RESEARCH

Swiss Finance Institute
Research Paper Series
N°22-24

Stripping the Discount Curve—a Robust Machine Learning Approach

Damir Filipović
Ecole Polytechnique Fédérale de Lausanne and Swiss Finance Institute

Markus Pelger
Stanford University

Ye Ye
Stanford University

Stripping the Discount Curve — a Robust Machine Learning Approach*

Damir Filipović[†] Markus Pelger[‡] Ye Ye[§]

This draft: March 10, 2023
First draft: March 15, 2022

Abstract

We introduce a robust, flexible and easy-to-implement method for estimating the yield curve from Treasury securities. Our non-parametric method learns the discount curve in a function space that we motivate by economic principles. We show in an extensive empirical study on U.S. Treasury securities, that our method strongly dominates all parametric and non-parametric benchmarks. It achieves substantially smaller out-of-sample yield and pricing errors, while being robust to outliers and data selection choices. We attribute the superior performance to the optimal trade-off between flexibility and smoothness, which positions our method as the new standard for yield curve estimation.

Keywords: yield curve estimation, U.S. Treasury securities, term structure of interest rates, non-parametric method, machine learning in finance, reproducing kernel Hilbert space
JEL classification: C14, C38, C55, E43, G12

*We thank Michael Bauer, Nicolas Camenzind, Darrell Duffie, Rüdiger Fahlenbrach, Kay Giesecke, Oliver Linton, Thomas Viehmann, and Sander Willems and seminar and conference participants at Stanford, ETH Zurich, Columbia University, Vienna University of Economics and Business, University of Waterloo, Swedish House of Finance, the World Online Seminars on Machine Learning in Finance, 11th World Congress of the Bachelier Finance Society, European Actuarial Academy online conference CONVENTION A, Hong Kong Conference for Fintech, AI and Big Data in Business, Swiss Re, Ettore Majorana Foundation and Centre For Scientific Culture, NVIDIA, INFORMS and the SoFiE Financial Econometrics Summer School for helpful comments. We thank the China Merchants Bank for generous research support. We thank Rose Wang for excellent research assistance.

[†]École Polytechnique Fédérale de Lausanne and Swiss Finance Institute, Email: damir.filipovic@epfl.ch
[‡]Stanford University, Department of Management Science & Engineering, Email: mpelger@stanford.edu.
[§]Stanford University, Department of Management Science & Engineering, Email: yeye@stanford.edu.

1 Introduction

The yield curve of U.S. Treasury securities is one of the most fundamental economic quantities and critical datasets for macroeconomic and financial research and applications. The yield curve or, equivalently, discount curve is a key factor for economists, traders, asset managers, central banks, and financial-markets regulators. Precise and robust yield estimates are needed for trading and making investment decisions, studying the term structure, predicting bond returns, analyzing monetary policy, and pricing assets, derivatives and liabilities. However, the yield curve is unobserved and needs to be inferred from a relatively sparse set of noisy prices of Treasury securities. The quality of the yield curve estimate directly impacts the quality of the output of applications that build up on it.

We introduce a robust, flexible, and easy-to-implement method, which sets the new standard for yield curve estimation. We show in an extensive empirical study that it uniformly dominates existing benchmarks in terms of out-of-sample yield and pricing errors. Our non-parametric estimator can explain complex yield curve shapes. It admits a closed-form solution as a simple kernel ridge regression, which is straightforward and fast to implement, and results in estimated discount bonds that are tradeable portfolios of Treasury securities. Our estimator is robust to outliers and data selection choices. We attribute the superior performance of our estimator to the optimal trade-off between flexibility and smoothness of the curve. We provide a publicly available and regularly updated dataset of daily zero-coupon Treasury yields based on our precise estimates.[1]

Our approach imposes minimal assumptions on the true underlying yield curve, using only the core elements that define the estimation problem. We first show that all arbitrage-free discount curves are twice differentiable. We then trade off pricing errors against an economically motivated smoothness reward of the discount curve. Setting up the objective function in terms of the aggregated pricing error and a smoothness measure uniquely determines the optimal basis functions that span the discount curve in a reproducing kernel Hilbert space. This type of function space is particularly important in machine learning. The smoothness parameter is chosen by cross-validation such that the estimated discount curve attains the lowest out-of-sample pricing error. The interpretability and generality of our estimator allows us to learn from the data the key features of the underlying yield curve. Our method also lends itself to a Bayesian interpretation. This perspective gives a distribution theory and implies confidence intervals for the estimated discount curve, yields and implied fixed income security prices.

The literature on yield curve estimation can be divided into two categories: parametric and non-parametric methods. The most important approaches in the first category are Nelson and Siegel (1987), Svensson (1994), and Gürkaynak, Sack, and Wright (2007), which is a specific implementation of the former two. These methods assume smooth parametric forms for the yield curve, and parameters are estimated by minimizing pricing errors. The most prominent benchmarks in the non-parametric category include Fama and Bliss (1987) and Liu and Wu (2021).

[1] Our estimated yields of zero-coupon Treasuries at daily and monthly frequency are available at `https://www.discount-bond-data.org`.

Non-parametric methods tend to be more flexible than parametric ones and have the potential to capture local as well as global variations in the yield curve. However, they come not entirely without assumptions, some of which could be rather unrealistic or restrictive. For example, Fama and Bliss (1987) assumes that the forward curve is piece-wise constant, while in Liu and Wu (2021), the price of a zero-coupon bond is given by a weighted average of first-order Taylor expansions of the yield curve at monthly spaced knots. Another strand of the literature uses spline-based methods, which approximate curves with local polynomial functions. Spline-based methods include regression spline techniques, e.g., McCulloch (1971) and Vasicek and Fong (1982), and smoothing splines, e.g., Fisher, Nychka, and Zervos (1995), Tanggaard (1997), Andersen (2007). The latter apply a roughness penalty to regularize the fitting cubic splines. Such ad-hoc assumptions can lead to overfitting and dynamic instability. Our approach falls into the non-parametric category, and includes smoothing splines as a special case. Our method is based on a novel perspective, that comes with minimal assumptions and an optimal trade-off between flexibility and smoothness, which prevents overfitting and ensures robustness.

We provide one of the most comprehensive empirical studies of U.S. Treasury securities in terms of data, benchmark methods and metrics. We find that our method strongly dominates all the leading parametric and non-parametric benchmarks, including Gürkaynak, Sack, and Wright (2007), Fama and Bliss (1987) and our own implementations of Svensson (1994) and Liu and Wu (2021). Our method achieves substantially smaller yield and pricing errors, both in- and out-of-sample, and for all maturity ranges. We confirm that the parametric benchmark models are misspecified. In contrast to the non-parametric benchmarks, our estimator is robust to outliers and stable over time. This also makes it the tool of choice to filter outliers from the data. In fact, our estimated discount curves are at the same time the most precise and smoothest curves. We attribute this property to the optimal trade-off between flexibility and smoothness.[2]

To illustrate the economic significance of our method, we examine implied forward rates and discount bond returns over time. We find systematic biases and instabilities in the time series that are generated by the leading benchmark methods. Our method produces the least biased and most stable time series, both unconditionally and conditional on economic states. This has important implications for downstream applications, such as regressions or GMM estimates, that use a panel of interest rates as input. It also directly impacts bond investment decisions in numerous ways,

[2]We thus extend the existing studies of Liu and Wu (2021), Jeffrey, Linton, and Nguyen (2006) and Bliss (1996), while also confirming their findings. Bliss (1996) finds that the estimator of Fama and Bliss (1987) performs best among the benchmarks, while the smoothing splines of Fisher, Nychka, and Zervos (1995) is a dubious choice for estimating the term structure. Both, in-sample and out-of-sample, smoothing splines performs poorly, particularly for short maturities. Bliss (1996) identifies as the problem that the smoothness award is applied uniformly across the term structure. Indeed, we show that a properly weighted smoothness award is crucial for reliable estimation. Jeffrey, Linton, and Nguyen (2006) compare a local kernel smoothing method, which is the predecessor of the Liu and Wu (2021) method, to Fama and Bliss (1987) and McCulloch (1971) regression splines. Overall, they find that the kernel smoothing is better for long maturities and Fama and Bliss (1987) perform the best for short maturities. We show that our estimator uniformly performs the best for all maturities. Liu and Wu (2021) show that their method is more precise than the parametric estimation of Gürkaynak, Sack, and Wright (2007), and emphasize the economic implications of using their data. We confirm the limitations of parametric estimators, and highlight the economic implications due to overfitting and dynamic instability of popular methods.

as well as our understanding of bond term premia, and fixed income asset pricing and derivatives. The broader economic implications of the choice of discount curve estimation method depend on the type of applications. We discuss the areas that are the most likely to be affected by a specific choice of estimator. We conclude that the methods currently used in the academic literature are suboptimal, imprecise and unstable for certain parts of the data, and can be problematic for some applications. Our method and data library solves all these issues, while having a formal theoretical foundation.

The basis functions of our estimator, which span the cross-section of discount bonds, have a clear economic interpretation. These basis functions are not selected ad-hoc, but are determined by the problem and optimally selected from the data. Their shapes reflect the patterns of level, slope, curvature and polynomials of increasing higher order. The smoothness reward optimally controls the degree of curvature by selecting the weights on higher order polynomials. The fact, that the leading principal components of a panel of discount bonds exhibit the same shape patterns, is simply the consequence that these are the optimal basis functions for spanning the discount curve.

We also discuss the limitations of non-parametric estimators for extrapolating yield curves beyond the observed maturities. Extrapolation requires additional assumptions on the extrapolation region, either in the form of imposed functional restrictions, or by an exogenous choice of some parameters for more flexible models. The advantage of our method is that the only exogenous choice parameter needed for the extrapolation has a clear economic interpretation as the infinite-maturity yield. Our confidence intervals confirm the large uncertainty that is associated with the extrapolation.

In summary, our paper makes several methodological and empirical contributions. First, we formally show that any arbitrage-free yield curve is twice differentiable. This motivates a smoothness trade-off for the yield curve estimation problem. By doing so we deduce the structure of the estimator from core principles, defined by the nature of the estimation problem. This is in contrast to the conventional other way around. Second, we provide the general theory behind our machine learning estimator, which results in a simple, transparent and closed-form solution and confidence intervals. Third, we perform one of the most extensive empirical comparison studies of yield curve estimation methods. As part of it, we develop a new protocol in terms of how to compare and evaluate different models. We show how the benchmark yield curve estimators differ from each other, and we find that our estimator outperforms them in all dimensions. Fourth, we provide a new zero-coupon Treasury yield curve dataset that overcomes the limitations of the popular datasets of Fama and Bliss (1987) and Gürkaynak, Sack, and Wright (2007) and improves upon the dataset of Liu and Wu (2021). Our dataset provides the most precise zero-coupon Treasury yield estimates for all maturity ranges while being robust to data selection choices.

The paper is organized as follows. In Section 2, we formulate the fundamental problem of estimating the discount curve based on the core principles of trading-off pricing errors and smoothness and develop the theory for our estimation approach. In Section 3, we perform an extensive empirical analysis and comparison study. In Section 4, we discuss the economic importance of our estimation method. In Section 5, we conclude. In the appendix, we provide the theoretical background and all proofs of our main results. We also collect additional empirical results, and provide a simulation study.

2 Estimating the Discount Curve

We first formulate the fundamental problem of estimating the discount curve. We then provide the general solution to this problem as a kernel ridge regression, and we show that most existing models are nested within our framework. We also give a Bayesian perspective along with a distribution theory.

2.1 Estimation Problem

The discount curve depicts the fundamental values $g(x)$ of zero-coupon Treasury bonds as a function of their time to maturity x. The discount curve is an important economic state variable, as it reflects term premia and expectations of market participants about future changes in interest rates. As such, it allows to price any fixed income security that generates a risk-less cash flow. As the collection of possible maturities includes essentially any future day, we represent the discount curve as a function $g : [0, \infty) \to \mathbb{R}$.[3]

The full discount curve is not observed and has to be estimated from the relatively sparse set of traded securities with noisy prices. Depending on the application, the researcher may further reduce this set by applying data filters to remove securities whose prices are likely to be contaminated by convenience yields or micro-structure effects.[4] Any estimation of the discount curve combines two elements: the law of one price and some form of regularization. Formally, we observe the prices of $P_1, \dots, P_M$ of M coupon bonds with cash flows summarized in the $M \times N$ matrix C. This means that C_{ij} denotes the cash flow of security i at date x_j, for a common set of cash flow dates $0 < x_1 < \cdots < x_N$. The law of one price implies that the fundamental value of security i is equal to

$$P_i^g = \sum_{j=1}^{N} C_{ij} g(x_j).$$

In other words, we can represent any coupon bond as a portfolio of zero-coupon bonds, where the cash flows denote the portfolio weights. Absent arbitrage this portfolio has the same price as the security. Due to market imperfections, such as the lack of a deep, liquid, and transparent market, as well as data errors, observed prices P_i deviate from the fundamental values P_i^g, resulting in pricing errors ϵ_i,

$$P_i = P_i^g + \epsilon_i. \tag{1}$$

[3]Equivalent expressions are the zero-coupon yield curve $y(x) = -\frac{1}{x} \log g(x)$ and forward curve $f(x) = -\frac{g'(x)}{g(x)}$. We write g' for the derivative of g. The corresponding expressions for the discount curve are $g(x) = \mathrm{e}^{-y(x)x}$ and $g(x) = \mathrm{e}^{-\int_0^x f(t)\,dt}$.

