha x} dx, \tag{3}$$

for some shape parameter $\delta \in [0, 1]$ and maturity weight $\alpha \geq 0$. This norm entails the two standard measures for tension, $g'(x)^2$, and curvature, $g''(x)^2$, of a function g . We denote by $k : [0, \infty) \times [0, \infty) \rightarrow \mathbb{R}$ the reproducing kernel related to $\mathcal{G}_{\alpha,\delta}$.² More details on the space $\mathcal{G}_{\alpha,\delta}$, including a closed-form expression for the kernel k , are given in [10, Theorem 2]. For completeness, we recall here the expression for $\alpha > 0$ and $\delta = 0$, which will prove to be the default setting in the empirical part:

$$k(x, y) = -\frac{\min\{x, y\}}{\alpha^2} e^{-\alpha \min\{x, y\}} + \frac{2}{\alpha^3} \left(1 - e^{-\alpha \min\{x, y\}}\right) - \frac{\min\{x, y\}}{\alpha^2} e^{-\alpha \max\{x, y\}}. \tag{4}$$

The above estimation problem is completely determined up to the hyperparameters λ , α and δ . These parameters will be estimated via an out-of-sample cross validation of the weighted pricing errors. This renders the KR method fully data-driven. With regards to the choice of the weights ω_i various possibilities arise. A common choice in the finance literature is

$$\omega_i = \frac{1}{M} \frac{1}{(D_i P_i)^2} \tag{5}$$

where D_i denotes the modified duration of security i . This choice of ω_i in (2) corresponds to a first order approximation of the mean squared yield fitting errors,

$$\omega_i (P_i - P_i^g)^2 \approx \frac{1}{M} (Y_i - Y_i^g)^2,$$

where Y_i and Y_i^g denote the yield to maturity (YTM) of the i -th security based on the quoted price P_i and fundamental value P_i^g , respectively. An infinite weight $\omega_i = \infty$ is also possible and corresponds to an exact pricing of the i -th security, which gives the flexibility for integration of external views on the yield curve, see Example 2.3 below.

The solution of problem (2) boils down to a simple kernel ridge regression, which is given in closed form. We recall here the corresponding result in [10, Theorems 2 and A.1], where we define the $N \times N$ -kernel matrix \mathbf{K} by $\mathbf{K}_{ij} = k(x_i, x_j)$ and we write $\mathbf{1} = (1, \dots, 1)^\top$:

Theorem 2.1 (Kernel-Ridge (KR) Solution) *The fundamental problem (2) has a unique solution \hat{g} , which is given in closed form by*

$$\hat{g}(x) = 1 + \sum_{j=1}^N k(x, x_j) \beta_j,$$

where $\beta = (\beta_1, \dots, \beta_N)^\top$ is given by

$$\beta = \mathbf{C}^\top (\mathbf{C} \mathbf{K} \mathbf{C}^\top + \Lambda)^{-1} (\mathbf{P} - \mathbf{C} \mathbf{1}), \tag{6}$$

² Note that $\mathcal{G}_{\alpha,\delta}$ is not a linear space, in view of the initial condition $g(0) = 1$. In fact, k is the reproducing kernel of the RKHS $\mathcal{H}_{\alpha,\delta}$ consisting of twice weakly differentiable functions $h : [0, \infty) \rightarrow \mathbb{R}$ with $h(0) = 0$ and finite norm (3). We then have $\mathcal{G}_{\alpha,\delta} = \{g = 1 + h \mid h \in \mathcal{H}_{\alpha,\delta}\}$.

where $\Lambda := \text{diag}(\lambda/\omega_1, \dots, \lambda/\omega_M)$ where we set $\lambda/\infty := 0$.³

The solution (6) boils down to the inversion of a $M \times M$ -matrix, which is computationally a simple task. The KR framework is extremely flexible and covers a wide range of possible solutions. In fact, many popular model curves such as Fama–Bliss, Nelson–Siegel–Svensson, and the insurance industry standard Smith–Wilson lie in the space $\mathcal{G}_{\alpha,\delta}$ for appropriate choices of α and δ , see [10, Theorem 2]. We elaborate on the Smith–Wilson curves in more detail in the following section.

Remark 2.2 The discount curve $g(x)$ is a function of the time to maturity x whose actual values depend on the choice of the day count convention. In this paper we assume the *ACT/365* convention. That is, $x_i = i/365$ where i denotes the number of calendar days between the spot date and the cash flow date. Other methods in the literature may be based on different day count conventions. E.g., the published SNB curve parameters are based on the *GERMAN 30/360* convention. Strictly speaking, one cannot directly compare the discount curves of different methods unless they are based on the same day count convention. To compare the methods, one would have to compare the model implied bond prices derived in the respective day count conventions. Numerically, however, the differences are not economically significant. We compared time series of yields $y(x)$ of fixed maturities x ranging from 5 to 40 years implied by KR and SNB curves under both *ACT/365* and *GERMAN 30/360* conventions. That is, we compared $y(x)$ with $y(x + \Delta x)$ where Δx reflects the shift due to leap years during x years. E.g., for $x = 10$, we set $\Delta x = 2/365$. We found that the differences in yields $y(x) - y(x + \Delta x)$ are of the order 10^{-6} for maturities up to $x = 40$, throughout the sample, and of order 10^{-5} for $x = 50$, in first part of the sample. We also found that the day count convention has a greater impact on the calculation of accrued interest, which relates economic dirty prices to quoted clean prices. For simplicity, the *ACT/365* convention is used below to derive the model-implied prices for all methods.

Example 2.3 As an example for the integration of external views on the yield curve, we consider here the practice of some central banks to force the implied short rate of the estimated curve to match the prevailing benchmark short rate r_{short} , e.g., SARON. This can simply be achieved in Theorem 2.1 by defining one of the instruments, say $i = 1$, as zero-coupon bond maturing the next day, at $x_1 = 1/365$, and specifying its price $P_1 = e^{-r_{\text{short}} \cdot x_1}$ and cash flow $C_{1j} = 1$ for $j = 1$ and $C_{1j} = 0$ for $j > 1$, and setting its weight $\omega_1 = \infty$.

3 Smith–Wilson curves

This section outlines the relationship of our KR method and Smith–Wilson (hereafter referred to as SW). We first introduce the theoretical foundation to reformulate the SW method as a problem of the form given in Eq. (2). SW plays an important role in the insurance industry. Regulatory bodies such as the FINMA rely on risk-free discount

³ In this case, we also assume that the submatrix of CKC^T that consists of all rows and columns i corresponding to the infinite weights $\omega_i = \infty$ is invertible.

curves based on SW, e.g., the interest rate curves used in SST to discount insurance companies' assets and liabilities. We then apply the theoretical findings and introduce how to generate KR based SST curves. We also specify how our method can perfectly replicate the SST curves given by FINMA. This allows in theory to generate SST curves on a daily basis while FINMA provides SST curves only once a year.

3.1 Relation to KR method

SW [26] has been the insurance industry standard in Europe for constructing the discount curve used in the regulatory Solvency II framework, see the technical documentations of the European Insurance and Occupational Pensions Authority [5], the European Systemic Risk Board [6], and [13, 14, 29]. SW considers discount curves of the form $g_{SW}(x) = e^{-y_\infty x} g_0(x)$, for some $g_0 \in \mathcal{G}_{0,\delta}$ with $\delta \in (0, 1)$, and $y_\infty = \log(1 + UFR) > 0$, for the so-called ultimate forward rate $UFR > 0$. The SW method assumes exact pricing of all bonds up to a certain maturity $x_N < \infty$, which is also called the last liquid point (LLP), and disregards all bonds with longer maturity.

Formally, SW solves the exact pricing problem with regularization

$$\begin{aligned} & \min \|g_0\|_{0,\delta}^2 \\ & \text{s.t. } P = Cg_{SW}(\mathbf{x}), \\ & g_{SW}(x) = e^{-y_\infty x} g_0(x), \\ & g_0 \in \mathcal{G}_{0,\delta}. \end{aligned} \tag{7}$$

This can be brought into the form (2) by rewriting $Cg_{SW}(\mathbf{x}) = \tilde{C}g_0(\mathbf{x})$ for the tilted cash flow matrix

$$\tilde{C} := C \text{diag}(e^{-y_\infty x}).$$

Problem (7) now reads as

$$\begin{aligned} & \min \|g_0\|_{0,\delta}^2 \\ & \text{s.t. } P = \tilde{C}g_0(\mathbf{x}), \\ & g_0 \in \mathcal{G}_{0,\delta}. \end{aligned} \tag{8}$$

This is just a special case of Theorem 2.1 where all weights are infinite, that is, $\Lambda = 0$, see Footnote 3. From the solution to (8) we thus obtain the SW discount curve

$$\hat{g}_{SW}(x) = e^{-y_\infty x} \hat{g}_0(x) \tag{9}$$

where

$$\hat{g}_0(x) = 1 + \sum_{j=1}^N k(x, x_j) \beta_j = 1 + e^{y_\infty x} \sum_{j=1}^N W(x, x_j) \underbrace{\frac{1}{\delta \rho} e^{y_\infty x_j} \beta_j}_{=\xi_j} \tag{10}$$

and

$$\beta = \tilde{C}^\top (\tilde{C} \mathbf{K} \tilde{C}^\top)^{-1} (P - \tilde{C} \mathbf{1}). \tag{11}$$

Here

$$\begin{aligned} k(x, y) &= \frac{1}{\delta} \min\{x, y\} + \frac{1}{2\delta\rho} \left(e^{-\rho(x+y)} - e^{\rho \min\{x, y\} - \rho \max\{x, y\}} \right) \\ &= \frac{1}{\delta} \min\{x, y\} - \frac{1}{\delta\rho} e^{-\rho \max\{x, y\}} \sinh(\rho \min\{x, y\}) \end{aligned}$$

with $\rho := \sqrt{\delta/(1 - \delta)}$, is the kernel given in [10, Theorem 2] for $\alpha = 0, \delta \in (0, 1)$. In the second equation, we used that $\max\{x, y\} - (x + y) = -\min\{x, y\}$. This is related to the “Wilson kernel function”

$$W(x, y) = e^{-y_\infty(x+y)} \delta\rho k(x, y).$$

Comparing this to [5, Paragraph 134], for the function $H(u, v) = \delta\rho k(u, v)$, we see that our parameters correspond to the SW parameters speed of convergence “ α ” and ultimate forward intensity “ ω ” by

$$“\alpha” = \rho, \quad “\omega” = y_\infty, \tag{12}$$

respectively. A further inspection shows that then the above expressions (9)–(11) are identical to the expressions in [5, Paragraphs 149–151], with coefficients $\zeta = “C\mathbf{b}”$ in (10).

Remark 3.1 The SW parameters in (12) can be interpreted as follows. The larger the speed of convergence “ α ” = ρ , the closer δ is to 1 in (3). As the auxiliary curve \hat{g}_0 minimizes $\|g_0\|_{0,\delta}$, the quicker $x \mapsto \hat{g}_0(x)$ flattens out and converges to a constant. In turn, the quicker the exponentially tilted SW curve $x \mapsto \hat{g}_{SW}(x) = e^{-y_\infty x} \hat{g}_0(x)$ converges to an exponential decay at rate “ ω ” = y_∞ , which is the ultimate forward intensity or, equivalently, the infinite maturity yield.

Remark 3.2 From (9) and [10, Lemma 8(i) and (iii) and Lemma 1(iv)], it follows that the SW curve g_{SW} lies in $\mathcal{G}_{\alpha,\delta}$ for any $\alpha \in [0, 2y_\infty)$. The converse is not true: not every curve $g \in \mathcal{G}_{\alpha,\delta}$ is of the SW-form $g(x) = e^{-y_\infty x} g_0(x)$ for some $g_0 \in \mathcal{G}_{0,\delta}$. Indeed, a counter-example is given by $g(x) = e^{-\gamma x}$, which is element in $\mathcal{G}_{\alpha,\delta}$, for any $\frac{\alpha}{2} < \gamma < y_\infty$. However, the only possible pre-image $g_0(x) = e^{y_\infty x} g(x) = e^{(y_\infty - \gamma)x}$ exhibits exponential growth for $x \rightarrow \infty$. Hence, in view of [10, Lemma 2], g_0 does not lie in $\mathcal{G}_{0,\delta}$. This implies that our KR curves are superior to SW in terms of our objective function (2).

3.2 SST curves

The Swiss Solvency Test (SST) [9] is a supervisory tool applicable in Switzerland in the insurance industry. The goal is to assess the capitalisation of an insurance company.