[4]We apply such standard filters in the empirical analysis below. Importantly, we show that in contrast to other methods our approach is robust to the choice of many filters.

The natural starting point to estimate the unknown discount curve g would be to minimize the weighted mean squared pricing errors of the observed securities,

$$\min_g \left\{ \sum_{i=1}^{M} \omega_i \left(P_i - P_i^g\right)^2 \right\}, \tag{2}$$

for some exogenous weights ω_i. However, solving this problem requires further structure due to two fundamental issues. First, problem (2) is under-determined. The number of observed prices M is substantially smaller than the number of cash flow dates N. For example, on a typical trading day we observe around $M \approx 300$ different Treasury bond prices, while a discount curve for 30 years requires estimates of around $N \approx 10{,}000$ daily zero-coupon bond prices. Second, the goal is to estimate the ground truth curve and not to overfit the noise. Flexible discount curves can match the observed prices in-sample exactly, but tend to perform poorly out-of-sample and produce inconsistent curve shapes and dynamics. Therefore, any estimation approach for the discount curve imposes some form of regularization to limit the degrees of freedom in (2).

This leads us to two key questions. First, what is the class of plausible functions for the discount curve? Second, what is the appropriate regularization to deal with the only moderately large and noisy data? We provide an answer to both questions based on the principle of smoothness. Most of the literature agrees on smoothness of the discount curve as an economically plausible assumption: bonds with similar maturities and cash flows should have similar prices. We qualify this assumption in the following theorem, which seems to be new to the literature and answers the first question. It shows that any arbitrage-free discount curve is twice weakly differentiable, and therefore a smooth curve. Hence, it is essentially without loss of generality to study the extremely large space of discount curves that are twice weakly differentiable.

Theorem 1
Assume that the discount curve g is arbitrage-free in the sense that it can be written as

$$g(x) = \mathbb{E}_{\mathbb{Q}}\left[\mathrm{e}^{-\int_0^x r_u du}\right],$$

for a short rate process of the form $dr_t = \mu_t\, dt + dM_t$ for some drift process μ_t and some locally square-integrable martingale M_t with respect to the risk-neutral pricing measure $\mathbb{Q}$.[5] *We further assume that $g(x)$ and the following moments are finite*

$$\mathbb{E}_{\mathbb{Q}}\left[\int_0^x \mathrm{e}^{-\int_0^t r_u\, du}\left(r_t^2 + |\mu_t|\right) dt\right] < \infty, \tag{3}$$

$$\mathbb{E}_{\mathbb{Q}}\left[\int_0^x \mathrm{e}^{-2\int_0^t r_u\, du}\, d\langle M, M\rangle_t\right] < \infty, \tag{4}$$

for any finite x, where $\langle M, M\rangle_t$ denotes the predictable quadratic variation of M_t. Then $g(x)$ is

[5] See (Jacod and Shiryaev, 2003, I.§4a) for the definition of a locally square-integrable martingale and its predictable quadratic variation.

twice weakly differentiable in x in the sense that it can be written as

$$g(x) = 1 - r_0 x + \int_0^x \int_0^t g''(s) ds\, dt, \tag{5}$$

for a locally integrable function $g''(s)$. Its first and second derivatives are given by the present values

$$g'(x) = \mathbb{E}_{\mathbb{Q}}\left[\mathrm{e}^{-\int_0^x r_u du} (-r_x)\right], \tag{6}$$

$$g''(x) = \mathbb{E}_{\mathbb{Q}}\left[\mathrm{e}^{-\int_0^x r_u du} (r_x^2 - \mu_x)\right], \tag{7}$$

of the contingent claims with payoffs $-r_x$ and $r_x^2 - \mu_x$ at x, respectively.

Theorem 1 makes two important statements. First, the economically meaningful arbitrage-free discount curves are included in the space of twice weakly differentiable functions. Second, it connects first and second derivatives of the discount curve to present values of contingent claims whose payoffs are linear and quadratic in future short rate levels and drift. Note, that the setup in Theorem 1 is very general. The short rate process r_t can exhibit jumps through its local martingale part M_t. Essentially all continuous-time arbitrage-free models are of this form, subject to the technical moment conditions (3) and (4) on r_t and its characteristics.[6] We henceforth assume that the discount curve g is arbitrage-free, and by Theorem 1 therefore twice weakly differentiable of the form (5).

To answer the second question, we introduce a general measure of the smoothness of a twice weakly differentiable curve g, which is given by the weighted average of the first and second derivatives,

$$\|g\|_{\alpha,\delta} = \left(\int_0^\infty \left(\delta g'(x)^2 + (1-\delta) g''(x)^2\right) \mathrm{e}^{\alpha x}\, dx\right)^{\frac{1}{2}}, \tag{8}$$

for some maturity weight $\alpha \geq 0$ and shape parameter $\delta \in [0,1]$. This measure encompasses the conventional tension and curvature measures for curves. Penalizing $g'(x)^2$ avoids oscillations, hence forcing the curve g to be tense, while penalizing $g''(x)^2$ avoids kinks, enforcing the curve g to be straight. The tension parameter δ balances between these two forces, while the weight function $\mathrm{e}^{\alpha x}$ allows the smoothness measure to be maturity-dependent. This allows for greater pricing flexibility at shorter maturities while enforcing a smooth long end, which is an important feature of our approach, as suggested by Bliss (1996).[7]

Accordingly, we define the extremely large space of discount curves given by the set $\mathcal{G}_{\alpha,\delta}$ of twice weakly differentiable discount curves g of the form (5) and finite smoothness measure (8).[8]

[6] Remarkably, (5) rules out the Fama–Bliss curves. Indeed, Fama and Bliss (1987) assume a piece-wise constant forward curve. This corresponds to short rate dynamics with deterministic fixed jump times predetermined by the discontinuities of the forward curve, which is not of the form as assumed in Theorem 1.

[7] Bliss (1996) finds that the smoothing spline of Fisher, Nychka, and Zervos (1995) does poorly fit at shorter maturities and suggests that this is due to their smoothness award being applied uniformly across the term structure. While much of the theory that we develop in Appendix A holds for general weight functions, the exponential function $\mathrm{e}^{\alpha x}$ is convenient as it leads to closed-form expressions.

[8] The finiteness of (8) corresponds to a transversality condition on the present values (6) and (7), which should

This framework turns out to be very flexible and nesting many existing methods as special cases, as we show in Theorem 3 below.

We conclude that the estimator of the discount curve should be based on a trade-off between minimizing pricing errors of the observed securities and rewarding its smoothness. This core principle alone leads us to the formulation of the estimation problem. Rewarding smoothness is motivated by economic principles. Smooth discount curves imply that bonds with similar cash flows and maturities have similar prices. Large sudden changes along the discount curve, on the other hand, can lead to trading strategies with extreme payoffs. Indeed, Theorem 1 formally shows that large values of the first or second derivative imply large prices for specific contingent claims based on the short-rate process. Hence, rewarding smoothness sets limits on payoffs for trading strategies of Treasury securities with nearby maturities, which are likely to be infeasible out-of-sample. Most parametric or non-parametric yield curve estimation methods implicitly or explicitly impose smoothness. However, in contrast to our approach, these smoothness requirements are typically ad-hoc and not transparent.

In summary, the formal estimation problem trades off the weighted mean squared pricing error against the smoothness of g, as expressed by the convex optimization problem

$$\min_{g\in\mathcal{G}_{\alpha,\delta}} \Bigg\{ \underbrace{\sum_{i=1}^{M} \omega_i (P_i - P_i^g)^2}_{\text{pricing error}} + \lambda \underbrace{\|g\|_{\alpha,\delta}^2}_{\text{smoothness}} \Bigg\}, \tag{9}$$

for some smoothness parameter $\lambda > 0$. Increasing the smoothness parameter has has three effects. First, it enforces smoothness by reducing excessive oscillations and curvature of the estimated curve. Second, it regularizes the problem as intuitively a smooth curve can be described by fewer parameters (we formalize this aspect later). Third, it leads to more robust estimates by penalizing outliers that lead to deviations. Given exogenous weights ω_i, the estimation problem is completely determined up to the three parameters α, δ, λ. While keeping the framework general, we will select these parameter values empirically from the data via cross-validation to minimize out-of-sample pricing errors. In particular, we test whether data awards smoothness ($\lambda > 0$) against a larger pricing error. In the limit $\lambda \to 0$, we obtain an exact fit.

2.2 General Solution

We provide a remarkably simple closed-form solution to the infinite-dimensional optimization problem (9). The solution builds on insights in functional analysis and machine learning by leveraging the structure of reproducing kernel Hilbert spaces. This type of function space generalizes splines and is particularly important in machine learning because of the celebrated representer theorem. It states that every function in a reproducing kernel Hilbert space (RKHS) that minimizes an empiri-

converge exponentially fast to zero as $x \to \infty$. In Lemma 8(i) we give sufficient conditions in terms of the forward curve. If $\delta = 1$, we only assume that $g(x) = 1 + \int_0^x g'(s)\,ds$ is once weakly differentiable with locally integrable derivative g'. This allows to nest the Fama–Bliss curves as a boundary case, see Theorem 3.

cal objective function can be written as a linear combination of the reproducing kernel evaluated at the training points. This is crucial as it effectively simplifies an infinite dimensional optimization problem to a finite dimensional one. This means by setting up the objective function (pricing error and smoothness measure) and the feasible set of curves (twice weakly differentiable), the representer theorem uniquely pins down the optimal basis functions that solve the non-parametric problem. The solution is linear in these basis functions and boils down to a kernel ridge regression. We recap the definition of a RKHS and provide references for more background and applications in Appendix A.2.

Our conceptual framework is transparent in the choices we make. Our approach is fundamentally different from most yield curve estimation methods, which would first specify an ad-hoc set of basis functions and then study the properties of the fitted curve. We reverse the order, by deriving the estimation problem in (9) from first principles and properties of the discount curve, which then implies the optimal basis functions that are selected by the data.

Theorem 2 gives the highly tractable solution to the estimation problem (9). It boils down to a kernel ridge regression (KR) that admits a closed-form solution, which is straightforward and fast to implement. Our smoothness measure is the norm of an RKHS whose reproducing kernel we find in closed form. The representer theorem then applies accordingly. Henceforth, we stack the prices into the column vector $P = (P_1, \ldots, P_M)^\top$, and we write $\mathbf{1}$ for the column vector consisting of 1s. For any two numbers a, b, we write $a \wedge b = \min\{a, b\}$ and $a \vee b = \max\{a, b\}$.

Theorem 2 (**Kernel Ridge (KR) Solution**)
The estimation problem (9) *has a unique solution* $\hat{g}$, *which is given in closed form by*

$$\hat{g}(x) = 1 + \sum_{j=1}^{N} k(x, x_j)\beta_j, \tag{10}$$

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

$$\beta = C^\top (C\boldsymbol{K}C^\top + \Lambda)^{-1}(P - C\mathbf{1}), \tag{11}$$

for the $N \times N$*-kernel matrix* $\boldsymbol{K}_{ij} = k(x_i, x_j)$, *and* $\Lambda = \operatorname{diag}(\lambda/\omega_1, \ldots, \lambda/\omega_M)$. *The kernel function* $k : [0, \infty) \times [0, \infty) \to \mathbb{R}$ *is given in closed form according to the five cases:*

(i) $\alpha = 0$, $\delta \in (0, 1)$*:*

$$k(x, y) = \frac{1}{\delta}(x \wedge y) + \frac{1}{2\delta\rho}\left(\mathrm{e}^{-\rho(x+y)} - \mathrm{e}^{\rho(x \wedge y) - \rho(x \vee y)}\right) \tag{12}$$

where we define $\rho = \sqrt{\delta/(1-\delta)}$*;*

(ii) $\alpha = 0$, $\delta = 1$*:*

$$k(x, y) = x \wedge y; \tag{13}$$

(iii) $\alpha > 0,\ \delta = 0$:

$$k(x, y) = -\frac{x \wedge y}{\alpha^2} \mathrm{e}^{-\alpha(x \wedge y)} + \frac{2}{\alpha^3} \left(1 - \mathrm{e}^{-\alpha(x \wedge y)}\right) - \frac{x \wedge y}{\alpha^2} \mathrm{e}^{-\alpha(x \vee y)}; \tag{14}$$

(iv) $\alpha > 0,\ \delta \in (0, 1)$:

$$\begin{aligned} k(x, y) = &-\frac{\alpha}{\delta \ell_2^2} \left(1 - \mathrm{e}^{-\ell_2 x} - \mathrm{e}^{-\ell_2 y}\right) + \frac{1}{\alpha \delta} \left(1 - \mathrm{e}^{-\alpha(x \wedge y)}\right) \\ &+ \frac{1}{\delta \sqrt{D}} \left(\frac{\ell_1^2}{\ell_2^2} \mathrm{e}^{-\ell_2(x+y)} - \mathrm{e}^{-\ell_1(x \wedge y) - \ell_2(x \vee y)}\right) \end{aligned} \tag{15}$$

where we define $D = \alpha^2 + 4\delta/(1 - \delta)$, $\ell_1 = \frac{\alpha - \sqrt{D}}{2}$, *and* $\ell_2 = \frac{\alpha + \sqrt{D}}{2}$;

(v) $\alpha > 0,\ \delta = 1$:

$$k(x, y) = \frac{1}{\alpha} \left(1 - \mathrm{e}^{-\alpha(x \wedge y)}\right). \tag{16}$$

The case $(\alpha, \delta) = (0, 0)$ *is not specified.*

The basis functions $k(\cdot, x_j)$, which span the discount curve (10) in our non-parametric approach, are fully determined by the smoothness measure (8). That is, the kernel selection is guided by the principle of a smooth discount curve as measured by (8), and the parameters of the optimal basis functions are learned from the market data. This is distinctively different from conventional kernel-smoothing methods in the literature, which select the kernel exogenously.

As the solution of a kernel ridge regression, the estimated discount curve (10) becomes linear in the observed prices. Hence, while the underlying estimation problem itself is non-linear, it is translated into a linear problem in the kernel space. The solution is fast to implement and reproducible by any researcher. The main computational task in (11) boils down to a left matrix division in dimension M, and a few matrix multiplications, which results in an overall computational complexity of order $O(M^3 + MN^2)$. We emphasize that, while our solution itself is simple, the underlying concepts and derivations are non-trivial. Appendix A provides the general functional analytic theory and proofs.

The discount bonds in (10) are replicated by a portfolio of observed coupon bonds and a cash investment. This aspect is of independent interest as it allows to create synthetic discount bonds of any desired maturity given the set of traded bonds. The portfolio weights for these synthetic discount bonds depend on the smoothness parameter λ. These synthetic discount bonds can be used among others for term structure asset pricing or to immunize an obligation against interest rate changes by matching its cash flows.

The kernel ridge regression imposes sparsity of β, in the sense that only those components β_j are non-zero that correspond to actual cash flow dates x_j. For a horizon of 30 years to maturity with daily cash flows, N is of the order 10,000. However, the matrix C for most coupon bonds is usually sparse. Indeed, with semi-annual coupon payments most columns of the cash flow matrix C are zero. Consequently, the corresponding components of β in Equation (11) are zero, too.

The smoothness parameter λ controls the ridge shrinkage. Larger shrinkage increases the relative importance of the dominant eigenvectors of the kernel space projected on the cash flows. As we will show in our empirical analysis, the dominant principal components of the kernel space are associated with lower order polynomial-type functions that capture level, slope and curvature shapes. Hence, shrinkage reduces the dimensionality of the parameter space by putting most weight on these basis patterns.

Our framework is fully flexible in terms of the maturity domain. While in the main text we assume a right-open unbounded maturity domain, $[0, \infty)$, we show in Appendix A.7 that for $\alpha > 0$ our results easily extend to infinite maturity, $x_N = \infty$. This is relevant, as we find that the parameter α corresponds to the yield of an infinite-maturity zero-coupon bond and hence has a clear economic interpretation, under some mild technical assumptions. On the other hand, we show in Appendices A.4.5 and A.5 that Theorem 2 carries over to bounded maturity domains $[0, \tau)$, for pre-specified finite time horizons $\tau < \infty$. This case corresponds exactly to the smoothing splines of Fisher, Nychka, and Zervos (1995), Tanggaard (1997), and Andersen (2007). However, the case of a pre-specified finite time horizon is not desirable. First, τ represents an additional choice parameter, which affects the results. Second, we find that the corresponding basis functions $k(\cdot, x_j)$ exhibit linear and quadratic polynomial components. This does not seem to be a suitable functional shape for spanning the discount curve, which could partly explain the bad performance of these smoothing splines reported in the literature.

The choice of exogenous weights ω_i allows researchers to include economic priors. Theorem 2 allows for individual weights be set to $\omega_i = \infty$, by setting the corresponding $\Lambda_i = \lambda/\infty = 0$. This corresponds to the exact pricing of security i, and gives the researcher great flexibility in formulating the estimation problem. As a special case, we obtain the smoothest twice weakly differentiable curve that fits all observed prices exactly by setting the entire matrix $\Lambda = 0$, which corresponds to the limit $\lambda \to 0$. In our empirical analysis, we follow the standard approach in the literature for the pricing error weights ω_i by setting them inversely proportional to the squared duration D_i of security i, that is

$$\omega_i = \frac{1}{M} \frac{1}{(D_i P_i)^2}. \tag{17}$$

Hence, the weighted mean squared pricing error in (9) equals approximately the mean squared yield fitting error,

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

where Y_i and Y_i^g denote the yields to maturity of security i, corresponding to the observed price P_i and fundamental value P_i^g, respectively. This is the same choice as for example in Gürkaynak, Sack, and Wright (2007) and discussed in more detail in Appendix A.8.1.

2.3 Special Cases

Most existing models for estimating the discount curve are nested within our framework, in the sense that their feasible discount curves are restricted to subsets of our general function space $\mathcal{G}_{\alpha,\delta}$. These subsets are specified by imposing additional ad-hoc assumptions on the curves, such as their shape (Fama–Bliss) or parametric forms (Nelson–Siegel–Svensson). Hence, if those assumptions and models were correct, we would recover them in our estimation. However, including the methods discussed below as benchmarks in our empirical analysis, we show that their restrictive ad-hoc assumptions are rejected by the data.[9]

Fama–Bliss:
The discount curve and implied discount portfolios of Fama and Bliss (1987) are widely used in academic research. Fama and Bliss (1987) propose a non-parametric estimator, which assumes a piece-wise constant forward curve with finitely many steps, that is $f_{FB}(t) = f_i$ for $t \in [t_i, t_{i+1})$ for $i = 0, \ldots, m$ with $0 = t_0 < \cdots < t_m < t_{m+1} = \infty$. This results in the discount curve $g_{FB}(x) = \mathrm{e}^{-\int_0^x f_{FB}(t)\,dt}$.

Nelson–Siegel–Svensson:
A widely used parametric method to estimate the discount curve is due to Nelson and Siegel (1987) and Svensson (1994). The Nelson–Siegel–Svensson (NSS) forward curve is parametrized as

$$f_{NSS}(x) = \gamma_0 + \gamma_1 \mathrm{e}^{-\frac{x}{\tau_1}} + \gamma_2 \frac{x}{\tau_1} \mathrm{e}^{-\frac{x}{\tau_1}} + \gamma_3 \frac{x}{\tau_2} \mathrm{e}^{-\frac{x}{\tau_2}}$$

for the parameters $\gamma_0, \gamma_1, \gamma_2, \gamma_3$ and $\tau_1, \tau_2 > 0$. This results in the discount curve $g_{NSS}(x) = \mathrm{e}^{-\int_0^x f_{NSS}(t)\,dt}$. Gürkaynak, Sack, and Wright (2007) is a specific implementation of the Nelson-Siegel-Svensson model, which we also include in our empirical study.

Smoothing splines:
Smoothing splines of Fisher, Nychka, and Zervos (1995) and Tanggaard (1997) are a special case of our framework for $\alpha = 0$ and $\delta = 0$ when restricted to a bounded maturity domain $[0, \tau)$. In that case, our kernel function $x \mapsto k(x, y)$ becomes a cubic spline in x. The extension of Andersen (2007) to tension splines corresponds to $\delta \in (0, 1)$ in our framework. Details are given in Appendix A.4.5.[10]

[9] We want to clarify that the benchmark methods are special cases in terms of the functions space, in which their solutions lie, but in general their implementation is different, that is, the estimators themselves are not special cases of our kernel ridge regression and have different statistical properties. Our method leads to a convex optimization problem with a unique global solution that is given in closed form as a simple ridge regression. Other methods can lead to non-convex optimization problems that can be solved for local minima at best, using numerical approximations of first order conditions or other gradient-search methods. This can also lead to non-robust solutions.

[10] In more detail, Fisher, Nychka, and Zervos (1995) fit a cubic spline that includes an equally weighted roughness penalty for the second derivative of the estimated discount curve. The choice of knot points and spline function is ad-hoc. Tanggaard (1997) state that smoothing splines do not require the specification of knot points, and should be set to all cash flow dates. They argue that cubic splines arise as basis functions for a curvature penalty, but do not provide a rigorous functional analytical derivation of the statement. We conjecture that the inferior performance of smoothing splines is due to two main reasons: First, as stated already by Bliss (1996) applying the smoothness

Theorem 3 states that all of the above model curves are contained in our function space. Part (i) reveals that the Fama–Bliss curves are a boundary case, namely for $\delta = 1$. Part (iii) illustrates the large size of our feasible set $\mathcal{G}_{\alpha,\delta}$, which includes all spline-based estimators.

Theorem 3 **(Special cases)**
The class of Fama–Bliss, NSS and spline-based discount curves are special cases of our function space for specific parameter choices:

(i) *The Fama–Bliss curve g_{FB} lies in $\mathcal{G}_{\alpha,\delta}$ for any $\alpha \in [0, 2f_m)$ and $\delta = 1$;*

(ii) *The NSS curve g_{NSS} lies in $\mathcal{G}_{\alpha,\delta}$ for any $\alpha \in [0, 2\gamma_0)$ and $\delta \in [0, 1]$;*

(iii) *Any twice weakly differentiable function $g : [0, x_N] \to \mathbb{R}$ on a finite interval, $x_N < \infty$, with $g(0) = 1$ can be extended to $[0, \infty)$ such that g lies in $\mathcal{G}_{\alpha,\delta}$ for any $\alpha \geq 0$ and $\delta \in [0, 1]$. This includes all splines of degree two or more.*

In Appendix A.4 we show that our framework is also consistent with arbitrage-free dynamic term structure models, in the sense that it contains all discount curves that are generated by stochastic models of the Heath, Jarrow, and Morton (1992) type. Other nested approaches include Smith and Wilson (2001), which is the insurance industry standard in Europe for constructing the discount curve used in the regulatory Solvency II framework, and Filipović and Willems (2018). In contrast, the following model is not nested.

Liu–Wu:
Liu and Wu (2021) propose a non-parametric estimator that constructs discount curves by combining a normal kernel-smoothing method with a special bandwidth selection.[11] This results in discount curves of the form

$$\hat{g}_{LW}(x) = \frac{\sum_{n=1}^{360} K_{h(x)}(n-x) \exp\left(-(y_n + (x-n)y'_n)x\right)}{\sum_{n=1}^{360} K_{h(x)}(n-x)}$$

for the normal kernel-weighting function

$$K_{h(x)}(n-x) = \frac{1}{\sqrt{2\pi h(x)^2}} \exp\left(-\frac{(n-x)^2}{2h(x)^2}\right)$$

with bandwidth $h(x)$.[12] The Liu–Wu discount curve $g_{LW}(x)$ can be interpreted as a local kernel-smoothing mixture of auxiliary discount curves, $\exp\left(-(y_n + (x-n)y'_n)x\right)$. The parameters (y_n, y'_n)

award uniformly across the term structure is too restrictive for the short end. Second, the boundary conditions at the finite time horizon τ result in unstable behavior. Both problems do not affect the KR estimator, that uses a maturity dependent smoothness measure and unbounded maturities.

[11] Their framework builds on the work of Linton, Mammen, Nielsen, and Tanggaard (2001), who introduced a kernel-smoothing approach in estimating the yield curve.

[12] Liu and Wu (2021) measure time in months. The sum over n up to 360 thus corresponds to a time horizon of 30 years to maturity.

are found by minimizing a kernel-weighted mean squared pricing error, which is a non-convex optimization problem. The bandwidth $h(x)$ depends on the cross-section of bonds, and as such is piece-wise continuous but not differentiable in x, as shown empirically in Section 3.3. Hence the Liu–Wu discount curves are not differentiable and thus not contained in $\mathcal{G}_{\alpha,\delta}$ in general. If the bandwidth $h(x)$ was chosen as a differentiable function, the resulting kernel smoothing estimator would be special case of our framework. Note that method of Liu and Wu (2021), and more generally kernel estimators, are "local" approaches, which determine the fit and smoothness locally using only bonds with nearby maturities. In contrast, our smoothness measure is a "global" approach, which trades off overall smoothness against the aggregate pricing error. This will be beneficial in avoiding overfitting and dealing with outliers as shown in our empirical study. We include the Liu–Wu method as a the most flexible non-parametric benchmark in our empirical analysis. We show that an optimal method needs to optimally trade off flexibility against smoothness, which we achieve with our method.

2.4 Bayesian Perspective and Distribution Theory

Our KR approach lends itself to a Bayesian interpretation. This perspective allows us to obtain a distribution theory and confidence intervals for our estimated discount curve, yields and implied security prices. Importantly, our estimator does not require any Bayesian assumptions, but under this additional perspective we can make stronger statements.

Thereto we view the discount curve g as a Gaussian process with prior mean function $m : [0,\infty) \to \mathbb{R}$ and covariance defined by the kernel k. That is, for every finite selection of maturities $z_1, \dots, z_n$ the distribution of the vector $(g(z_1), \dots, g(z_n))$ is Gaussian with mean $(m(z_1), \dots, m(z_n))$ and covariance matrix $k(z_i, z_j)$. This is not an ad-hoc assumption, but makes the standard connection between Gaussian processes and RKHS estimators with a firm foundation in statistics and machine learning. For simplicity of notation, we will stack all the cash flow dates into the column vector $\boldsymbol{x} = (x_1, \dots, x_N)^\top$ and evaluate functions elementwise. Hence, the Gaussian process view assumes the Gaussian prior distribution

$$g(\boldsymbol{x}) \sim \mathcal{N}\left(m(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}^\top)\right).$$

Next, we impose assumptions on the errors in the pricing equation (1). The pricing errors ϵ_i are viewed as independent centered Gaussian random variables with variance parameters σ_i^2, that is, the vector $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \Sigma^\epsilon)$ with $\Sigma^\epsilon = \operatorname{diag}(\sigma_1^2, \dots, \sigma_M^2)$.