Table 1 Historical SST parameters

SST year	LLP	UFR (%)	α
2022	15	1.95	0.1
2021	15	2.1	0.1
2020	15	2.25	0.1
2019	15	2.4	0.1
2018	15	2.55	0.1
2017	15	2.7	0.1
2016	15	2.7	0.1
2015	15	2.9	0.1
2014	15	2.9	0.1
2013	15	2.9	0.1
2012	15	2.9	0.1

This table contains the historical SST parameters since 2012. Source: FINMA

To value a company's assets and liabilities a risk-free discount curve is required. For this FINMA publishes once a year the SST curve. As outlined in the technical documentation "Technische Beschreibung SST-Bilanz, risikolose Zinskurven und FDS" (only available in German and French) on [9], since 2012, the SST curve is based on the SW method described above with underlying Swiss government bond market data taken from the SNB [23], which is also the data for our empirical study below. Concretely, the SST curve matches the discount bond prices (computed from the zero-coupon yields published by the SNB) with maturities 1 year, 2 years, ..., 10 years and 15 years, and where 15 years is taken as the LLP. Additionally, FINMA publishes the speed of convergence, " α ", and the UFR. Table 1 contains historical SST parameters, where "UFR" here in fact denotes the continuously compounded ultimate yield y_∞ .

4 Model selection

In this section we describe the data and the evaluation metrics for our empirical analysis. We then derive and discuss the baseline values for the hyperparameters λ , α and δ , and weights ω_j . We study their robustness and compare local and global optimal hyperparameter values thereafter.

4.1 Data

For our empirical study we use public available data from the SNB [23]. The SNB collects clean prices of Swiss government bonds on a daily basis excluding weekends and national Swiss public holidays. To ensure price continuity a waterfall logic is applied, [21, page 68], [22]. Concretely, every day at 10:30 Swiss time (until December 2020 and 11:00 from January 2021 onward) available quotes from a data provider are captured. First choice is a traded price on the respective day. If no transaction was executed the mid price between the bid and ask price is used. If the bid price is missing

the ask price minus 25bps is reported. On the other hand, if the ask price is missing the bid price is used. In the rare event that both prices, bid and ask, are missing the last available traded price is stored in the SNB data set. The methodology was originally published in 2002. All consecutive changes and revisions are documented and publicly available, see [19]. From available clean prices and meta data the accrued interest can easily be calculated to obtain dirty prices. We retrieved the accrued interest from Bloomberg for each day assuming trade and settlement day are on this very same day. Our empirical study is then carried out on dirty prices.

We noticed that the SNB removes matured bonds retrospectively from their website [23]. The SNB provided us with the complete daily data on request, but only from September 2018 for licensing reasons. As a result we use two samples, one long and one short, in the following. The long sample contains daily prices from 1 January 2010 to 30 June 2022 of 22 Swiss government bonds, but exhibits missing data,⁴ The short sample contains all daily prices from 1 September 2018 to 30 June 2022 including three additional Swiss government bonds.⁵ We use the long sample for the long-term analysis and selection of the KR hyperparameters, and the short sample for the comparison study in the next section. Table 2 summarizes the key meta data of all bonds in scope.

The maturity profile of all bonds is shown in Fig. 1. Each black line represents the time to maturity of a particular bond. The red line indicates the bond with longest remaining time to maturity. Before 2014 this was steadily decreasing from slightly less than 40 years. In June 2014 a new 50 years bond was issued, which has remained the longest maturity bond in the subsequent years. Moreover, it is evident from Fig. 1 that prior to 2013 the sample does not include bonds with maturities less than 10 years. This reflects the aforementioned missing data in the long sample. We have highlighted the additional three bonds provided by the SNB that are part of our short sample in blue. Note that also in the long sample period each coupon bond exhibits annual cash flows (coupon payments). At any point in time these cash flows support the estimation of the yield curve in shorter maturity buckets.

The data set contains only fully taxable and only non-callable bonds. This selection is consistent with the standard filters applied in [7, 15]. In [12] they also exclude bonds with maturity less than 90 days due to data quality. Figure 1 reveals that all bonds in our long sample comply with this filter: their maturities are more than 90 days ahead. For our short sample we apply the SNB filter: until the end of 2020, they excluded bonds with a maturity less than 1 year, from 2021 onwards they only exclude bonds with a maturity less than 3 months. Another frequently used filter differentiates between on and off the run bonds and excludes the two mostly recently issued securities (with maturities 2 years, 3 years, 4 years, 5 years, 7 years, 10 years, 20 years, and 30 years) as proposed in, e.g., [12]. However, due to the relative small universe in the Swiss government bond market this filter is not feasible. In fact, the concept of on and off the run bonds does not apply.

⁴ According to [4, Table 1] 12 Swiss government bonds matured between 2010 and 2019, which are not in our sample.

⁵ It also includes the Swiss government bond with maturity date 2019-05-12, which however is filtered out of the short sample as its remaining time to maturity is less than one year.

Table 2 Bond meta data

ISIN	Coupon (%)	Maturity	First coupon date
CH0021908907	2.25	2020-07-06	2006-07-06
CH0111999816	2.0	2021-04-28	2011-04-28
CH0127181011	2.0	2022-05-25	2012-05-25
CH0008435569	4.0	2023-02-11	1999-02-11
CH0127181177	1.25	2024-06-11	2013-06-11
CH0184249990	1.5	2025-07-24	2014-07-24
CH0224396983	1.25	2026-05-28	2015-05-28
CH0031835561	3.25	2027-06-27	2008-06-27
CH0008680370	4.0	2028-04-08	1999-04-08
CH0224397346	0.0	2029-06-22	2017-06-22
CH0224397171	0.5	2030-05-27	2016-05-27
CH0127181029	2.25	2031-06-22	2012-06-22
CH0344958688	0.5	2032-06-27	2019-06-27
CH0015803239	3.5	2033-04-08	2004-04-08
CH0440081393	0.0	2034-06-26	2020-06-26
CH0557778310	0.25	2035-06-23	2022-06-23
CH0024524966	2.5	2036-03-08	2007-03-08
CH0127181193	1.25	2037-06-27	2013-06-27
CH0440081401	0.0	2039-07-24	2020-07-24
CH0127181169	1.5	2042-04-30	2013-04-30
CH0344958498	0.5	2045-06-28	2018-06-28
CH0009755197	4.0	2049-01-06	2000-01-06
CH0344958472	0.5	2055-05-24	2018-05-24
CH0224397338	0.5	2058-05-30	2017-05-30
CH0224397007	2.0	2064-06-25	2015-06-25

This table contains all the relevant meta data of the Swiss government bonds used in the empirical part. The coupon rate of all bonds are paid annually. In blue highlighted are static data of bonds we received from the SNB for the short sample. Source: Bloomberg Finance L.P

4.2 Evaluation

Throughout the paper we use the maturity buckets <1 year, 1 year–5 years, 5 years–10 years, 10 years–15 years, 15 years–25 years and ≥ 25 years to report various result on these aggregated levels. The choice of the maturity buckets is coarser as in [10, 15] due to the sparser data compared to the U.S. Treasury case. Figure 1 shows that bonds are not evenly distributed across these buckets. More bonds fall into the longer end buckets.

We apply the same weighted price and YTM errors as in [10], which are either reported as time series or aggregated into the maturity buckets above. Specifically, we

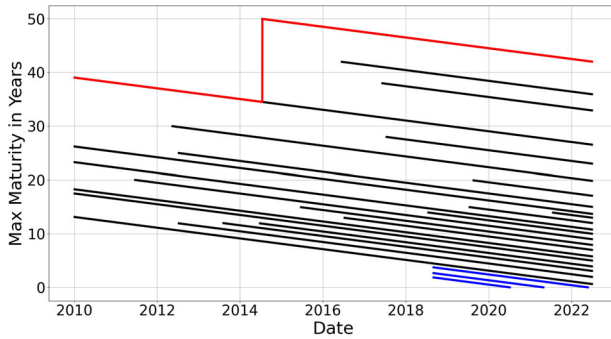


Fig. 1 Maximal time to maturity for the full sample. The figure plots the available bonds and their respective remaining time to maturity over time. The red line indicates the longest time to maturity available in the data set at each point in time, i.e. it is the longest time to maturity of the outstanding bonds in the sample at a particular point in time. Blue lines indicate bonds provided by the SNB used in our short sample

define the root mean squared error at time t as

$$RMSE_t := \sqrt{\sum_{i=1}^{M_t} \omega_{i,t} (P_{i,t} - \hat{P}_{i,t})^2}$$

and the time average root mean squared error as

$$RMSE := \frac{1}{T} \sum_{t=1}^T RMSE_t.$$

Here $\hat{P}_{i,t} = P_{i,t}^{\hat{g}_t}$ denotes the model implied price of instrument i derived from the estimated discount curve \hat{g}_t at time t , and $\omega_{i,t}$ the corresponding weight for the price error. Similarly, we write $\hat{Y}_{i,t} = Y_{i,t}^{\hat{g}_t}$ for the estimated model implied yield to maturity. We use three different error metrics: a duration weighted error with weights given by (5), a relative pricing error that correspond to weights $\omega_{i,t} = \frac{1}{M_t P_{i,t}^2}$, which normalize all bond prices to one, and a YTM based error that is given as the RMSE of model implied yields to maturity,

$$\sqrt{\frac{1}{M_t} \sum_{i=1}^{M_t} (Y_{i,t} - \hat{Y}_{i,t})^2}.$$

The YTM RMSE is the preferred error metric for the estimation of the discount curve. In fact, Fig. 2 shows box plots of the bid-ask implied bond YTM spreads across maturity buckets.⁶ It reveals that bid-ask implied spreads of bond YTM are essentially uniform across maturity buckets, and range in the order of 4–7 bps, for maturities ≥ 1 year.

⁶ Since the SNB does not publish bid-ask bond prices we have used daily bid-ask prices from Bloomberg and converted them to bid-ask yields.

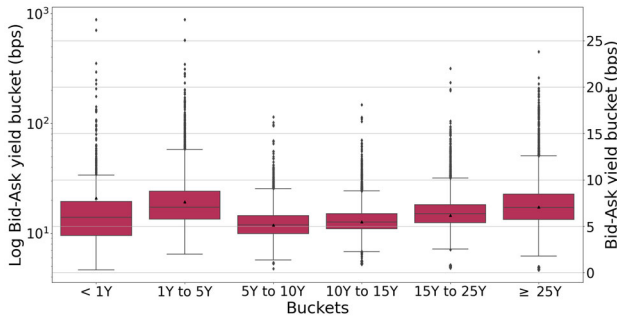


Fig. 2 Yield difference based on Bid-Ask spreads. The plot shows box plots per maturity buckets of bid-ask implied yields. A box plot shows the quartiles of the data set, while the Whiskers show the rest (minimum/maximum) of the distribution, except for the black points, which were determined to be outliers using a method that is a function of the interquartile range. Bucket <1 year which uses a y log-scale on the left while all other buckets use a normal y scale on the right. Source bid-ask prices: Bloomberg Finance L.P. Sample window corresponds to our long sample

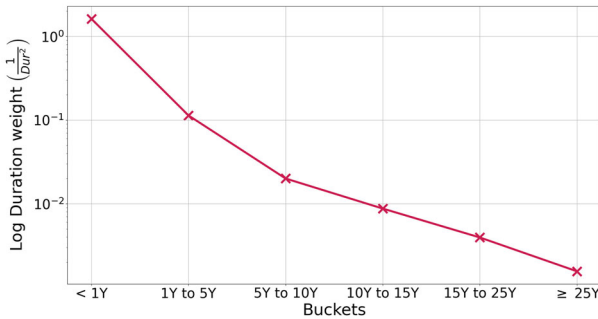


Fig. 3 Logarithmic duration weights. The figure displays the time averaged logarithmic duration weights. Based on the long sample

The duration weights convert price errors into YTM errors, accurately up to first order, so that no additional modification of the weights ω_i is required, see Fig. 3. We therefore use duration weights in (2) for the estimation of KR curves in all subsequent results.

We illustrate the empirical results by plotting the estimated yield time series for 1 months, 3 months, 6 months, 1 year, 5 years, 10 years, 15 years, 20 years, 30 years, 40 years, 50 years, 80 years, 100 years to maturity, see Figs. 13 and 15 below and figures in the online appendix. The 50 years, 80 years, 100 years points are based entirely on extrapolation.⁷ Besides the time series of single yields we plot the entire yield curves on four representative example days, 2010-06-15, 2014-06-16, 2018-06-15 and 2022-06-15. These example days are equally spaced over the sample period and cover different

⁷ The 50 years point falls within the interpolation range on the day in June 2014 when the new 50 years bond was issued.

interest rate regimes since 2010. We avoid month end or year end data, as they could be biased.⁸

4.3 Choice of hyperparameters

The KR method outlined in Sect. 2 requires to choose hyperparameters λ , α and δ . We find these optimal parameters in a purely data-driven way. On each day of the sample period we perform a leave-one-out cross-validation, LOOCV, for a large grid of values of λ , α and δ . The parameters run each through a predefined grid, $\lambda \in \{0.01, 0.05, 0.1, 0.5, 1, 5, 10, 50, 100\}$, $\alpha \in \{0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.10\}$ and $\delta \in \{0, 1e-06, 1e-05, 1e-04, 1e-03, 0.01, 0.1, 1\}$. We build all possible combinations based on these grids. In [10], the regularization parameter λ is scaled by the time-varying factor $1/(365 \cdot x_N)$, where x_N is the longest available time to maturity in years of the quoted bonds any given day. E.g., for $x_N = 30$, this amounts to dividing λ by 10,950. Here we modify the scaling factor and set it to the fixed value 10^{-4} . Both scaling factors are of similar order and allow convenient representation. For the ease of notation we report unscaled values of λ in all plots (e.g., “ $\lambda = 10$ ” refers to a regularisation parameter value of 10^{-3}). We use the long sample without any further filtering for the optimal hyperparameter choice.