Bayesian updating implies that the conditional distribution of g, given the observed prices P, is still Gaussian with posterior mean function

$$m^{\text{post}}(z) = m(z) + k(z, \boldsymbol{x})^\top C^\top (C\boldsymbol{K}C^\top + \Sigma^\epsilon)^{-1}(P - Cm(\boldsymbol{x})) \tag{18}$$

and posterior variance given by the posterior kernel

$$k^{\text{post}}(y,z) = k(y,z) - k(y,\boldsymbol{x})^\top C^\top (C\boldsymbol{K}C^\top + \Sigma^\epsilon)^{-1} Ck(\boldsymbol{x},z) \tag{19}$$

for $\boldsymbol{K}_{ij} = k(x_i, x_j)$. We arrive at the following theorem.

Theorem 4 (Bayesian perspective)
Assume the prior mean function of g is constant, $m(x) = 1$, the covariance kernel k is as in Theorem 2, and the pricing error variance equals $\sigma_i^2 = \lambda/\omega_i$. Then, the posterior mean function in Equation (18) *coincides with the KR discount curve estimator in Equation* (10).

Moreover, we obtain a confidence range for the discount curve: for every finite selection of maturities $z_1, \dots, z_n$, the conditional distribution of the vector $(g(z_1), \dots, g(z_n))$, given the observed prices P, is Gaussian with mean $(m^{post}(z_1), \dots, m^{post}(z_n))$ and covariance matrix $k^{post}(z_i, z_j)$.

Thus, we obtain the posterior distribution, given observed prices P, of any fixed income instruments with given cash flows. In particular, the model implied coupon bond prices have the Gaussian posterior distribution $Cg(\boldsymbol{x}) \sim \mathcal{N}\left(Cm^{\text{post}}(\boldsymbol{x}), Ck^{\text{post}}(\boldsymbol{x}, \boldsymbol{x}^\top)C^\top\right)$.

The Bayesian perspective offers an alternative interpretation of our estimator. We start with an uninformative prior of the mean, while the kernel represents our prior of the deviations of this mean.[13] Observed prices with larger noise are associated with a larger roughness penalty. When updating the prior mean, we put more weight on the observed prices with smaller noise or higher correlation with the dominant principal components of the kernel projected on the cash flows. The posterior variance is larger for noisier observations, which corresponds to a larger penalty, or when the prices are not spanned by the dominant principal components of the kernel projected on the cash flows.

Appendix A.6 provides a formal discussion and the proofs for the Bayesian approach. It also discusses the extension to more general priors. In particular, it shows that our method can actually accommodate any prior, other than $m(x) = 1$. However, in unreported experiments we found that different priors lead to essentially the same posterior mean function. This is in line with findings in the machine learning literature on Gaussian processes, which often sets priors equal to the zero function. Furthermore, we show that the posterior mean function is invariant with respect to scaling the kernel and the error variance, and hence our KR discount curve estimator does not change. However, such a scaling impacts the posterior variance of the estimator. We propose to select the scaling to maximize the empirical log-likelihood, which results in the prior variance that best describes the observed data.

2.5 Improving Given Discount Curves

Our method also provides a conceptual framework to improve a given discount curve estimate. In this sense, it can be understood as a tool to enhance any given method or model.

[13] A constant prior mean function, $m(x) = 1$, corresponds to a zero prior yield curve.

As discussed in the previous section, our KR estimator corresponds to a Bayesian estimator with an uninformative prior of $m(x) = 1$ for the discount curve. In other words, the KR estimator in Equation (10) estimates deviations from 1 that optimally trade off pricing errors and smoothness of the resulting discount curve. With minimal modification, our approach can also determine the optimal deviations from any given candidate model. The uninformative prior of 1 in Equation (10) must simply be replaced by the estimate of a candidate model. Examples are parametric estimates such as the Nelson-Siegel Svensson model or curve estimates with KR from the previous day. In the latter case, KR would estimate changes that are sufficiently smooth and result in small pricing errors. As the roughness penalty is applied to the changes relative to the prior, a meaningful prior curve should be sufficiently smooth. We can interpret this extension as reducing the bias of a candidate estimator while controlling the smoothness of the change. Empirically, we find that our uninformative prior of 1 yields essentially the same results as another prior such as a GSW curve.[14] However, an informative prior could be beneficial for estimating the discount curve of bond markets that have sparser and noisier data than the U.S. Treasury market.

3 Empirical Results

We perform an extensive empirical analysis. We first describe the data and introduce the evaluation metrics based on which we select the parameters. We then conduct a comprehensive comparison study to the most important benchmarks. We discuss robustness to outliers and over time, curve extrapolation, principal component analysis, and the economic implications of our method.

3.1 Data and Evaluation

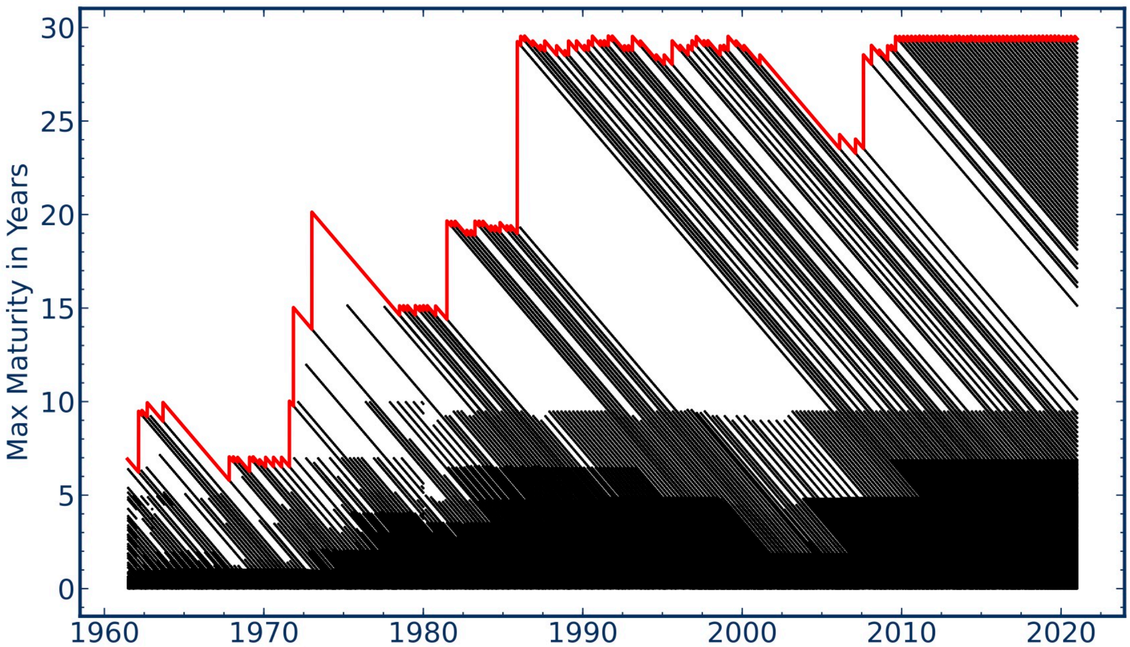
We use the standard set of CUSIP-level coupon-bearing Treasury bond data from the CRSP Treasuries Time Series. Our main sample are daily observations from June 1961 to December 2020. For each bond, we observe the end-of-day bid and ask prices and its features including the maturity and coupon payments. We use ex-dividend bid-ask averaged mid-prices for the bonds and our main analysis focuses on the end of month prices. Throughout our analysis we measure time in years.

We apply standard data filters to remove issues that trade at a premium due to their specialness or liquidity. First, our sample only includes fully taxable, non-callable, and non-flower bond issues.[15] This step ensures that our sample does not include bonds with tax benefits and option-like features. This is the same standard filter as applied in Fama and Bliss (1987), Gürkaynak, Sack, and Wright (2007) and Liu and Wu (2021). Second, we exclude on-the-run issues due to their liquidity and specialness. In more detail, we follow Gürkaynak, Sack, and Wright (2007) and Liu and Wu (2021) and exclude the two most recently issued securities with maturities of 2, 3, 4, 5, 7, 10, 20, and 30 years for securities issued in 1980 or later. We confirm that our dataset is very close to Liu and

[14] The results are available upon request.

[15] This means that CRSP ITYPE equals 1, 2, or 4. We also remove the 13 issues of securities whose time series of prices terminate because these bonds are "all exchanged".

Figure 1: Maximal time to maturity in years



This figure shows the maximal time to maturity in years for all bonds in our data. The black lines correspond to different bonds. The red line indicates the maximal maturity available on a specific day.

Wu (2021). After applying the filters, this gives us a total of 5,353 issues of Treasury securities and 121,088 end of month price quotes for 715 months.

While we show that our method is robust to outliers, we also consider the impact of various outlier filters. The first outlier filter removes all bonds with maturities under 90 days as suggested in Gürkaynak, Sack, and Wright (2007), as those prices seem to have the largest number of outliers relative to estimated curves.[16] The next two filters remove outliers in a data-driven way. We use either the estimated yield curve with the Nelson-Siegel-Svensson method or our KR method to remove those securities whose yield errors exceed three standard deviations based on the cross-section of yield errors for that day. Our baseline setup estimates models on the data after applying the 90-day maturity filter. However, our baseline evaluation uses no filters, and also includes the prices of securities with maturities under 90-days. We also show the evaluation of the models after applying the 90-day maturity, NSS- or KR-outlier detection filters. Our comprehensive analysis shows that our results are not driven by outliers and are robust to all of these choices.

Figure 1 shows the maximum maturities of all Treasury securities over time. First, we note that the bonds have a very unequal maturity distribution. The maximum available maturity in the first 20 years is below 20 years. While the later time periods include bonds with 30 years of maturity, the middle spectrum of maturities between 10 to 20 years is only sparsely represented. Both empirical facts have an impact on the estimation of the yield curve. As we will show, the estimation of the

[16] Note that the Treasury notes and bonds with longer maturity still have coupon payments within the first 90 days, which allows us to estimate the short end of the discount curve.

yield curve beyond the maximum available maturity becomes an extrapolation problem, which is different from the interpolation between observed values. Furthermore, the estimation of the yield curve in the sparsely populated maturity range is a challenging problem.

We compare our method (KR) with the estimates from Fama and Bliss (1987) (FB), Gürkaynak, Sack, and Wright (2007) (GSW), and our own implementations of the Nelson-Siegel-Svensson model (NSS) and the Liu and Wu (2021) (LW) approach. These are the leading benchmark methods and discussed in detail in Section 2.3. The FB curve is constructed from piece-wise constant forward rate estimates. Robert Bliss has shared a granular dataset of estimated forward rates from June 1961 to December 2013, which we use to construct the yield curves and to price Treasury bonds.[17] We implement the NSS model on the same underlying dataset as our KR method following the procedure in Svensson (1994). We circumvent the known issue that the estimation of a non-linear model can be numerically unstable, by using multiple numerical solvers to ensure convergence. GSW is a specific implementation of NSS, but estimated on a more restricted dataset.[18] The GSW parameter estimates are available on the authors' website at daily frequency. We implement the LW approach on our data following exactly the same approach as in the original paper.[19]

We do not include spline estimators of the yield curve for the following reasons: First, Bliss (1996) and Jeffrey, Linton, and Nguyen (2006) have already shown that the FB estimator and simpler variants of the LW estimator dominate spline estimators. Second, there is no well-defined benchmark implementation or reference dataset. Regression spline estimators come with a large number of tuning parameters, in particular for selecting the knot points. The yield estimates of the Federal Reserve Bank based on splines are only publicly available for a few selected maturities, and as the implementation choices are not reported, it is unclear how to replicate them. Third, the most widely used methods in the empirical finance and macroeconomic literature are by far the four benchmark methods that we consider.

We evaluate various pricing metrics for the end of month prices for the time period from June 1961 to December 2020. However, as we have FB data only available from June 1961 to December 2013 we also show a comprehensive comparison analysis on this shorter sample, which allows us to include FB as a benchmark method. We show all results in- and out-of-sample. The in-sample results estimate the discount curve on the same day that we use for the evaluation. Flexible non-parametric methods are expected to perform better, but might overfit in-sample. The out-of-sample analysis evaluates the discount curve on the business day after the estimation day. The

[17]We thank Robert Bliss for sharing the data with us. This data is more detailed and includes more maturities than the version of the FB dataset available on CRSP. In this granular data the knots of the yield curve are 1-month apart up to 3 years, 6-month apart from 3 to 10 years and 1-year apart for maturities longer than 10 years. We interpolate FB knot points using the underlying assumption of the FB model that the daily forward rate curve should be piece-wise constant.

[18]Prior to 1980, GSW uses the Nielson-Siegel functional form, which sets γ_3 to zero as this more restricted model explains the sparse and shorter maturity data better. After 1980, GSW uses the general NSS form without setting γ_3 to zero. For comparability, we follow GSW's convention and use the restricted functional form prior to 1980 for our NSS estimates as well.

[19]We thank Cynthia Wu for sharing their implementation code with us. We use their optimal tuning parameter for the adaptive bandwidth construction.

underlying assumption is that the discount curve does not change much over consecutive days and that pricing errors are only weakly dependent from day to day. Under this assumption, the next day fit is a valid out-of-sample analysis, which measures how well the true discount curve is estimated. As a robustness test, we also include a cross-sectional out-of-sample analysis. Similar to our cross-validation analysis described in more detail below, we split the bond data on a day into ten stratified samples, which have the same maturity distribution, and use nine folds for estimation and the remaining fold for out-of-sample evaluation. This is repeated over all splits to obtain an out-of-sample evaluation for each price. Both out-of-sample analyses lead to the same conclusions. While the cross-sectional out-of-sample analysis does not make assumptions about the dynamics of the fitted curves and errors, we cannot use it for the already fitted curves from FB and GSW and therefore our main analysis uses the next-day out-of-sample analysis.

We provide detailed results for different maturity ranges. More specifically, we report results separately for bonds that are in the following ten maturity buckets: 0-3M, 3M-1y, 1y-2y, 2y-3y, 3y-4y, 4y-5y, 5y-7y, 7y-10y, 10y-20y and > 20y.[20] The number of available bonds in these buckets is not evenly distributed and we observe substantially more prices for short maturity bonds. We also use these maturity buckets for our stratified cross-validation analysis. In order to study the effect of the parameters λ, α and δ, we apply 10-fold stratified cross-validation at the last day of the quarter. This means that we sample randomly without replacement ten non-overlapping folds for each day, such that each fold has the same maturity distribution as the overall sample, i.e. the proportion of bonds in the ten maturity buckets is the same for all folds. Then, we use nine folds for estimation and the remaining fold for evaluation. We repeat this ten times such that each data point is used exactly once for evaluation. Importantly, a naive cross-validation that would sample prices randomly is not appropriate because of the highly unbalanced maturity distribution.

We report average pricing and yield errors for different maturity buckets and time periods. Given an estimated discount curve we report the root-mean-squared errors (RMSE) in weighted prices averaged over time and maturities. More specifically, we estimate the discount curve for day t and report aggregated pricing errors over the full sample and for each day t:[21]

$$\text{RMSE} = \frac{1}{T}\sum_{t=1}^{T}\text{RMSE}_t, \qquad \text{RMSE}_t = \sqrt{\sum_{i=1}^{M_t}\omega_{i,t}\left(P_{i,t} - \hat{P}^g_{i,t}\right)^2}.$$

First, we report the duration weighted pricing errors, see (17), which approximate yield errors, and are also used in the objective function of KR in Equation (9). We label this the duration weighted pricing RMSE. Second, we list the percentage pricing errors of coupon bonds by normalizing the bond prices to $P_{i,t} = 100$ and using equal weights $\omega_{i,t} = 1/M_t$, or equivalently using the weights $\omega_{i,t} = 1/(M_t P_{i,t}^2)$. We label this the relative pricing RMSE. Third, we calculate the yield RMSE

[20] The definition of maturity buckets follows closely the one by the Treasury and Liu and Wu (2021). It takes into account that the available bonds with longer maturities are much sparser.

[21] We indicate the parameters and quantities for a specific day by the additional subscript t.

between the observed yields of coupon bonds and the model implied yields, $\sqrt{\frac{1}{M_t}\sum_{i=1}^{M_t}\left(Y_{i,t}-\hat{Y}_{i,t}^{g}\right)^2}$. The yield and duration weighted pricing errors account for the effect of longer maturities. Fourth, we report a maturity bucket weighed yield error based on the ten maturity buckets defined above. The maturity weighted yield error can be interpreted as an average yield error for the hypothetical case that the number of short maturity bonds would be the same as the number of long maturity bonds.

Out-of-sample RMSE of yields and discount bond prices are a meaningful way to evaluate how close the estimated discount curves are to the true discount curve. The observed Treasury prices represent the true no-arbitrage prices plus noise. A flexible model can obtain in-sample RMSE of zero by overfitting the noise. In fact, our KR estimator obtains a perfect in-sample fit for a vanishing smoothness parameter λ, which corresponds to setting $\Lambda = 0$ in Theorem 2. However, we show that such an overfitted model has a large out-of-sample error, and hence does not provide a good representation of the true discount curve. We show that KR is the closest to the true discount curve among the benchmark methods, since it has uniformly smaller out-of-sample RMSE among all maturity buckets. This result does not depend on specific assumptions of the noise, as it holds for both, the next-day and cross-sectional, out-of-sample analyses.

3.2 Parameter Selection

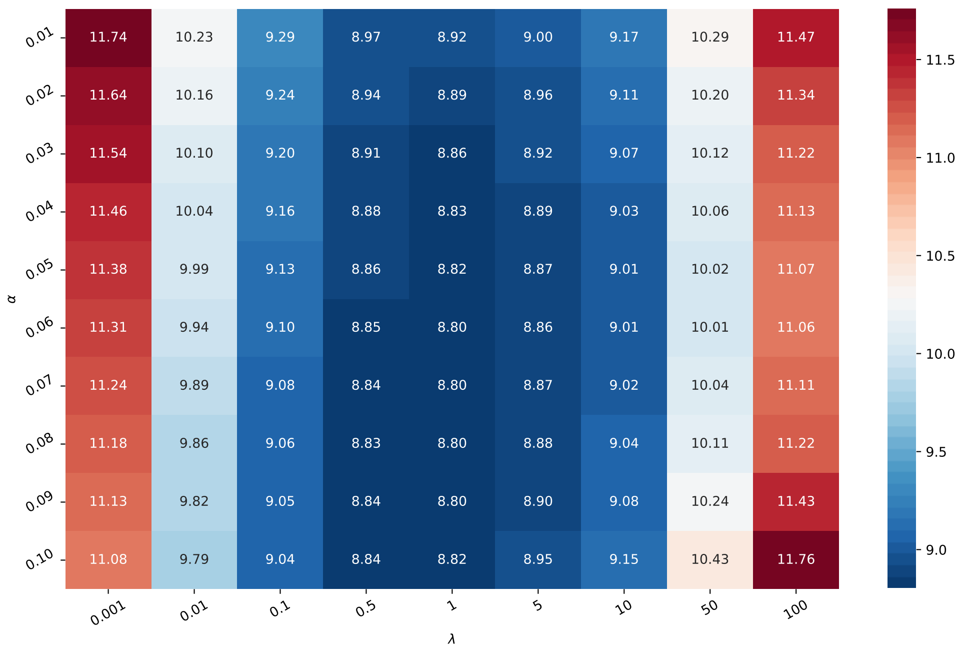
We start our analysis by selecting the optimal parameter values for our KR method. The KR method is completely specified up to the smoothness, maturity weight and tension parameters λ, α and δ. We select the optimal set of parameter values in a data-driven way by applying 10-fold stratified cross-validation, which ensures the same maturity distribution in each fold.[22] For presentation purposes we show the cross-validation results for combinations of two of the three parameters, while the third parameter is set to the overall optimal value. We have verified that our results remain optimal when searching over a three-dimensional grid. We use quarterly data to speed up the calculation.[23]

Figure 2 shows the RMSE in yields for different choices of λ and α, while we set the tension parameter to $\delta = 0$. First, the choice of α has a negligible effect on the fitted yields if it within a reasonable range. This is expected as α corresponds to the yield of an infinite-maturity discount bond as formalized in Theorem A.2 in the appendix. As we will study in more detail, the interpolation fit for finite maturities for a flexible estimator is not affected by the choice of the infinite-maturity yield, while it mainly affects the extrapolation for maturities that are not observed. In contrast, the choice of the smoothness parameter λ has a large effect on the estimation. The optimal value is attained for $\lambda = 1$. The findings carry over to the duration weighted and relative pricing errors

[22] In order to render the smoothness parameter λ comparable across the time series, we normalize it for each day as described in Appendix A.8.2.

[23] For presentation purposes we only show the results for selected grid points. We have confirmed that the results are robust to a finer grid. We have also estimated the optimal parameter choice for each day separately, that is, we have allowed for different λ_t, α_t and δ_t for each day. The results are essentially the same as for the case of using the same parameters for each day. All of these robustness results are available upon request.

Figure 2: Cross-validation YTM RMSE for λ and α



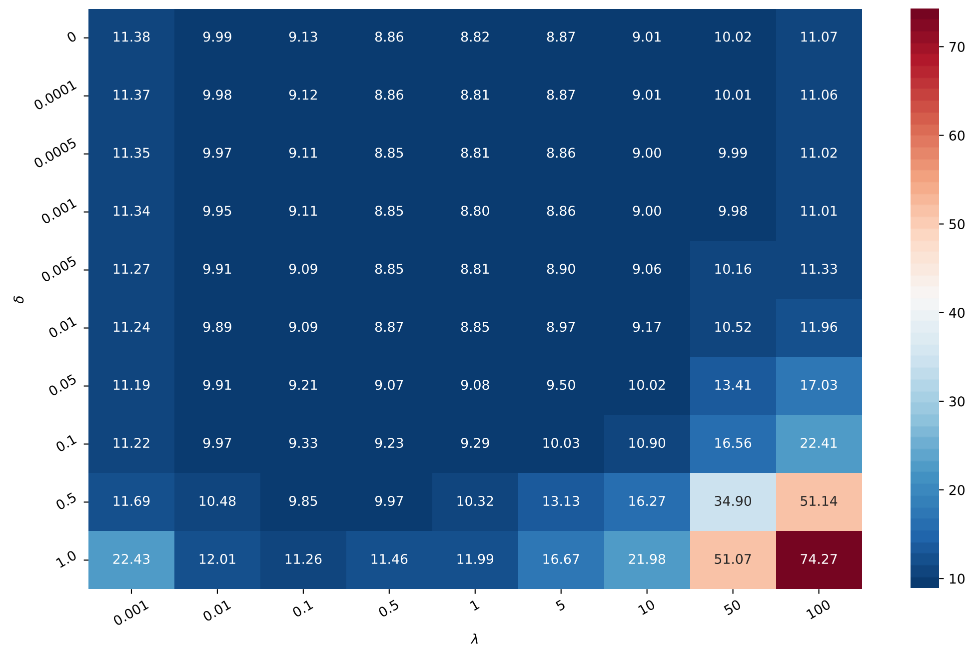
This figure shows the cross-validation yield RMSE in basis points (BPS) for our KR method as a function of the smoothness parameter λ and maturity weight α. The tension parameter is set to $\delta = 0$. The results are calculated using quarterly data from June 1961 to December 2020.

as shown in Figures A.1 and A.2 in the appendix. We conclude that an infinite-maturity yield of $\alpha = 0.05$ and smoothness parameter of $\lambda = 1$ provides a (close to) optimal model with tension parameter zero.

Figure 3 shows the effect of the tension parameter for varying λ for fixed maturity weight $\alpha = 0.05$. The cross-validated yield error is increasing for most values of δ, while the value of $\lambda = 1$ remains optimal. Interestingly, a very small tension parameter of around $\delta = 0.001$ can lead to minor improvements. However, the effect is economically negligible and a parsimonious model that is close to optimal would set $\delta = 0$. It makes sense that the tension parameter δ is not relevant for fitting a discount curve. A larger value of δ only penalizes the first derivative of the discount curve, while a smaller value of δ penalizes more the second derivative of the discount curve, which enforces a smoother yield curve without kinks. In other words, a good term structure model needs a smooth yield curve. Figures A.3 and A.4 in the appendix show that the same results hold for the duration weighted and relative pricing errors.

Throughout the paper we illustrate the various estimated models on the three representative days 1961-06-30, 1986-06-30 and 2013-12-31. Figure 4 shows the effect of the parameters λ, α and δ on estimated yield curves for 1986-06-30, while Figure A.5 in the appendix collects the results for the other two example days. The effect of the smoothness parameter λ is as expected. Smaller values of λ lead to yield curves with more curvature, which are more likely to overfit the data. In contrast excessive values of λ generate curves which are not sufficiently flexible. The infinite-maturity yield α does not affect the shape of the yield curve for shorter maturities and only has

Figure 3: Cross-validation YTM RMSE for λ and δ



This figure shows the cross-validation yield RMSE in basis points (BPS) for our KR method as a function of the smoothness parameter λ and tension parameter δ. The maturity weight is set to $\alpha = 0.05$. The results are calculated using quarterly data from June 1961 to December 2020.

a subtle effect on long maturity yields. Increasing the tension parameter δ leads to "kinks" in the yield curve, which are likely to overfit outlier prices.

Our findings show that discount curves with small out-of-sample errors need to be sufficiently smooth. This suggests that the true discount curve itself has to be sufficiently smooth with respect to our smoothness measure. A perfect in-sample fit by setting $\lambda = 0$ results in large out-of-sample errors. On the other hand, inflexible curves with large λ also lead to large out-of-sample errors. This suggests that the true discount curve cannot be too simplistic neither.

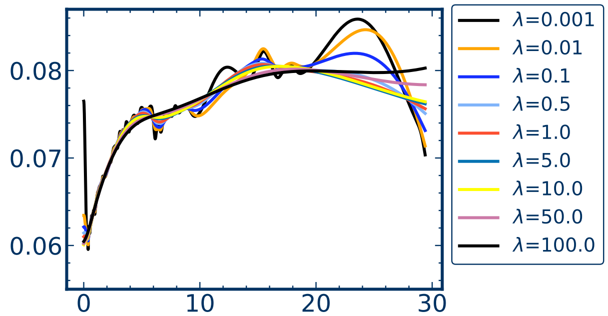
Based on the cross-validation, we define our baseline KR model as $\lambda = 1$, $\alpha = 0.05$ and $\delta = 0$, which we use in the next sections for our comparison study.