The minimal LOOCV based YTM RMSE is attained at the hyperparameter values $\lambda = 10$, $\alpha = 0.02$ and $\delta = 0$. Figures 4 and 5 show the heatmaps for fixed δ and α , respectively. We find similar optimal values as in [10] for the U.S. Treasury market. In particular, the differences in YTM RMSE for very small δ are negligible and the choice of $\delta = 0$ simplifies the model: the smoothness measure in the objective function (2) only involves the second derivative. For λ we find an optimal value of 10, which matches closely the optimal value in [10] for the U.S. Treasury market ($\lambda = 1$). Recall that λ is chosen on a logarithmic scale and here we use a slightly different scaling factor, as explained above. The optimal value for α is found at 0.02 and is smaller as in the U.S. market ($\alpha = 0.05$). As shown in [10] α can be interpreted as the infinity maturity yield. Since the U.S. data exhibits structurally higher interest rates this difference in the optimal value of α might be not surprising. However, a word of caution is warranted as this interpretation is only valid under certain technical assumptions and it does not make any statement about the speed of convergence nor the behaviour of the resulting yield on any finite maturity in the extrapolation area. In summary, we fix as baseline values $\lambda = 10$, $\alpha = 0.02$ and $\delta = 0$ in the following. This corresponds to the kernel (4).

Figures 4 and 5 show that our model is robust with respect to small deviations from the baseline values. However, these heatmaps represent an aggregated view over time only. The following more granular analysis shows how the local optimal parameters behave over time. We refer to “local optimal” for the optimal values of λ , α and δ on a specific day while “global optimal” refers to our aggregated optimal baseline values. Interestingly, we can observe that local optimal values of λ vary frequently, while for α and δ the dispersion seems to be much smaller.

⁸ Month ends tend to distort prices as index rebalancing takes place. Year end is avoided due to low liquidity in the Swiss market.

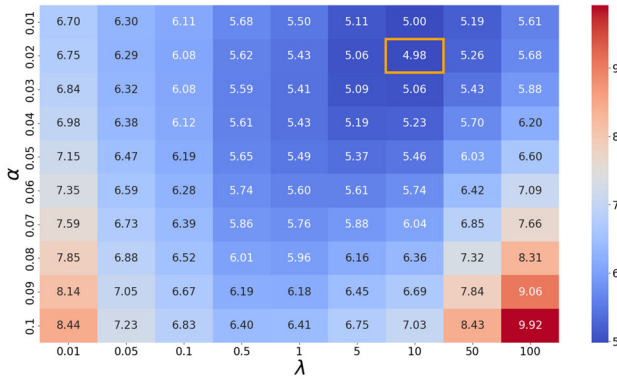


Fig. 4 LOOCV YTM RMSE for λ and α . Based on daily LOOCV the YTM RMSE for a specific grid of λ , α and δ is shown. The figure fixes the optimal δ and shows the two dimensional heatmap varying only λ and α . The orange square indicates the lowest YTM RMSE for corresponding hyperparameters. Based on the long sample

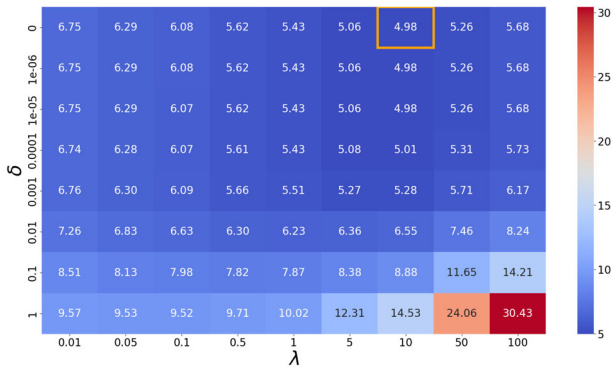


Fig. 5 LOOCV YTM RMSE for λ and δ . Based on daily LOOCV the YTM RMSE for a specific grid of λ , α and δ is shown. The figure fixes the optimal α and shows the two dimensional heatmap varying only λ and δ . The orange square indicates the lowest YTM RMSE for corresponding hyperparameters. Based on the long sample

Do the hyperparameters capture relevant economic information about the discount curve? If that were the case, we would see systematic patterns in the time series of the local optimal hyperparameter values. Figures 6, 7 and 8 show that this does not seem to be the case. These plots compare local optimal hyperparameters over time against global optimal values. Local optimal ones might change on a daily basis while global optimal ones remain fixed. Local optimal values are plotted as blue dots along with their medians.⁹ For α , which is on a linear scale, we also show the mean. We also add the global optimal values. The lighter the blue dots the less constant the local optimal value is. The dashed vertical black lines indicate dates, where a new bond was added to the universe. We have also compared the local optimal values over time against some common economic indicators, including the SNB policy rate, curve steepness or GDP.

⁹ In the online appendix we show histograms of each hyperparameter distribution.

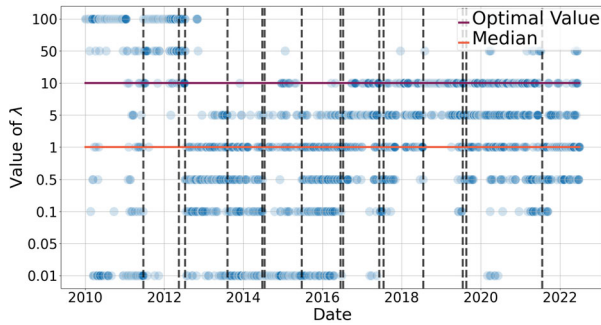


Fig. 6 Global vs local optimal hyperparameters for λ . The figure shows the local optimal values for λ over time and the corresponding median. We show no mean due to the logarithmic scale. The black dashed lines indicate the points in time where a new bond became available in the sample. Based on the long sample

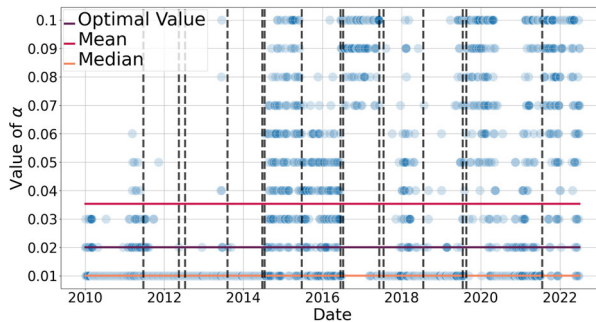


Fig. 7 Global vs local optimal hyperparameters for α . The figure shows the local optimal values for α over time and the corresponding mean and median. The black dashed lines indicate the points in time where a new bond became available in the sample. Based on the long sample

For none of them we found any systematic pattern. We conclude that the fluctuations in the local optimal hyperparameter values are mainly due to noise, which speaks for the robustness of our method. All essential economic information of the bond market in turn is captured by the KR curve.

To gauge how well the global optimal solution performs over time against the local optimal solution, we plot daily YTM RMSEs of the two methods based on LOOCV in Fig. 9. By construction, the local optimal errors are smaller than the global optimal ones. The difference between the errors appears to be larger in the first half of the sample. However, the magnitude is in the order of less than 5bps. Overall we find that the global optimal solution matches closely the YTM RMSE of the local optimal solution on a daily basis. There are no significant outliers in the differences in YTM RMSE. This again speaks for the robustness and stability of our method, which is based on global setting of baseline values for the hyperparameters.

Figure 9 also reveals some spikes of YTM RMSEs at the end of Q1 in 2020. This is a period of extreme market turmoil due to the Corona virus. The online appendix takes a closer look, showing that the spikes are due to outliers at the longer end of the term structure.

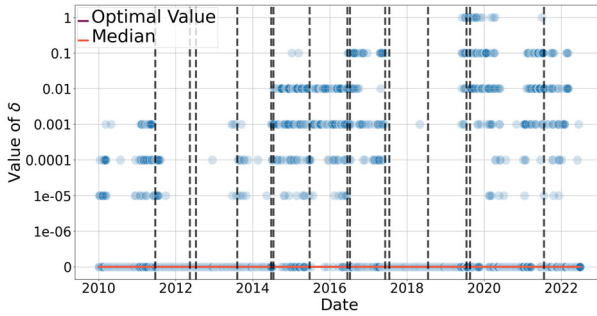


Fig. 8 Global vs local optimal hyperparameters for δ . The figure shows the local optimal values for δ over time and the corresponding median. We show no mean due to the logarithmic scale. The black dashed lines indicate the points in time where a new bond became available in the sample. Based on the long sample

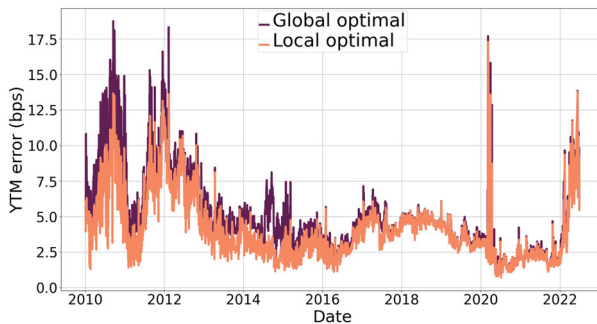


Fig. 9 Global vs local optimal solution - LOOCV. The figure shows the out of sample (LOOCV) YTM RMSE for the global optimal and local optimal hyperparameter values of the KR method. Based on the long sample

4.4 Example days

To better understand the impact of different choices of values for the hyperparameter λ , α and δ we plot on each example day the resulting yield curve as a function of one hyperparameter. In each figure we set the non-varying hyperparameters to the global optimal value found via LOOCV. Yield curves are shown up to 50 years. This goes slightly beyond the longest available maturity on any day. The latter is indicated with a vertical dashed red line. This representation is motivated as 50 years is the maximal time to maturity available (on one single day) in the sample period and FINMA provides yields up to 50 years for its SST curves. Figures 10 details the impact of different values for the hyperparameters.

Since λ acts as a smoothing parameter the larger the value the smoother the resulting yield curve, see Figs. 10a, d, g and j. The optimal value of δ is found to be 0, so that the smoothness penalty term only involves the second derivative of g in (3). It illustratively shows the trade off between pricing error minimization and smoother curves in terms of the norm $\|\cdot\|_{\alpha, \delta}$.

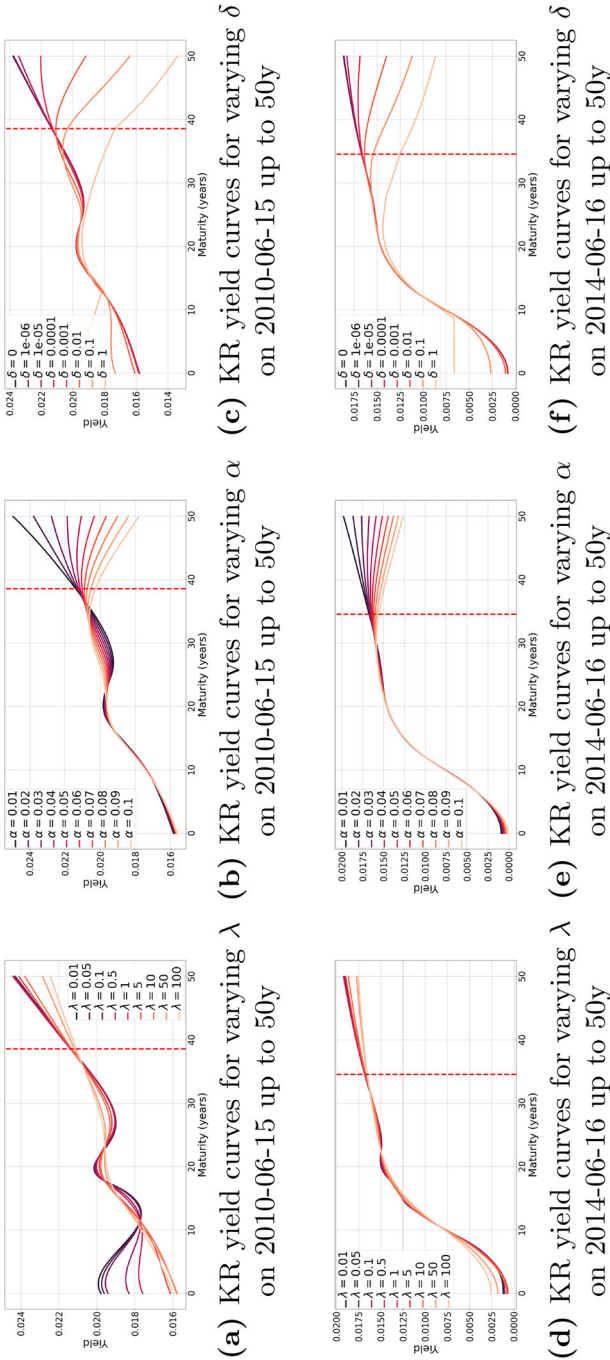


Fig. 10 KR yield curves for varying hyperparameter choices. The figures show the impact of varying λ , α and δ on the KR method. Resulting yields are drawn up to 50 years. The vertical dashed red line indicates the beginning of the extrapolation to the right. Based on the long sample

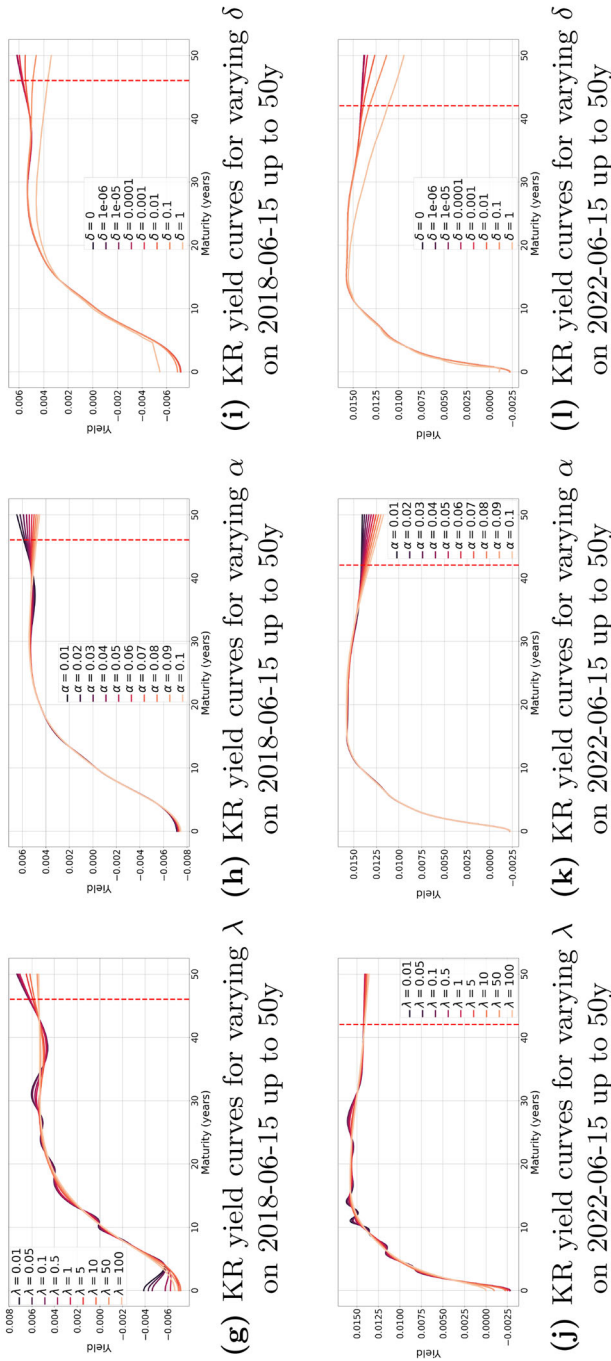


Fig. 10 continued

Figures 10b, e, h and k show the impact of varying α . Compared to λ , α affects the curve mainly at longer maturities within the extrapolation range. This is also consistent with the aforementioned link of α to the infinity maturity yield.

Figures 10c, f, i and l show the impact of varying δ . As describe above, the optimal value for δ is 0 leading to smoother curves. A large value for δ assigns a larger weight to the first derivative in (3). This results in more kinks and less smooth curves when compared to smaller values of δ .

5 Comparison study

After the selection of the base model, we now compare the KR method with the current standard models in the Swiss market. We apply the same evaluation metrics defined in Sect. 4.2 that were used to determine the optimal hyperparameters. First, we introduce the most common benchmark methods in more detail. We then present a sophisticated fitting error analysis, which clearly shows that our KR method performs best in all criteria. We also compare the yield time series of the different methods, and look at particular features on the example days across methods.

5.1 Benchmark methods

The current standard benchmark is from the SNB. The SNB itself fits a Nelson–Siegel–Svensson (hereafter referred to as NSS) yield curve

$$y^{NSS}(x) = B_0 + B_1 \left(\frac{1 - e^{-\frac{x}{T_1}}}{\frac{x}{T_1}} \right) + B_2 \left(\frac{1 - e^{-\frac{x}{T_1}}}{\frac{x}{T_1}} - e^{-\frac{x}{T_1}} \right) + B_3 \left(\frac{1 - e^{-\frac{x}{T_2}}}{\frac{x}{T_2}} - e^{-\frac{x}{T_2}} \right)$$

and publishes estimated parameters B_0, B_1, B_2, B_3 and $T_1, T_2 > 0$ on a daily basis, [20, 22]. A technical documentation regarding the estimation of these parameters was also published in [2]. In short, the SNB uses a classical NSS [17, 25] with parameter constraints to match the prevailing short rate r_{short} . Concretely, they set

$$B_0 + B_1 = r_{\text{short}}. \tag{13}$$

Until the end of 2020, they set r_{short} to the LIBOR spot next, and as of 2021, they set r_{short} to the SARON 1 month-swap rate. This way, SNB creates an additional anchor point at the short end of the yield curve, which is in contrast to the KR model.¹⁰ As explained in Example 2.3, we could easily modify the KR method to include such an anchor point as well. The resulting KR curves with SNB constraint are analyzed in the appendix.

Below, we also compare to our own implementation of NSS, which we refer to as “NSS”. We use a similar objective function for parameter estimation as in (2), $\sum_i \omega_i (P_i - \hat{P}_i^{NSS})^2$, where \hat{P}_i^{NSS} is the implied price of security i using the estimated NSS discount curve \hat{g}_t^{NSS} at time t . In this setting we use the duration weights

¹⁰ SNB also used the 3 months, 6 months, and 12 months LIBOR as additional data points until the end of 2020, and the SARON 3 months-swap rate instead of LIBOR from 2021 on.

ω_i to minimize the approximated YTM errors. We do not apply any constraint at the short end of the yield curve for parameter estimation like the SNB. The NSS curves are parsimonious parametric, and parameter estimation boils down to a highly non-convex optimization problem. To guarantee numerical convergence in our NSS implementation we use different standard solvers available, e.g., BFGS.

We also compare to our own implementation of the SST method as described in Sect. 3.2, which we refer to as “SST”. In this way, we calculate daily SST curves as of 2012. We back tested and compared our own calculated SST curves with the published annual FINMA SST curves, and we found that we can replicate the FINMA curves exactly up to the basis point. Since the SST curves are known to have been biased towards a relatively large UFR during the low interest regime, we do not report them in all performance comparisons. We mainly include them to compare the shapes of the resulting yield curves.

All metrics to compare KR, SNB, SST, and our own NSS, are calculated on a daily basis. We distinguish between in- and out-of-sample errors. For the former, we use all available data from the same day for estimation and evaluation. For the latter, we estimate KR and NSS on any given day and take the available SNB NSS parameters. Evaluation is then performed on the next following business day. The underlying assumption is that the yield curve does not change significantly over the course of one day. We use this procedure because for the SNB curves we only have access to their estimated model parameters so that a cross-validation within the same day is not feasible.

5.2 Fitting error

For the performance comparison, we use the YTM error, duration weighted error and relative price error, which we introduced in Sect. 4.2. By definition the duration weighted error should closely match the YTM error as a first order approximation. This is confirmed in the results. In this section the SNB method is the only benchmark method in scope. We use here the short sample to be as much aligned as possible to the universe the SNB used while fitting their model parameters.

Figure 11 displays the aggregated errors by maturity bucket, both in- and out-of-sample. On each day we split the available bonds into the corresponding maturity buckets. The respective error of each bond is then assigned to this bucket. We perform an average per day per bucket to derive a daily average error type per bucket. We then average these daily averages over time to obtain aggregated numbers. As we can see our KR estimates outperform the SNB in each maturity bucket for all error types in- and out-of-sample. In-sample errors show a similar pattern as the out-of-sample errors on a lower absolute level. As shown in Fig. 1, bond data for the maturity < 1 year bucket is very scarce. There are periods in the short sample during which this bucket is empty.

We also provide the daily mean of the out-of-sample error for each bucket over time in Fig. 12. This time series view confirms that the KR method is outperforming the SNB consistently. There are no time periods in which KR would systematically

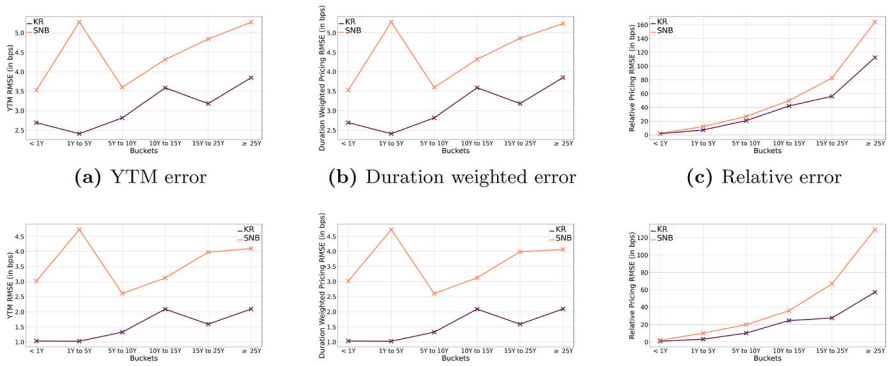


Fig. 11 In- and out-of-sample error comparison. The figures show the YTM error, duration weighted pricing error and the relative pricing error per maturity bucket aggregated over time in bps. The first row shows out-of-sample while the second row in-sample errors. Based on the short sample

underperform the SNB in any bucket. Similar results hold for the long sample, as shown in the online appendix.

5.3 Yield time series

In this section, we study the time series of fixed points on the estimated yield curve up to 30 years, which are within the maturity range that is covered by the bond data. Below we also show the time series of yields with larger maturities. Note that very short matured yields are mainly estimated from the coupon cash flows of the bonds as we use the long sample. In contrast to the previous section we also include SST yields in this comparison.

Figure 13 shows time series of the 1 month, 1 year, 10 years and 30 years yield in the left column. In the right column we show the rolling volatility of the yield estimates. Let $\hat{y}_t(x) = y_t^{\hat{g}}(x)$ denote the estimated yield with maturity x at time t derived from the estimated discount curve \hat{g}_t . We then define the rolling volatility as square root of the realized quadratic variation

$$\sigma_t(x) = \sqrt{\frac{252}{L} \sum_{s=0}^{L-1} (\hat{y}_{t-s}(x) - \hat{y}_{t-s-1}(x))^2},$$

where L refers to the lookback measured in business days (and we assume a year has 252 business days). Here we set $L = 21$, which is 1 month lookback.

For the 1 month yield in Fig. 13a we see a large discrepancy between the KR and the other two yields, which is due to missing short maturity bonds in the long sample and the additional constraint imposed at the short end used by SNB (and thus inherited by SST). Even for the two benchmark methods that use these additional anchor points we see some questionable spikes, which get also fed through the rolling volatility plot in Fig. 13b. A similar picture still emerges for the 1 year yield in Fig. 13c and d.