3.3 Comparison Study

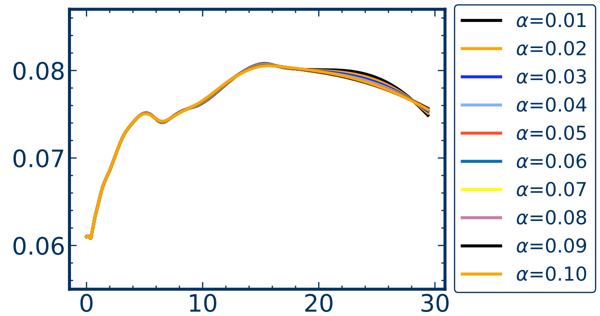
Our KR method uniformly dominates all other approaches in any metric and for any selected subsample. Figure 5 reports the yield RMSE and duration weighted and relative pricing RMSE in- and out-of-sample for the different maturity buckets. First, we observe that the KR method obtains smaller pricing and yield errors than all the other methods for any maturity bucket.[24] Second, the qualitative results for the in- and out-of-sample analysis are the same, while as expected the relative difference between the flexible non-parametric methods and the parametric methods is larger for the

[24] The out-of-sample yield errors for bonds with maturity larger than 20 years are essentially identical for KR and LW. However, a cross-sectional out-of-sample analysis reveals smaller yield errors for KR for this maturity bucket.

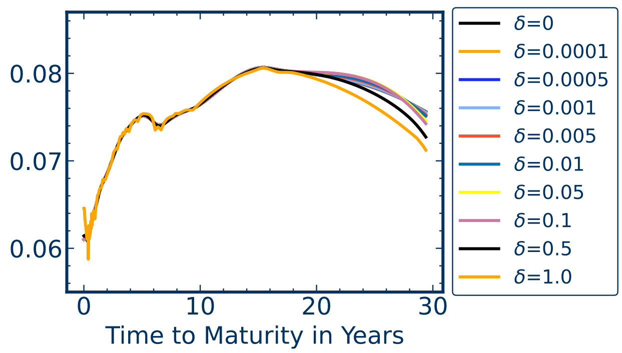
Figure 4: KR yield curve estimates as a function of parameters



(a) Varying λ, fixed $\alpha = 0.05$ and $\delta = 0$



(b) Varying α, fixed $\lambda = 1$ and $\delta = 0$



(c) Varying δ, fixed $\alpha = 0.05$ and $\lambda = 1$

This figure shows yield curve estimates with KR for various combination of parameters on the example day 1986-06-30. Subfigure (a) varies the smoothness parameter λ for fixed values $\alpha = 0.05$ and $\delta = 0$. Subfigure (b) varies the maturity weight α for fixed values $\lambda = 1$ and $\delta = 0$. Subfigure (c) varies the tension parameter δ for fixed values $\lambda = 1$ and $\alpha = 0.05$.

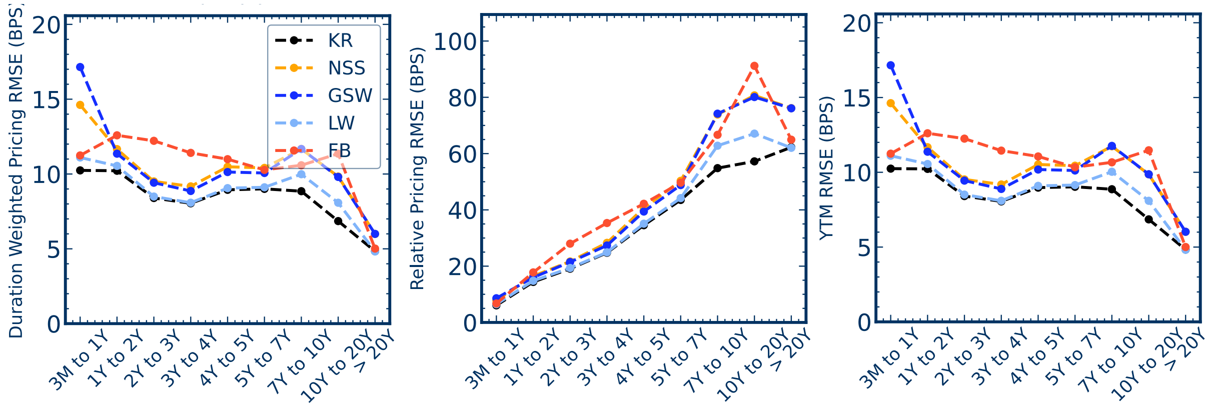
in-sample fits. Third, the second-best performance is observed for the LW method, which indicates that flexible methods are needed to adequately approximate the yield curve. Importantly, the LW method has larger errors for a maturity range between 7 to 20 years, where the observed prices are sparse. We will revisit and explain this shortcoming when discussing the instability of the LW method. Fourth, FB performs relatively well for short maturities, but its performance deteriorates for longer maturities.[25] Fifth, we confirm the known observation that the NSS and GSW models are not well suited to estimate the short end of the yield curve. Last but not least, we emphasize the economic importance of the superior fit of the KR method as even small errors in yields can have large economic effects.

As expected the duration weighted pricing RMSE and yield RMSE are essentially identical. By construction, the relative pricing RMSE puts larger emphasis on longer maturities as the same yield error has a larger effect for a longer maturity bond. Hence, different metrics put different weights

[25] The FB estimated forward curve, that was shared with us, includes bonds with maturity less than 3 months in the estimation. Hence, not surprisingly it can obtain slightly smaller pricing errors in the evaluation for those bonds. For the maturity bucket up to 3 months, our KR method has the smallest pricing errors among all the methods that do not include those bonds in the calibration.

Figure 5: Pricing errors for different maturities

Panel A: Out-of-sample

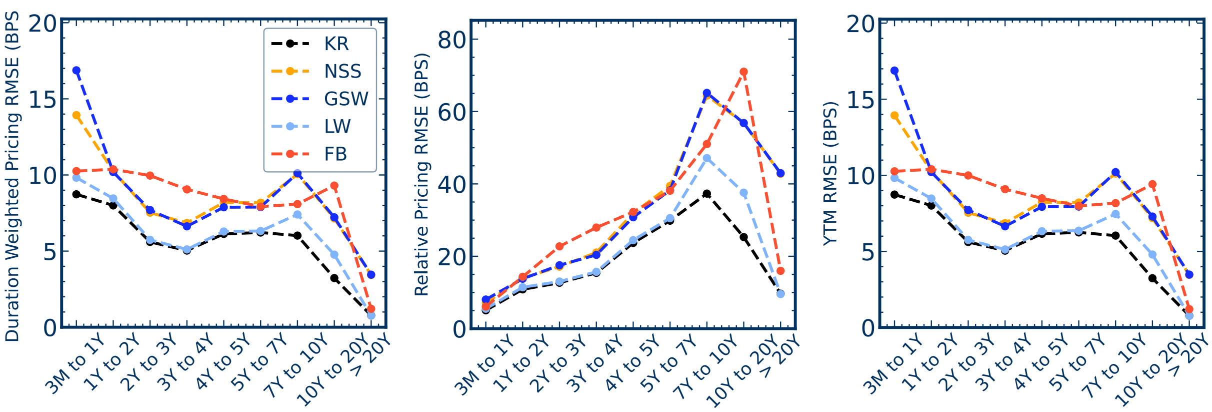


(a) Duration Weighted Pricing RMSE

(b) Relative Pricing RMSE

(c) YTM RMSE

Panel B: In-sample



(d) Duration Weighted Pricing RMSE

(e) Relative Pricing RMSE

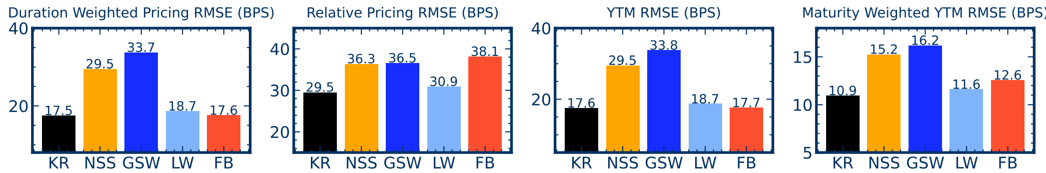
(f) YTM RMSE

This plots shows evaluation metrics calculated in- and out-of-sample on the short sample from June 1961 to December 2013. Out-of-sample errors are calculated using curves estimated at t to price securities observed on the next business day. The top panel shows the out-of-sample results and the bottom panel the in-sample results. The evaluation metrics are the RMSE of duration weighted and relative pricing errors and yield errors. All errors are in basis points (BPS). We evaluate the model for different maturity buckets. The KR method uniformly outperforms all benchmark estimators.

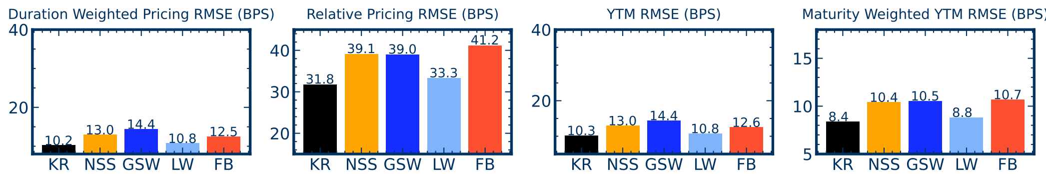
on different parts of the maturity spectrum. The comparison analysis in the main text focuses on the shorter sample until 2013 in order to include FB. Tables A.1 and A.2 in the appendix give a comprehensive summary and show that all the results carry over to the full sample until 2020. Table A.2 also shows that the cross-sectional out-of-sample analysis leads to the same findings as the next-day out-of-sample analysis in the main text.

Our results are not driven by outliers and are robust to various filters. Figure 6 shows the aggregate out-of-sample RMSE for different evaluation metrics for different evaluation samples. Most importantly, our KR continues to uniformly dominate all other methods. First, we consider

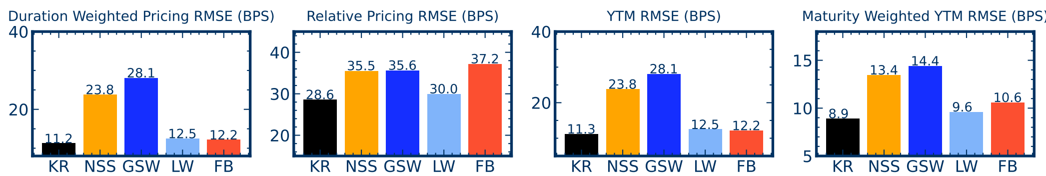
Figure 6: Out-of-sample results by evaluation metric for different filters



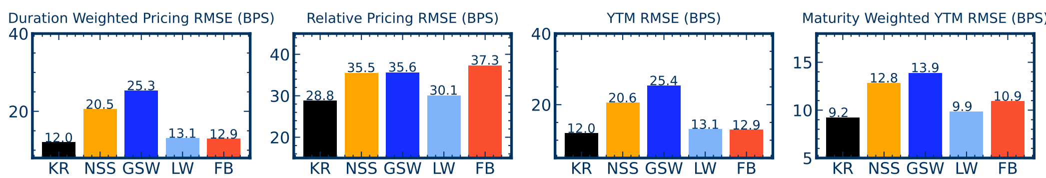
(a) Full data (no outlier removal)



(b) 3-Month Filter



(c) KR Filter



(d) NSS Filter

This plot shows aggregated evaluation metrics calculated out-of-sample on the short sample from June 1961 to December 2013. Out-of-sample errors are calculated using curves estimated at t to price securities observed on the next business day. Columns correspond to duration weighted pricing RMSE, relative pricing RMSE, YTM RMSE, and maturity-weighted YTM RMSE. All numbers are in basis points (BPS). The top panel correspond to results evaluated on the full data without filtering. The second panel shows results evaluated on data where securities maturing within three months are removed. In the third panel, results are evaluated on the sample for which an NSS filter is used to remove outlier securities, whose YTM fitting errors are at least three standard deviation away from the average YTM fitting error calculated using NSS curves in the same cross-section. The last panel collects the results evaluated on the sample for which KR is used to remove outlier securities with the same three-standard-deviation rule as for the NSS filter. KR outperforms other methods in term of out-of-sample fitting quality according to all four evaluation metrics on datasets with and without outlier removal.

the aggregate metrics on the full data including the first 3 months of maturities. This increases in particular the yield errors of the parametric NSS and GSW estimates. Second, we report the aggregate statistics with the 3-month filter, which particularly benefits the parametric approaches. The relative performance stays the same, while we note that all methods seem to have relatively large yield errors for the very short maturity bonds. The relative pricing errors are the largest for FB as this method has the worst fit for long maturity bonds. As the number of bonds with longer maturities is relatively small, the equally weighted yield error for FB does not accurately reflect this issue. The maturity-bucket-weighted yield error combined with the 3-month filter reveals that the FB method provides actually the worst fit of the yield curve when excluding the short end. The third and fourth panel of Figure 6 remove outliers based on KR or NSS filters. The results are qualitatively similar, confirming that it is not particular outlier values that drive our findings. By construction, the NSS filter should improve the fit of the NSS method, but it does not affect the relative ranking. Similarly, the KR filter improves the fit of the non-parametric methods, but does not change the relative ranking. Last but not least, our estimate of NSS and the estimate of GSW are close. However, it seems that using our data and a careful implementation of the optimization of NSS lead to slight improvements. Figures A.6, A.7 and A.8 in the appendix show that the findings are the same in-sample and for different maturity buckets. Our simulation study in Section C in the appendix further supports the empirical findings.

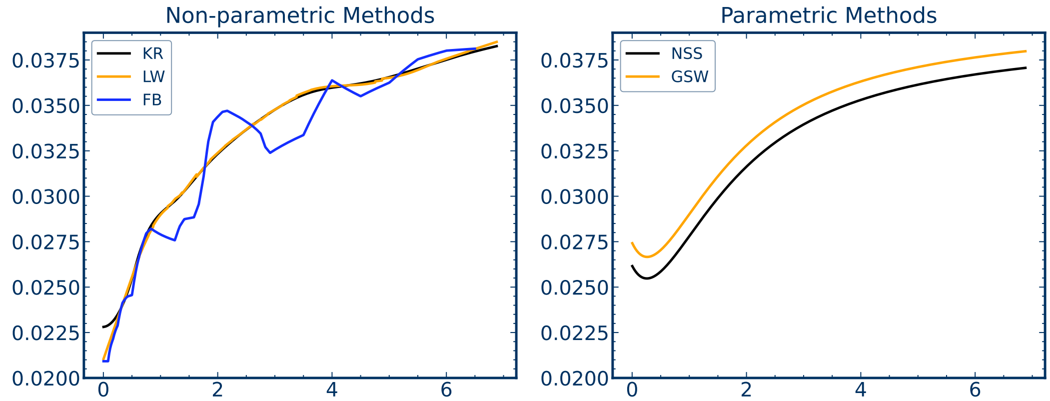
The estimated yield curves in Figure 7 for the three representative example days shed further light on why we observe these pricing results. The left subplots show yield curves for the three non-parametric methods KR, LW and FB and on the right we show the estimates for the parametric GSW and NSS models. The first example day is in the early part of the sample and has shorter maturity bonds. The piece-wise constant FB forward curve leads to visible kinks and the least smooth yield curve. Figure A.9 in the appendix shows the observed and fitted bond yields for the three days. The FB method is visibly overfitting individual bonds. By construction, the NSS and GSW curves are very smooth, but misspecified for the short end and lack the flexibility for more complex shapes. As expected the NSS and GSW curves are very close, while the level of the NSS yield curve is slightly closer to the non-parametric benchmarks. Importantly, the first and third day illustrate the known issues of the NSS model for the very short end. The functional form strongly deviates from the non-parametric estimates and the actually observed coupon bond prices and explains the extreme pricing errors at the short end.

The KR method optimally trades off smoothness against flexibility. The LW method is also flexible and seems to capture similar shapes. However LW curves may not be smooth because of the jumps in its bandwidth selection, and as we will show LW is more prone to overfitting the data. The jumps for the LW yield curve are particularly visible on 1986-06-30 for the sparse maturity range of around 17 years. Figure A.10 in the appendix shows the LW bandwidth estimates. Whenever the density of observed coupon prices changes, it leads to excessive changes in the bandwidth.

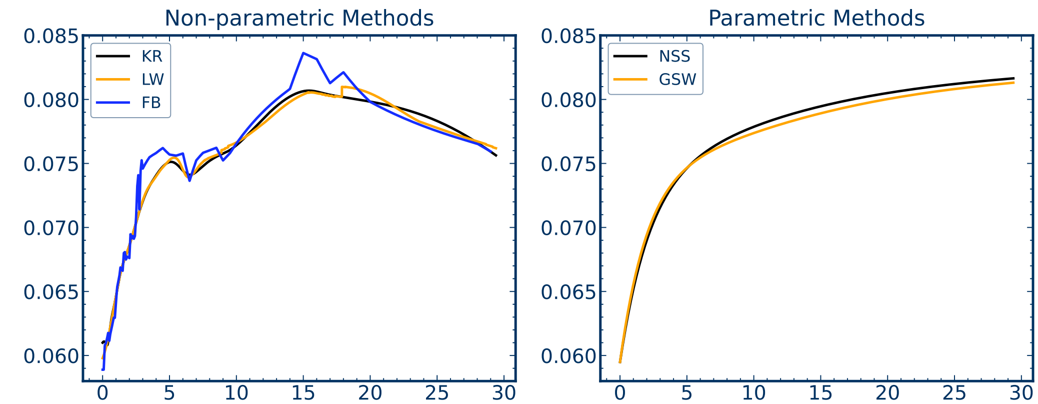
Our KR method gives the overall smoothest curve, while at the same time has the smallest pricing errors. Figure 8 shows the aggregate tension and curvature measures for the different estimation approaches for different maturity buckets.[26] All yield curves are continuous, which results in fairly similar first derivatives of the discount curve. As expected, the FB curve is the least smooth as technically it is not even twice weakly differentiable. The second least smooth curve is LW due to its non-continuous bandwidth selection. Interestingly, the parametric NSS and GSW are less smooth than KR at the very short end, due to the misspecified estimates as illustrated in

[26]We compute these measures by numerically discretizing the integrals $\frac{1}{|\mathcal{R}|}\int_{\mathcal{R}} g'(x)^2\,dx$ and $\frac{1}{|\mathcal{R}|}\int_{\mathcal{R}} g''(x)^2\,dx$ over the maturity ranges $\mathcal{R}$, where $|\mathcal{R}|$ denotes the length of $\mathcal{R}$, and where we discretize the derivatives g' and g'' by the respective difference quotients.

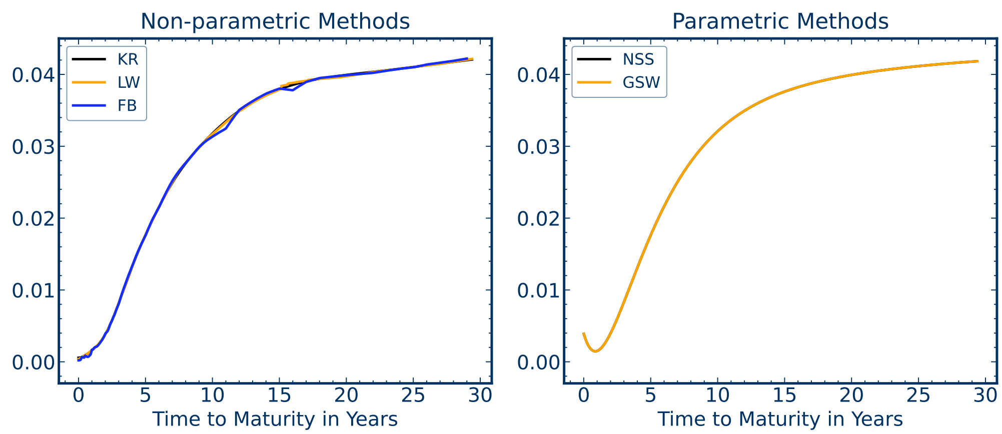
Figure 7: Yield curve estimates of different methods



(a) 1961-06-30



(b) 1986-06-30

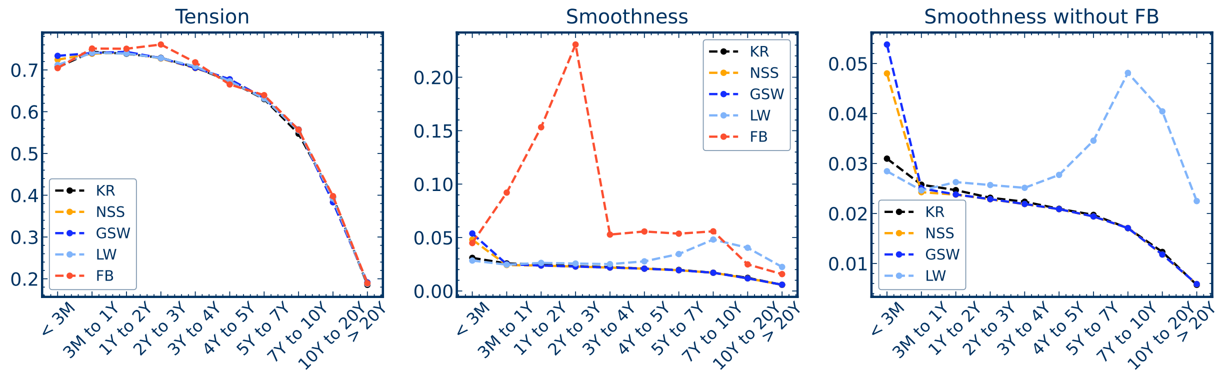


(c) 2013-12-31

This figure shows the yield curve estimates for the three representative example days: 1961-06-30 (top panel), 1986-06-30 (mid panel), and 2013-12-31 (bottom panel). The left and right columns show estimates for non-parametric and parametric methods.

Figure 7. In summary, the objective function of KR results in yield curves that are as smooth as a parametric model, while having substantially smaller pricing errors. The discretized derivatives in

Figure 8: Tension and curvature for different maturities



This plot shows the discretized measures for tension (left panel) and curvature (mid panel) for different maturity ranges. The right panel shows the curvature measure for KR, NSS, and GSW only. The discrete derivatives use monthly granularity. Results are calculated on the short sample from June 1961 to December 2013.

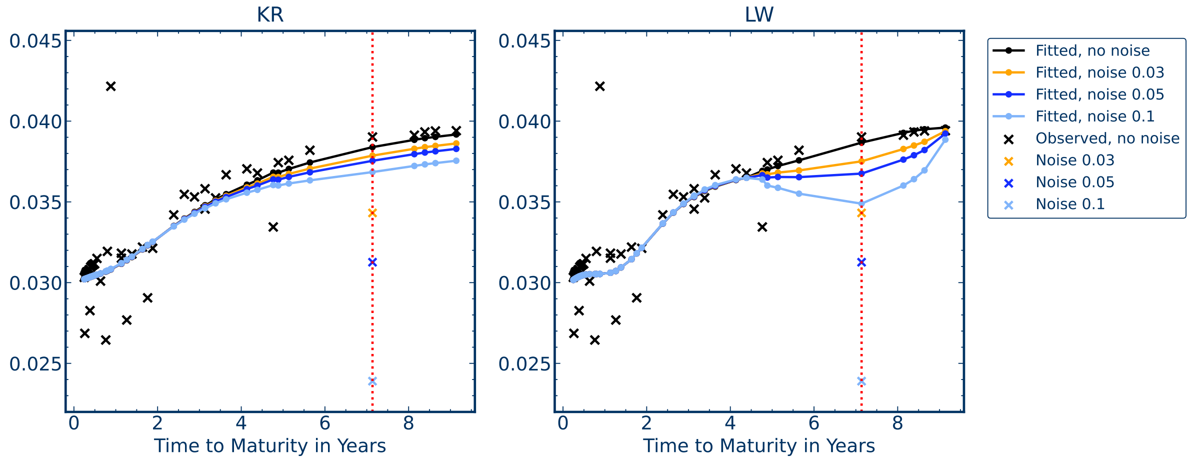
Figure 8 are based on monthly maturities, while the corresponding results based on daily maturities are in Figure A.11 in the appendix. By construction, the piece-wise constant FB forward curve has daily numerical derivatives that are zero for most days and hence appears to be smoother under this measure. In contrast, the daily bandwidth changes for LW blow up the daily derivative of its yield curve. The smoothness measures of the KR curve are not affected by moving from monthly to daily discretized derivatives.

3.4 Robustness

Our KR estimator combines flexibility with robustness to outliers. Figure 9 illustrates the effect of the non-parametric estimation after contaminating the data with a single large outlier. We report the observed and the model implied coupon bond prices for KR and LW for the representative example day 1963-06-28. The discount curves are once estimated on the actual observed prices and once after increasing a single bond price by 3, 5 or 10%. The left subplot shows that the estimates of the KR method are barely changed after adding this single outlier. In contrast, the long maturity yield estimates of the LW method in the right subplot are strongly biased downwards after adding a single outlier data point.

Non-parametric kernel estimators like the LW method are inherently local in nature. Intuitively, the LW method takes a weighted average of the yields of eight nearby bonds. If one of these bonds is an outlier, it can lead to a massive bias for the estimates of the neighboring points. This problem becomes particularly severe in only sparsely populated maturity regions where fewer data points are available and one outlier can contaminate a large fraction of the yield curve. The local nature of kernel estimators also implies that the fit for short maturities is decoupled from the fit for long maturities. In contrast, our KR method is conceptually different as it takes advantage of a global smoothness reward. Therefore, the KR estimator only allows for more curvature if this in turn reduces the overall pricing error. Our KR estimator shares the robustness properties of regularized estimators. We conclude that our KR method provides a robust estimation, while local estimators

Figure 9: Yield estimation with outlier contamination



This plot shows the observed and fitted yields given by KR (left panel) and LW (right panel). The discount curves are either estimated on the observed prices or after increasing the price of a single bond by 3, 5 or 10%. The maturity of the contaminated security is marked with red vertical lines. The results are for the representative example day 1963-06-28. KR is substantially more robust to noise compared to LW.