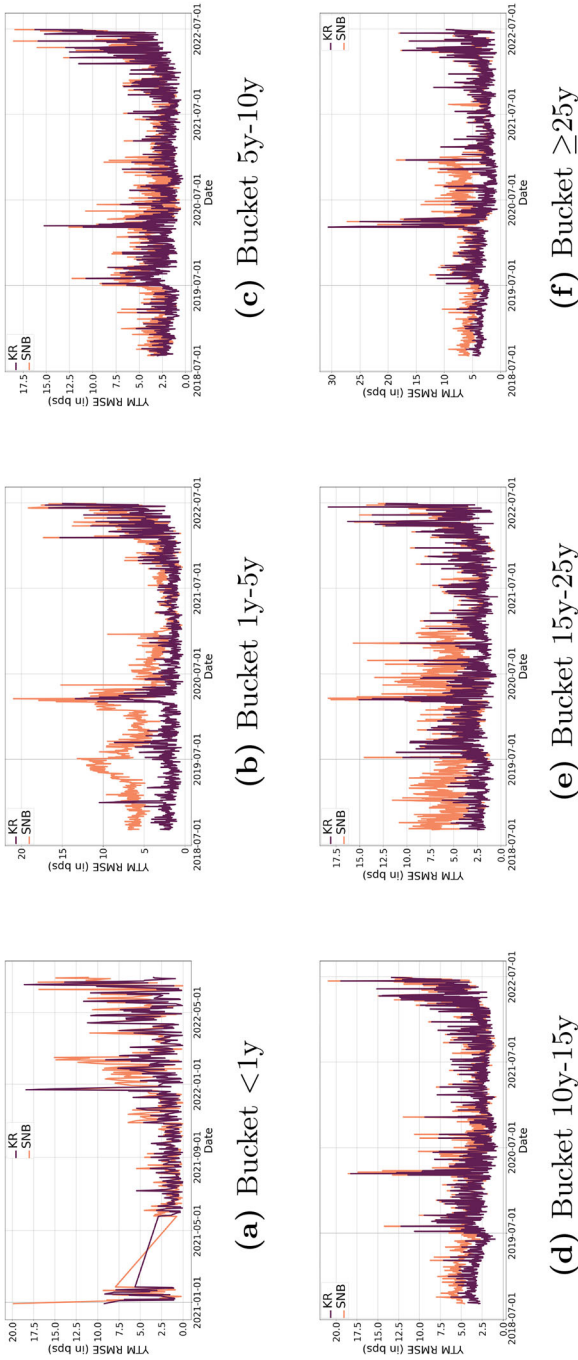


Fig. 12 Time series YTM errors per maturity bucket. The figures show the YTM error as daily time series per maturity bucket. In all figures out of sample errors are used. Straight lines in **a** indicate a period during which bucket < 1 year is empty. Based on the short sample

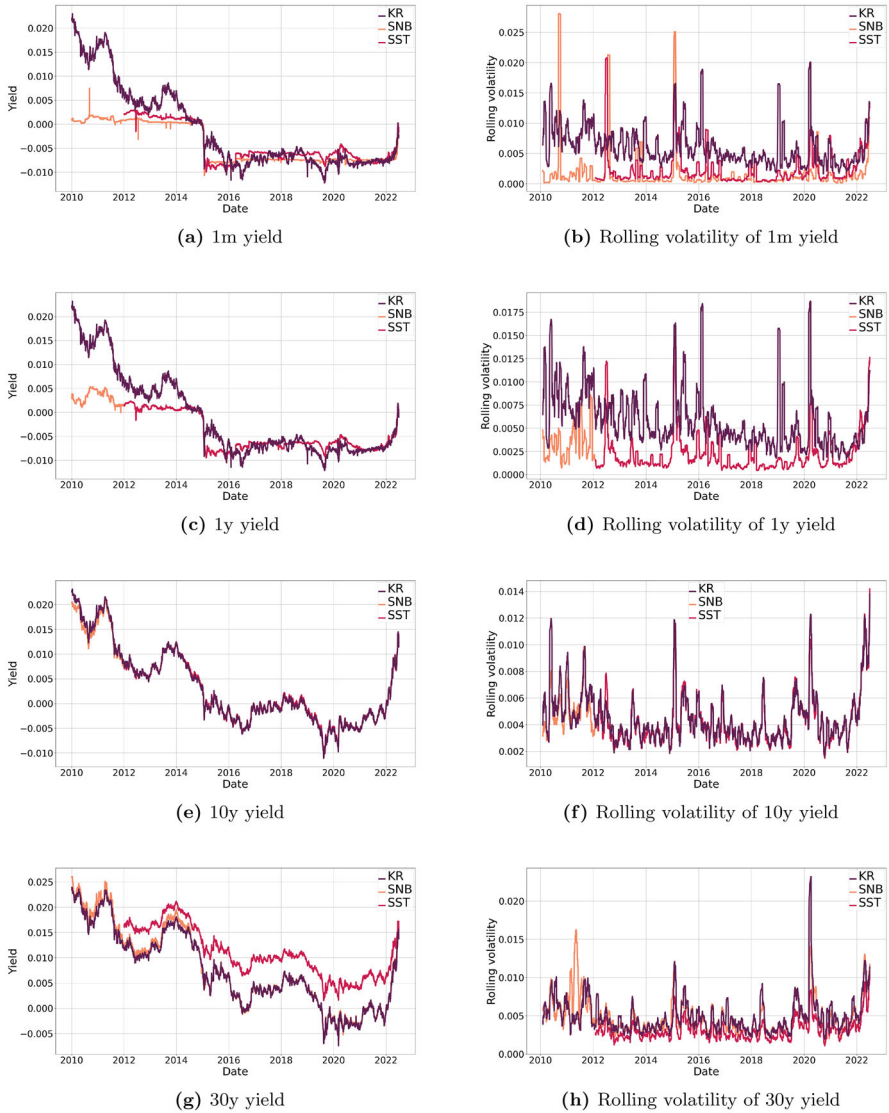


Fig. 13 Yield time series and rolling volatilities. The figures show the constant 1 month, 1 year, 10 years and 30 years yield time series on the left hand side and the respective rolling volatility on the right hand side. Based on the long sample

Remarkably, the KR yield estimates for 1 month and 1 year are close to SNB and SST in the second half of the sample, despite the fact that KR is based entirely on bonds with maturities way beyond one year, which seems to indicate that bond and money markets are integrated.

The longer dated yields, e.g., 10 years, in Fig. 13e and f, behave similarly across methods. The same observation holds for the 30 years yield in Fig. 13g except for

SST. This is not surprising because beyond its LLP of 15 years, the SST curve lies systematically above KR and SNB during the low interest rate environment. This is due to the exogenous choice of the UFR, which is larger than the market yields at the long end. However, the repricing of the interest rate market towards the end of the sample period is such that the gap almost disappears between KR, SNB and SST for the 30 years yield in Fig. 13g. As a sanity check, we also observe that the SST yield perfectly aligns SNB for the 1 year and 10 years point. The online appendix contains the time series for additional maturities up to 40 years.

In summary, we find that the level and volatility of the yield time series of the KR and SNB methods are similar for maturities between 5 and 40 years. Beyond 40 years, we see differences in the first part of the sample, before the introduction of the 50 years bond in 2014. Moreover, no periodic pattern is observed in the level or volatility of the yield time series. In particular, there are no visible year-end effects.

5.4 Example days

All estimation methods lead to smooth yield curves on the example days shown in Fig. 14. On 2010-06-15 we can observe the additional short maturity bonds and the constraint to match the prevailing short term rate (at that time it was CHF LIBOR) for SNB. Upon availability the SST curves indirectly use this constraint, too. The input parameter for the estimation of the SST curves are the estimated yields from the SNB. Thus, by adding this short term rate constraint to the SNB it gets automatically feed through SST. The KR method only uses coupon cash flows of longer maturity bonds to estimate the discount curve on the short end in the long sample. However, it is remarkable that already in 2014, where shorter maturity bonds were still missing, the KR closely matches the SNB and SST curves below 10 years. It should be kept in mind that no bond with maturity less than 1 year is available in the long sample before 2022. Thus, the existence of a bond in 2014 with time to maturity of approximately 8y already increases the goodness of the fit (assuming the additional anchor points used by the SNB are valid proxies to short term government bond yields). To some extent this is almost an extrapolation exercise (only coupon cash flows are available) on the short end of the curve. We have already observed this behaviour in Fig. 13.

Estimating the NSS parameters is a highly non-convex problem. In our own implementation of NSS we tested different standard solvers. We found that the estimates depend significantly on the seeds of the optimizers.

To visualize this issue, we have included in Fig. 14 our own implementation of NSS. We plot ten different curves, which result from slightly modified initial values of the optimization algorithms. Concretely, we took the SNB NSS parameters prevailing at that day and randomly perturbed each one by multiplying by $\exp(0.2 \cdot Z)$, where Z follows a standard normal distribution. The perturbed parameter values were entered into the optimizer as initial values. The resulting curves are significantly spread for maturities less than 10 years, and remarkably so at the long end for the first sample day. In summary, we find that NSS curves are hardly reproducible, which is due the non-convexity of the estimation problem.

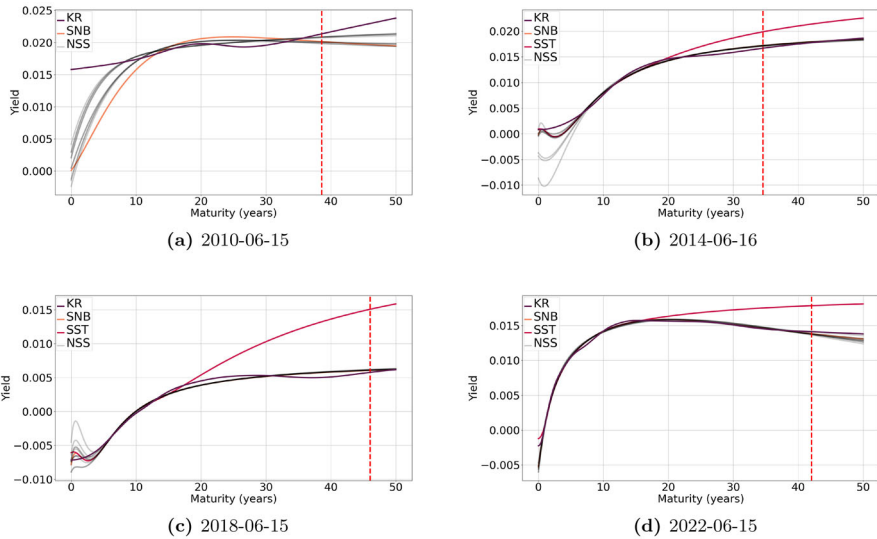


Fig. 14 Yield curve method comparison on example days. The four figures show the KR, SNB, NSS and the SST curve where applicable for the example days. The grey lines are our own implementation of NSS where we have slightly perturbed the initial values of the optimization algorithm (ten times). The vertical dashed red line indicates the beginning of the extrapolation to the right. Based on the long sample

Figure 14c also shows a large discrepancy between the SST and the other curves at the long end. This is due to the exogenous choice of the UFR during the low interest rate regime. At the end of the sample period, this gap has significantly narrowed as fixed income markets have undergone an aggressive repricing of interest rates. Figure 14 also confirms that the SST and SNB curves coincide at the maturity points 1 year, ..., 10 years and 15 years, by construction, on all four example days. After the LLP of 15 years, the curves diverge quickly as the SST's remaining anchor point is the UFR while the KR and SNB curves are based on longer maturity bonds' prices.

6 Extrapolation

So far we have focused on the time span up to 50 years. During most of the long sample period this is close to the longest maturity bond available in the sample universe. In this section we examine the behaviour and comparison in the extrapolation range beyond 50 years. We sketch results up to 100 years, which may be of particular interest in the actuarial science where, e.g., long-term liability cash flows need to be discounted. Thus, a reliable and robust curve estimation method is of utmost importance. However, any extrapolation of the Swiss discount curve is subject to great uncertainty, as the longest maturity of Swiss government bond is less than 50 years. This is the case in most comparable bond markets. Further below, we provide an outlook on ongoing research that addresses the challenge of long-term extrapolation.

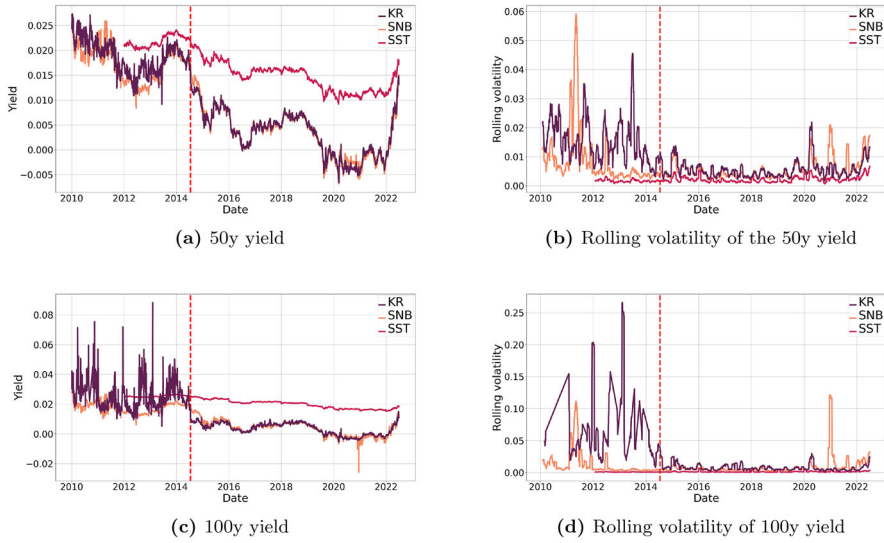


Fig. 15 Longterm yield time series and rolling volatilities. The figures show the 50 years and 100 years yield time series on the left hand side and the respective rolling volatility on the right hand side. The red dash line indicates the issuance day of the Swiss government bond with longest maturity (2064-06-25). Based on the long sample

6.1 Yield time series

To better understand the behaviour of extrapolated yields we extend the analysis of Sect. 5.3. Here, we focus on yields that lie far in the extrapolation area, namely 50 years and 100 years.

Figure 15 shows the time series of these yields and their 1 month rolling volatilities. The yields in the left column once again show the artificially high UFR for the SST curve. We observe large differences in the first part of the sample for absolute levels and rolling volatility, prior to the introduction of the 50 years bond in June 2014. The volatility of the KR yield time series drops significantly and KR and SNB yield levels match closely after that date.

The time series of the SNB yield for 100 years exhibits some extreme spikes in late 2020 and early 2021, which are also captured by the rolling volatility. Figure 16 takes a closer look and shows the yield curves on some of these extreme days. The first row includes the SST while the second omits it for better visualization. The SNB yield curves are visibly downward biased in the extrapolation region, which is an artifact of their rigid parametric form.

6.2 Example days

Figure 17 shows the impact of varying values of the hyperparameters λ , α and δ on the extrapolated curves. These are extended plots from Fig. 10 for the same example days. The effects described in Sect. 4.4 are magnified in the extrapolation region. In

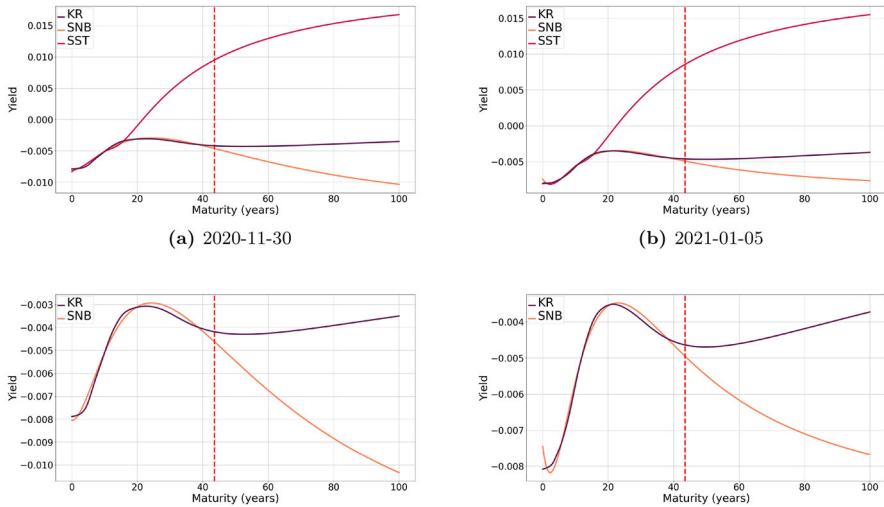


Fig. 16 Extreme SNB NSS forecast. The figures show the yield curves for the example days that lead to the large increase in the rolling volatility for the 100 years yield for the SNB NSS in Fig. 15d. The first row includes the SW SST curve while the second row only shows KR and SNB NSS to better visualize the difference of the two. The vertical dashed red line indicates the beginning of the extrapolation to the right. Based on the long sample

particular, the choice of α has a much more pronounced impact on the yield curve beyond 50 years. Some of the extrapolated yield curves diverge. This is because KR is a linear estimator of the discount curve, which can become negative in the extrapolation region. The yield curve is a logarithmic transform of the discount curve and therefore explodes when the discount curve approaches zero. The extrapolated curves behave well for the last two example days, after the introduction of the 50 years bond in the sample.

Figure 18 shows the extrapolated yield curves for all methods. These are extended plots from Fig. 14. Also here, all effects described in Sect. 5.4 are magnified in the extrapolation region. Notably, our own NSS curves exhibit a wide spread around the SNB curve beyond 50 years. This again highlights the non-reproducibility of the NSS estimates due to the critical non-robustness of the NSS method with respect to the choice of initial parameters in the optimizer.

We conclude that none of the methods in scope can provide reliable and robust extrapolations of the yield curve. Extrapolation is a choice and depends on additional assumptions. Since many actuarial and other applications require yield curves with horizons up to 100 years and beyond, we outline here two possible approaches to obtain such extreme extrapolations.

The first approach is based on a multi-curve extension of the KR method. Hereby, one jointly estimates the discount curves of several markets, including fixed income markets with longer dated instruments. The method learns similarities between different market curves, by regularizing their spreads, and thus provides additional anchor points for specific markets (segments), where data quality is poor or not existing at all. For example, the Austrian government bond market has a bond outstanding with

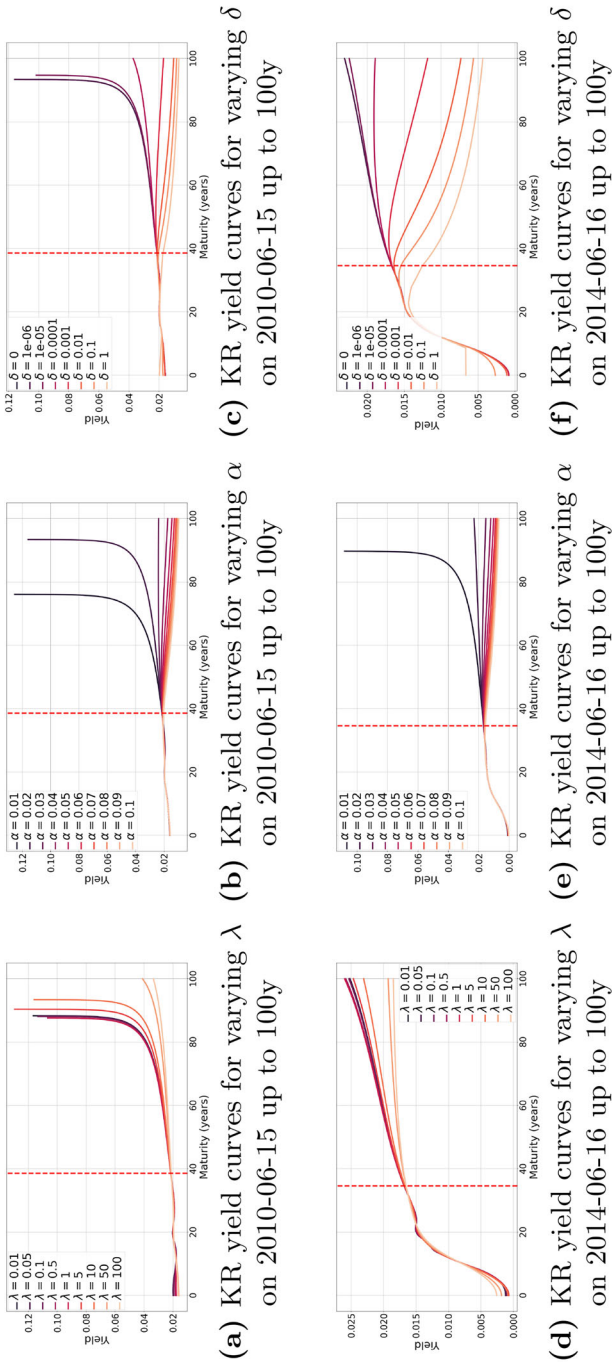


Fig. 17 KR yield curves for varying hyperparameter choices beyond 50 years. The figures show the impact of varying λ , α and δ on the KR method. Resulting yields are drawn up to 100 years. The vertical dashed red line indicates the beginning of the extrapolation to the right. Based on the long sample

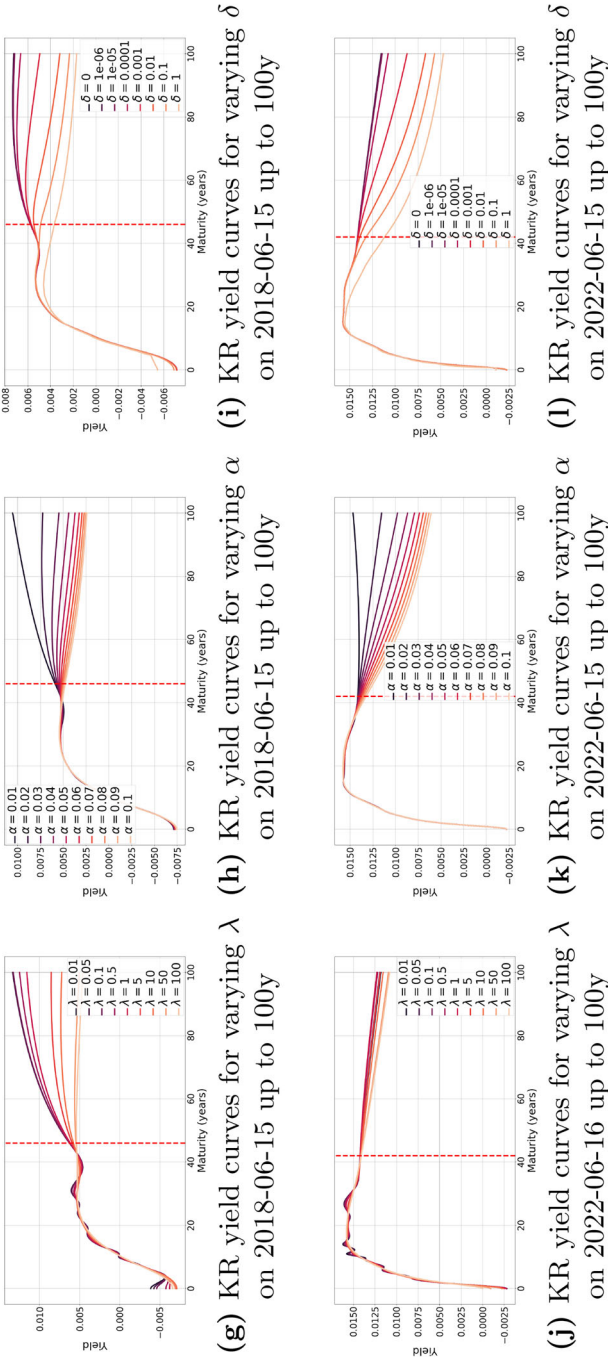


Fig. 17 continued

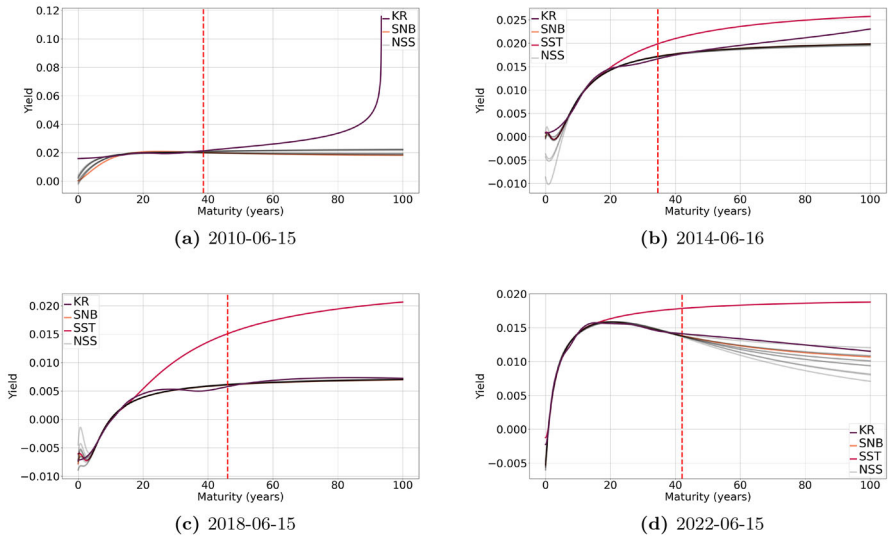


Fig. 18 Yield curve method comparison on example days. The four figures show the KR, SNB, NSS and the SST curve where applicable for the example days. The grey lines are our own implementation of NSS where we have slightly perturbed the initial values of the optimization problem (ten times). The vertical dashed red line indicates the beginning of the extrapolation to the right. Based on the long sample

a maturity of June 30, 2121.¹¹ A joint estimation of the Swiss and Austrian discount bond curves benefits the quality of the Swiss curve in the extrapolation region. Not only other government bond markets can be used but also similar instrument markets, e.g., swap markets. This is work in progress, see [3].

The second approach is based on a dynamic arbitrage-free interest rate model of choice. In its simplest form, this could be a constant short rate $r_t \equiv r$. A more flexible and economically reasonable model is, e.g., the two-factor Gaussian affine model for the short-rate process r_t with stochastic mean-reversion level γ_t , as introduced and estimated in [11]. The model parameters, say θ , can be efficiently estimated using a past sample of bond data. Discount bond prices in this model are given in closed form $g_{r_t, \gamma_t, \theta}(x)$ depending on the prevailing values r_t, γ_t and the parameter θ . We can then extrapolate the KR curve $g(x)$ beyond the last observed maturity x_N by setting

$$g^{extra}(x) = g(x_N) \cdot g_{r_{x_N}, \gamma_{x_N}, \theta}(x - x_N), \quad x > x_N. \tag{14}$$

Under the hypothetical assumption that the future values r_{x_N}, γ_{x_N} are known today, this extension yields an arbitrage-free and well-behaved discount curve for all $x \geq 0$, see [8, Section 2.2.3]. The model-based extrapolation (14) is fully transparent and explainable. In fact, the role of the model parameters θ is well known. A plausible choice of the future short rate is to set $r_{x_N} = -g'(x_N)/g(x_N)$ to be equal to the forward

¹¹ Source: Bloomberg Finance L.P.

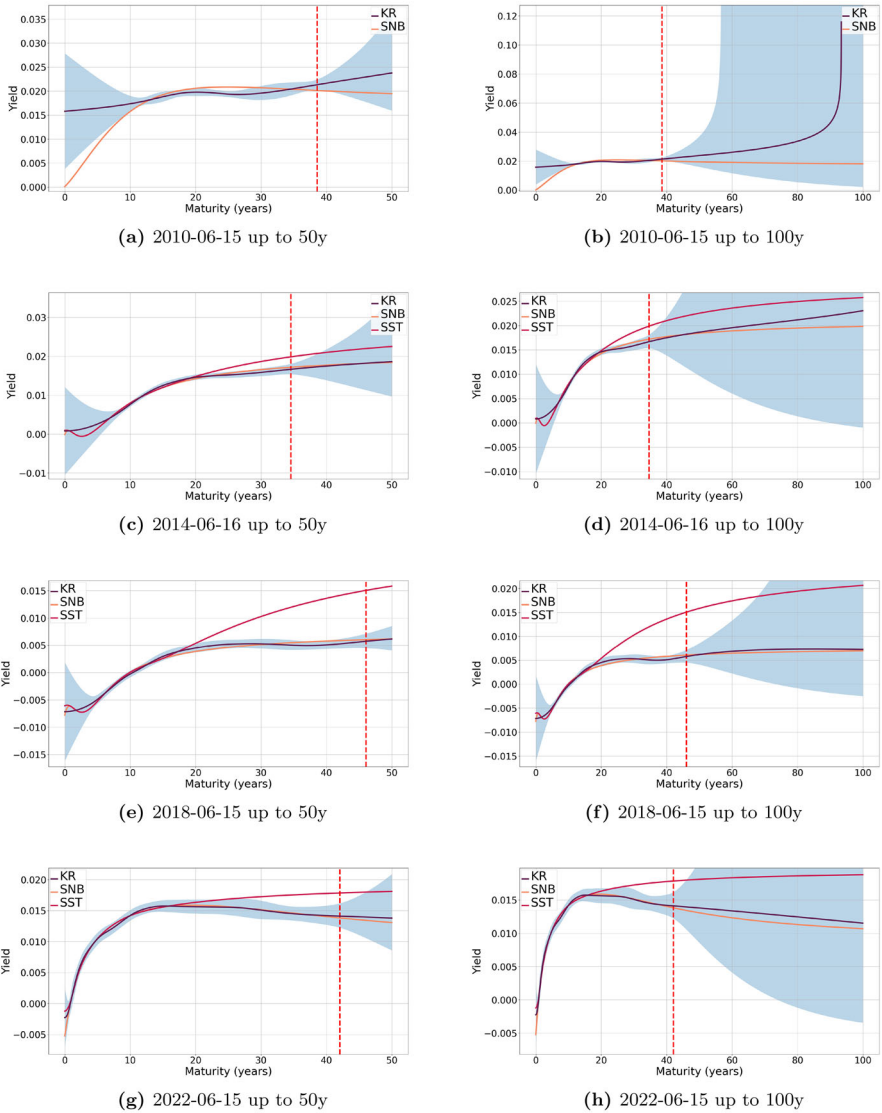


Fig. 19 KR 3σ -confidence bands on example days. The figure shows yield curve estimates and confidence bands (3σ) based on the KR method. Resulting yields are drawn up to 50 years on the left and up to 100 years on the right. The vertical dashed red line indicates the beginning of the extrapolation to the right. Based on the long sample

rate implied by the KR curve at x_N . This gives a smooth pasting such that $g^{extra}(x)$ is twice weakly differentiable. The future mean-reversion state γ_{x_N} can be set equal to its risk-neutral mean-reversion level, which reflects risk-neutral stationarity. This is work in progress.

7 Statistical inference

There is a well known correspondence between kernel ridge regression and Gaussian processes allowing for a Bayesian interpretation, see [10, Section 2.4] for more details. Here we assume that the discount curve g is a Gaussian process with mean function $m : [0, \infty) \rightarrow \mathbb{R}$ and covariance given by the kernel k . That is, $g(\mathbf{x}) \sim \mathcal{N}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}^\top))$. We also assume that the pricing errors in (1) are independent centered Gaussian random variables, $\epsilon \sim \mathcal{N}(0, \Sigma^\epsilon)$, with $\Sigma^\epsilon = \text{diag}(\sigma_1^2, \dots, \sigma_M^2)$. The posterior distribution of g given the observed prices P is again Gaussian with posterior covariance function k^{post} given by

$$k^{\text{post}}(y, z) = k(y, z) - k(y, \mathbf{x}^\top)C^\top(C\mathbf{K}C^\top + \Sigma^\epsilon)^{-1}Ck(\mathbf{x}, z).$$

If the prior mean function is $m(x) \equiv 0$, and the pricing error variances equal $\sigma_i^2 = \lambda/\omega_i$, then the posterior mean function is equal to the KR estimate \hat{g} . We can now use the posterior covariance function to compute confidence intervals around \hat{g} .

We illustrate this for the example days. Figure 19 shows KR yield curves with corresponding 3σ -confidence bands, along with the SNB and SST curves, with and without extrapolation. The wide confidence bands indicate regions with scarce or missing data (short end) and price dispersion (middle ranges). In fact, it is remarkable how the KR method detects the range of missing data and adequately estimates a confidence band. Extrapolation regions exhibit large uncertainty which is reflected and quantified by the wide and expanding confidence bands. This uncertainty can be drastically reduced by applying the multi-curve extension using debt markets that exhibit longer dated bonds, see [3]. It is worth noting that SST curves sometimes lie outside the 3σ -confidence bands, reflecting their bias towards the UFR.

8 Conclusion

An accurate and robust estimation of a discount curve is of vital importance for academic, industry, and regulatory purposes. The KR method developed in [10] proves to satisfy all desirable characteristics of a preferred estimation method. (i) KR is simple and fast to implement. The estimation boils down to a simple kernel ridge regression. (ii) KR is transparent and reproducible. The kernel ridge regression admits a unique solution, which is given closed form and linear in the data. (iii) KR is fully data-driven. All hyperparameters are globally chosen by cross-validation. (iv) KR provides a precise representation of the term structure taking into account all market signals. It is a fully flexible non-parametric method trading off between minimal the fitting error and smoothness of the curve. (v) KR is robust to outliers and data selection choices. Rewarding smoothness of the curve renders the estimates robust. (vi) KR is flexible for integration of external views. The user can easily force single points of the curve to match exogenously given yields, for example, at the short end or in the extrapolation region. (vii) KR is consistent with finance principles. It reprices all fixed income instruments based on the law of one price, and the smoothness of the curve is moti-

vated by the economic principle of limits to excessive payoffs of trading strategies in bonds with nearby maturities.

We apply the KR method to the Swiss government bond market. We find that the KR method outperforms the SNB and SST benchmarks in all dimensions. Extrapolating the yield curve beyond the observed maturity range remains an open challenge. We propose two possible approaches, namely multi-curve learning and dynamic stochastic models, which will be the subject of future research.

This paper provides a technical input to the regulatory process to find a method to improve the current insurance industry standard Smith–Wilson. It also offers itself as a new method of choice for central banks.

I KR curves with SNB short end constraint

In this section, we analyze KR curves with the binding constraint that the implied short rate of the estimated curve match the prevailing short rate r_{short} , as explained in Example 2.3. Concretely, we apply the same constraint as used by the SNB curves, which is encoded in the NSS parameters (13) whose values we retrieve from [20]. We refer to the KR curves with this SNB type constraint as “KR–SNB” in the following.

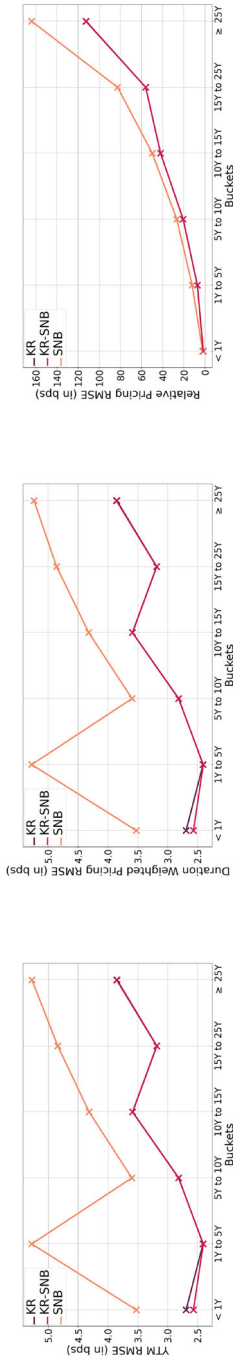
Figure 20 shows that the out-of-sample fitting errors of KR–SNB curves are essentially the same as for the unconstrained KR curves. In fact, a slightly larger error is only observed in the first maturity bucket < 1 year. This suggests that the KR curves are very robust to local data perturbations.

This local robustness of the KR method is confirmed by the curves on the example days shown in Fig. 21. In fact, KR–SNB and SNB curves coincide at the short end by construction. This is particularly true for 2010-06-15, where the absence of short-dated bonds was most pronounced, demonstrating the remarkable benefit of the additional anchor point. The KR–SNB curves then quickly converge to the unconstrained KR curves. Notably, the characteristic wiggles of the SNB and NSS curves in the < 5 year maturity range are not present in the KR–SNB curves. This suggests, once more, that these wiggles are due to the functional rigidity of NSS and do not have any economic content.¹²

The functional misspecification of the NSS curves also reveals itself at the long end. Figure 22, which corresponds to Fig. 16, shows that the KR–SNB curves coincide with the unconstrained KR curves in the extrapolation range also on some specific days where the SNB curves are ill-behaved.

Figure 23, which corresponds to Fig. 13, shows the time series of yields of various maturities. It confirms that the KR–SNB curves interpolate the exogenous short end with the unconstrained KR curves and the convergence takes place in the range < 10 years. Notably, the 1 month yield time series of the KR–SNB curve is less spiky than the one of the SNB curve. One notable exception is on the 15 January 2015, when the SNB discontinued the minimum EUR/CHF exchange rate and lowered policy rate to -75 bps. This economic shock is consequently also captured by the KR–SNB curve.

¹² As mentioned in Sect. 5.1, SNB uses 3 months, 6 months, 12 months LIBOR and SARON 3 month-swap rates, respectively, as additional data points, which we do not and which may partly explain the differences between the SNB and KR–SNB curves in the < 1 year maturity range.



(a) YTM error

(b) Duration weighted error

(c) Relative error

Fig. 20 Out-of-sample error comparison. The figures show the YTM error, duration weighted pricing error and the relative pricing error per maturity bucket aggregated over time in bps. We only report out-of-sample errors. Based on the short sample

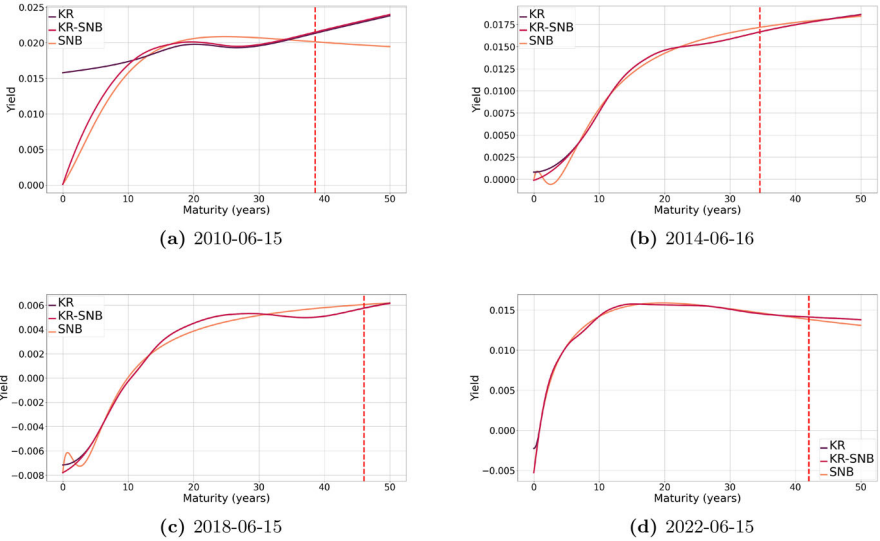


Fig. 21 Yield curve method comparison on example days. The four figures show the KR, KR–SNB and the SNB curve for the example days. The vertical dashed red line indicates the beginning of the extrapolation to the right. Based on the long sample

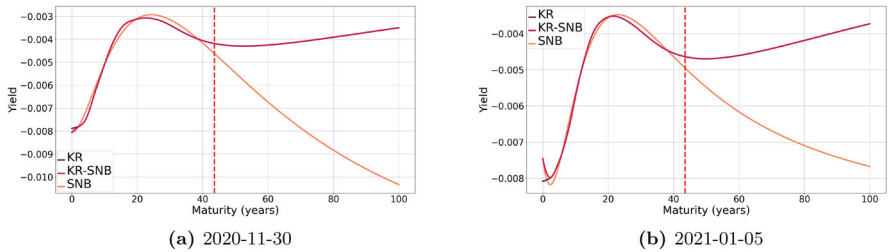


Fig. 22 Extreme SNB NSS forecast with SNB short end constraint. The figures show the yield curves for the example days that lead to the large increase in the rolling volatility for the 100 years yield for the SNB NSS in Fig. 15d. Here we include the KR–SNB constrained version, too. The vertical dashed red line indicates the beginning of the extrapolation to the right. Based on the long sample

Another exception is on the 8 September 2010, when the SNB parameters $B_0 + B_1$ reported on [20] spike up by 70 bps, which might be a data error for which we could not find any other explanation.

The short end constraint has an obvious impact on the confidence bands around the KR–SNB curves. There is no uncertainty at the short end. Again it is remarkable how the KR method adequately indicates the ranges of missing data. This is confirmed in Fig. 24, which corresponds to Fig. 19.

In summary, we find the KR–SNB curves provide a valuable alternative to the SNB curves, as they combine the short end constraint of the SNB method with all the advantages of the KR method.

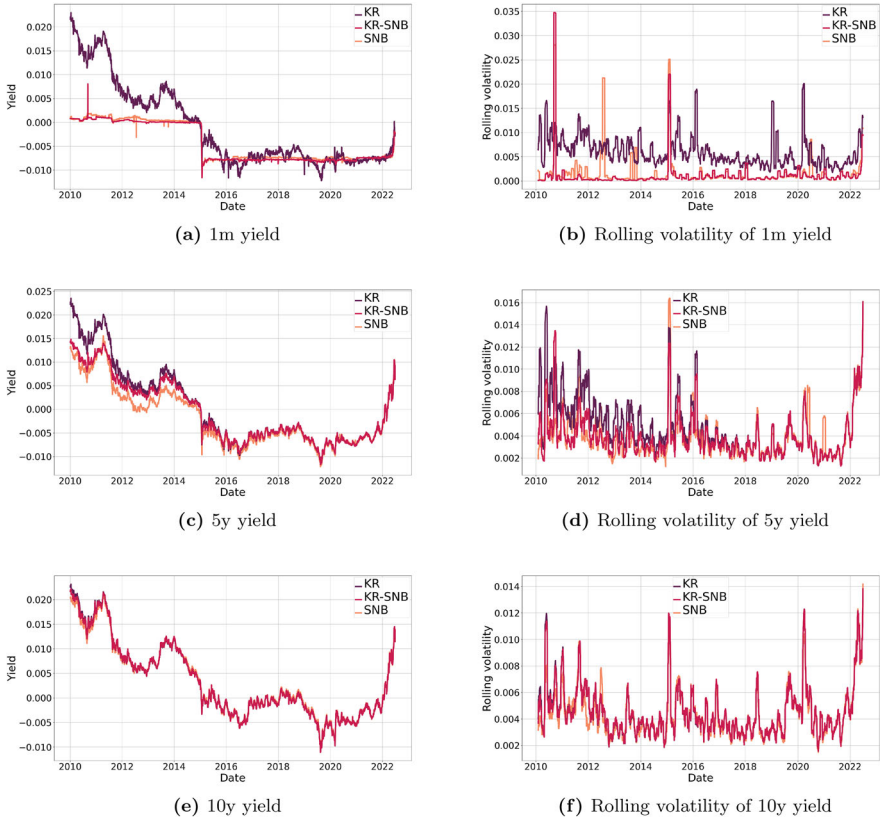


Fig. 23 Yield time series and rolling volatilities with SNB short end constraint. The figures show the constant 1 months, 5 years and 10 years yield time series on the left hand side and the respective rolling volatility on the right hand side. The vertical dashed red line indicates the beginning of the extrapolation to the right. Based on the long sample

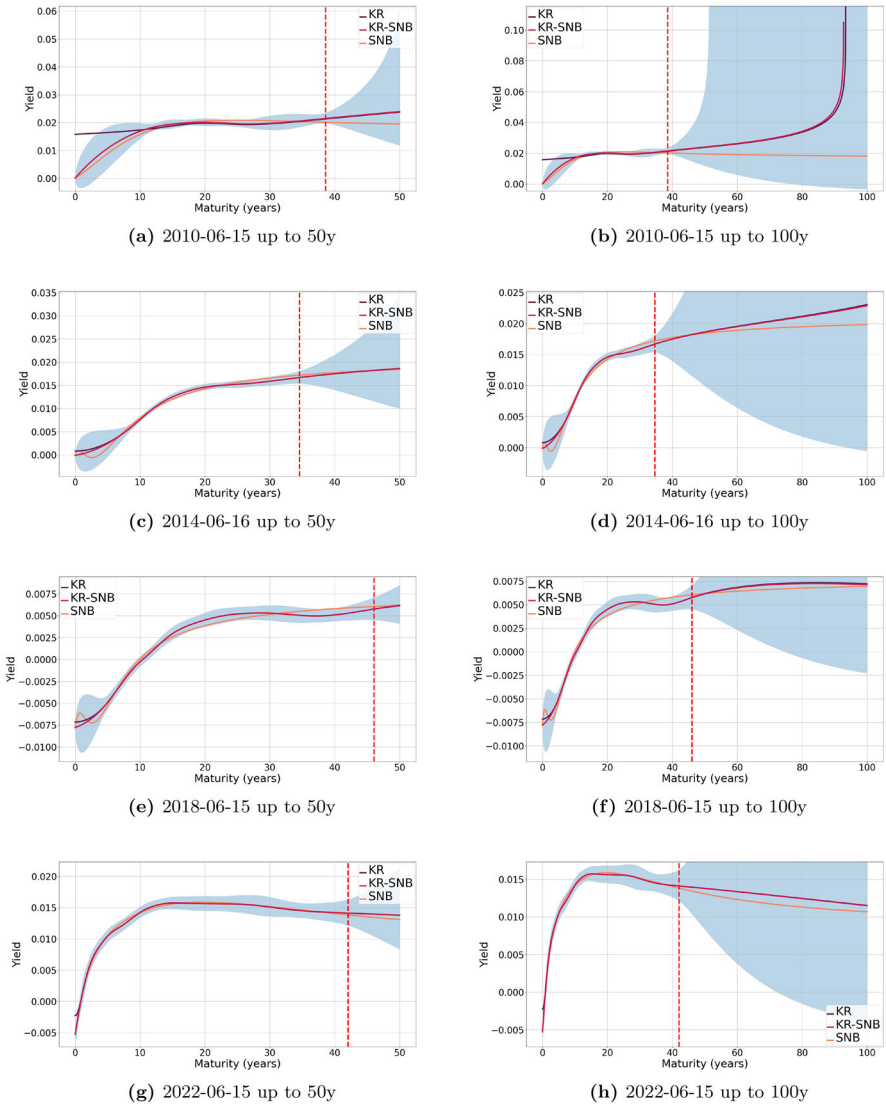


Fig. 24 KR 3σ -confidence bands on example days with SNB short end constraint. The figure shows yield curve estimates and confidence bands (3σ) based on the KR–SNB method. Resulting yields are drawn up to 50 years on the left and up to 100 years on the right. The vertical dashed red line indicates the beginning of the extrapolation to the right. Based on the long sample

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s13385-024-00386-4>.

Funding Open access funding provided by EPFL Lausanne.

Data availability For our empirical study we use public available data from the SNB [23].

Declarations

Conflict of interest The authors declare that they have no Conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Andersen L (2007) Discount curve construction with tension splines. *Rev Deriv Res* 10(3):227–267
- BIS (2005) Zero-coupon yield curves: technical documentation. BIS Papers 25, Bank for International Settlements. <https://www.bis.org/publ/bppdf/bispap25.htm>
- Camenzind N, Filipović D, Pelger M, Wang R (2023) Joint learning of international yield curves. In: Presented at SIAM conference on financial mathematics and engineering
- Christensen JHE, Mirkov N (2022) The safety premium of safe assets. Working Paper 2019–28. Federal Reserve Bank of San Francisco. <https://doi.org/10.24148/wp2019-28>
- EIOPA (2021) Technical documentation of the methodology to derive EIOPA's risk-free interest rate term structures. Technical report, European Insurance and Occupational Pensions Authority. https://www.eiopa.europa.eu/tools-and-data/risk-free-interest-rate-term-structures_en
- ESRB (2017) Regulatory risk-free yield curve properties and macroprudential consequences. Technical report, European Systemic Risk Board. https://www.esrb.europa.eu/pub/pdf/reports/esrb-reports170817_regulatoryriskfreeyieldcurveproperties.en.pdf
- Fama EF, Bliss RR (1987) The information in long-maturity forward rates. *Am Econ Rev* 77(4):680–692
- Filipović D (2009) Term-structure models: a graduate course. Springer finance. Springer, Berlin
- FINMA (2023) Swiss Solvency Test (SST). <https://www.finma.ch/en/supervision/insurers/cross-sectoral-tools/swiss-solvency-test-sst>. Accessed Sept 2023
- Filipovic D, Pelger M, Ye Y (2022) Stripping the discount curve—a robust machine learning approach. Swiss Finance Institute Research Paper No. 22–24. <https://ssrn.com/abstract=4058150>
- Filipovic D, Trolle AB (2013) The term structure of interbank risk. *J Financ Econ* 109(3):707–733
- Gürkaynak RS, Sack B, Wright JH (2007) The U.S. treasury yield curve, (1961) to the present. *J Monet Econ* 54(8):2291–2304
- Jørgensen PL (2018) An analysis of the Solvency II regulatory framework's Smith–Wilson model for the term structure of risk-free interest rates. *J Bank Financ* 97:219–237
- Lagerås A, Lindholm M (2016) Issues with the Smith–Wilson method. *Insur Math Econ* 71:93–102
- Liu Y, Wu JC (2021) Reconstructing the yield curve. *J Financ Econ* 142(3):1395–1425
- McCulloch JH (1971) Measuring the term structure of interest rates. *J Bus* 44(1):19–31
- Nelson CR, Siegel AF (1987) Parsimonious modeling of yield curves. *J Bus* 60(4):473
- Paulsen VI, Raghupathi M (2016) An introduction to the theory of reproducing kernel Hilbert spaces. Cambridge studies in advanced mathematics. Cambridge University Press, Cambridge
- SNB (2002) Changes and revisions—interest rates, yields and foreign exchange market. https://data.snb.ch/en/topics/ziredev/doc/changerev_ziredev
- SNB (2002) Nelson–Siegel–Svensson parameters. <https://data.snb.ch/en/topics/ziredev/cube/rendopar>
- SNB (2002) Quartalsheft 2/2002. https://www.snb.ch/de/mmr/reference/quartbul_2002_2_komplett/source/quartbul_2002_2_komplett.de.pdf
- SNB (2002) Notes—interest rates, yields and foreign exchange market. https://data.snb.ch/en/topics/ziredev/doc/explanations_ziredev. Accessed Sept 2023

23. SNB (2023) Price, yield and remaining period to maturity of individual Swiss Confederation bond issues. <https://data.snb.ch/en/topics/ziredev/cube/rendoeid>. Accessed Sept 2023
24. Schölkopf B, Smola AJ (2018) Learning with kernels: support vector machines, regularization, optimization, and beyond. The MIT Press, Cambridge
25. Svensson L (1994) Estimating and interpreting forward interest rates: Sweden 1992–1994. NBER Working Papers 4871. National Bureau of Economic Research, Inc
26. Smith A, Wilson T (2001) Fitting yield curves with long term constraints. Working paper
27. Tanggaard C (1997) Nonparametric smoothing of yield curves. *Rev Quant Financ Acc* 9(3):251–267
28. Vasicek OA, Gifford FH (1982) Term structure modeling using exponential splines. *J Financ* 37(2):339–348
29. Viehmann T (2019) Variants of the Smith–Wilson method with a view towards applications. Working paper

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.