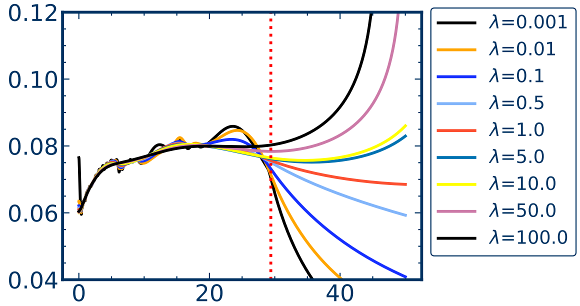
like LW are more prone to overfitting singular outliers. As a consequence of its superior pricing and robustness properties, the KR estimator naturally lends itself as a tool to filter outliers in a data-driven way.

3.5 Extrapolation

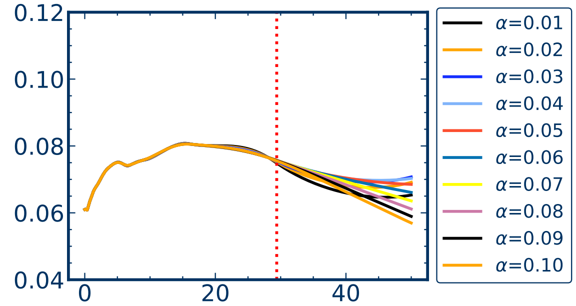
The yield curve estimation up to the maximal observed maturity can be interpreted as an interpolation problem. In this section we study the problem of estimating a yield curve beyond the maximal observed maturity, which is conceptually an extrapolation problem. Figure 10 shows the estimated yield curves for KR as function of the different parameters on 1986-06-30, while Figure A.14 in the appendix collects the results for the other example days. We plot the yield curve for up to 50 years. The maximal observed maturity is indicated by the red horizontal line, and hence the curves beyond it are extrapolated. We observe that all three parameters have a much more pronounced effect on the extrapolation region.

Recall that the maturity weight α corresponds to the yield of a discount bond with infinite maturity. As the yield of such a bond is obviously not observed, it becomes a choice parameter. Panel (b) in Figure 10 shows that the choice of α essentially only affects the extrapolation region, but does not change the interpolation part. This makes sense as intuitively the KR method optimally joins the observable bond yields with the infinite-maturity yield, which cannot be learned from the observed prices. The extrapolation depends on our prior for the infinite-maturity yield and hence is a choice by the researcher. The situation is different for λ and δ as those parameters can be optimally estimated from the observed prices. Our optimal baseline model uses $\lambda = 1$, $\delta = 0$ and $\alpha = 0.05$. In this case, the extrapolated curves would smoothly connect the observed yields with the infinite-maturity yield of 0.05.

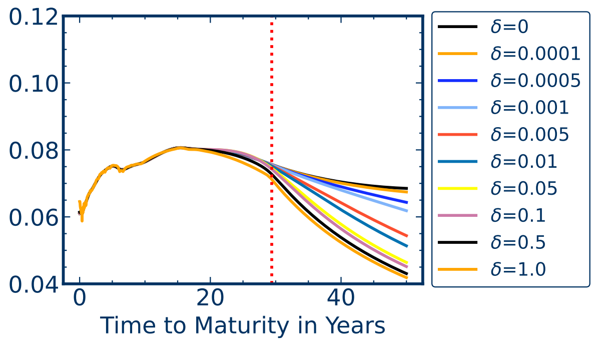
Figure 10: Extrapolated KR yield curves



(a) Varying λ, fixed $\alpha = 0.05$ and $\delta = 0$



(b) Varying α, fixed $\lambda = 1$ and $\delta = 0$



(c) Varying δ, fixed $\alpha = 0.05$ and $\lambda = 1$

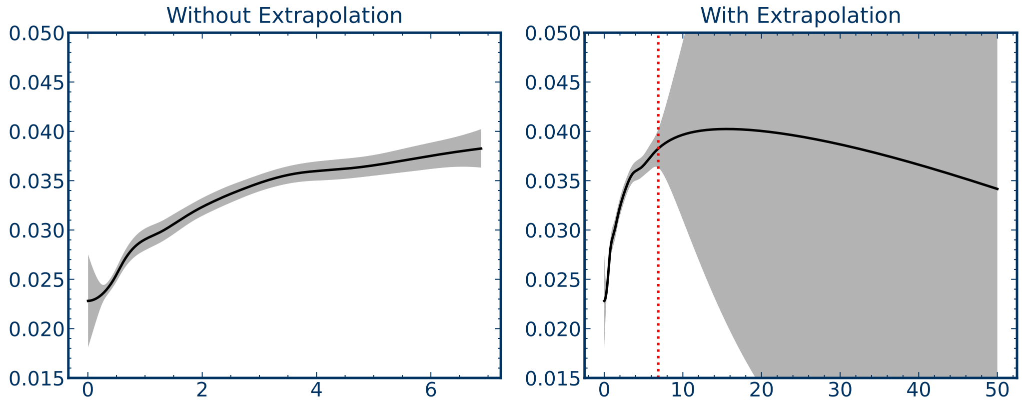
This figure shows yield curve estimates with extrapolation to 50-year maturity for KR as a function of parameters on the example day 1986-06-30. The region to the right of the red dashed vertical line is the extrapolation region. Subfigure (a) varies the smoothness parameter λ for fixed values $\alpha = 0.05$ and $\delta = 0$. Subfigure (b) varies the maturity weight α for fixed values $\lambda = 1$ and $\delta = 0$. Subfigure (c) varies the tension parameter δ for fixed values $\lambda = 1$ and $\alpha = 0.05$.

Non-parametric estimators are designed for interpolation. We want to emphasize that this is not a weakness of our approach but a conceptual point. Extrapolation requires extra assumptions on the extrapolation region, either in the form of imposed functional restrictions as in the NSS model, or by an exogenous choice of α in the KR model. However, such assumptions can generally not be verified by the observed data in the interpolation region. The advantage of our KR method is that the only exogenous choice parameter needed for the extrapolation has a clear economic interpretation as the infinite-maturity yield.

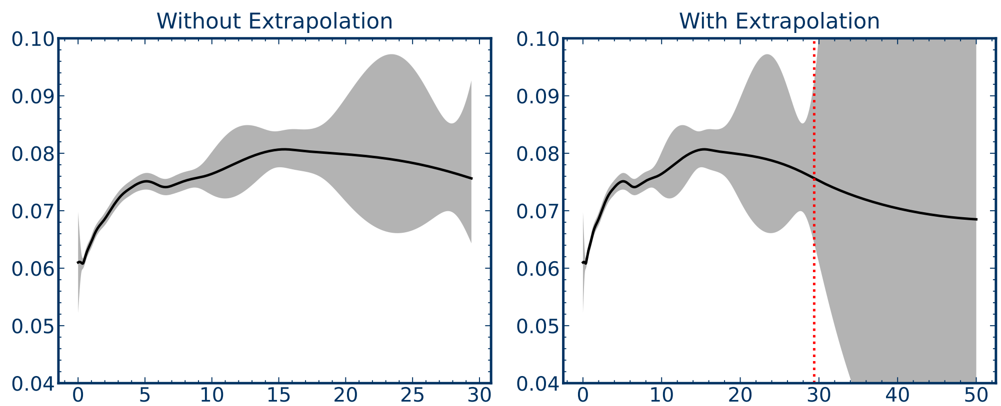
3.6 Statistical Inference

Our distribution theory provides guidance on the quality of our point estimates. Figure 11 shows the 99% confidence bands for the yield curve estimates based on Theorem 4 for the three representative example days. The left subplots depict the results up to the maximal observed maturity. We observe wider confidence intervals for maturity regions with less observed prices, for example the maturities over 20 years on 1986-06-30. The confidence intervals also increase if there is larger dispersion in prices, which is particularly prominent for the very short end. The wider confidence intervals also

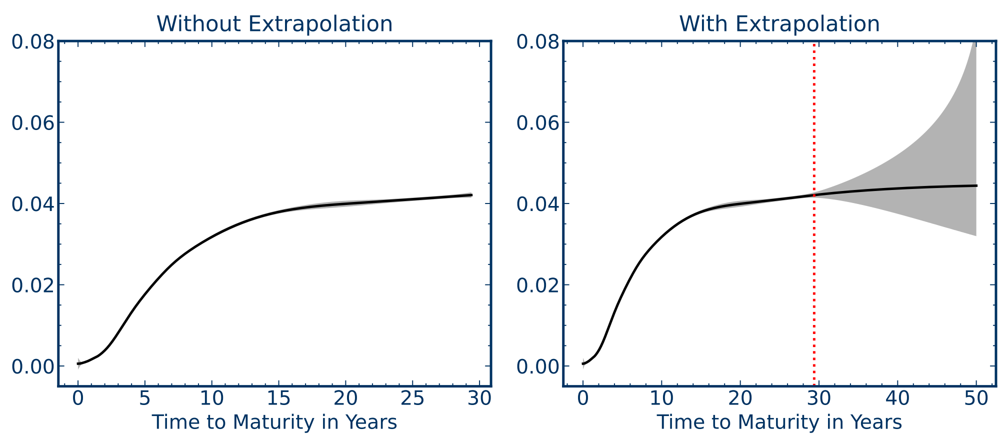
Figure 11: KR yield curve confidence bands



(a) 1961-06-30



(b) 1986-06-30



(c) 2013-12-31

The figure shows 3-standard-deviation confidence bands for yield curve estimates given by the KR model under the Gaussian process assumption. The panels correspond to the example dates 1961-06-30, 1986-06-30, and 2013-12-31. The left column shows results without extrapolation, and the right column includes extrapolation results for up to 50-year maturity.

reflect the regions where different estimators disagree. The third example day is comparatively "easy" to fit and hence all models agree on it. As a result the confidence bands are very tight. Figure A.15 in the appendix shows the corresponding confidence bands for the discount curve estimates with the same findings. The confidence intervals for the implied prices of individual coupon bonds in Figure A.16 further highlight the uncertainty for sparsely populated maturity regions.

The distribution theory also confirms the conceptual point that yield curve estimation is an interpolation and not extrapolation problem. The right subplots in Figure 11 display the confidence intervals for yield curve estimates up to 50 years. Given the observed prices, the extrapolation quickly leads to exploding confidence bands. In other words, the observed data can tell us little about yields with very long maturity.

3.7 Results over Time

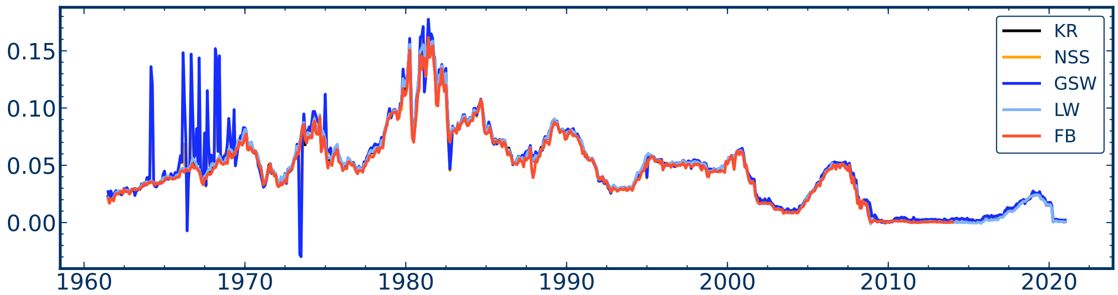
The relative comparison results are robust over time. Figures A.12 and A.13 show the yield errors for different maturity buckets for each month. The parametric GSW and NSS estimates lead to particularly large yield errors for bonds with maturities shorter than one year. While these errors are more pronounced in the earlier part of the sample, they also persist throughout the full sample. The pricing errors of FB are visibly larger for longer maturity bonds. The non-parametric KR and LW methods have the smallest pricing errors throughout the full sample. The variation in pricing errors for longer maturities is positively related to the overall magnitude of yields over time.

Figure 12(a) shows the estimates of the 30-day yield over time for all methods. The one-month yield is an important input for academic research and a crucial economic indicator. The most obvious observation is that the GSW and NSS estimates for the one-month yield cannot be used in good faith for the first half of the sample. The FB estimates are closer to the KR estimates, but exhibit some smaller irregularities. We conclude that we need to use one of the three non-parametric methods KR, LW or FB for economically meaningful short rate dynamics.

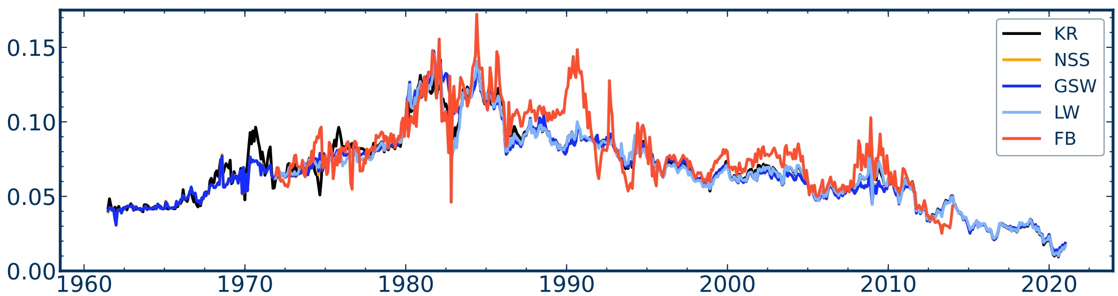
The dynamics of long maturity rates are also sensitive to the choice of estimator. Figure 12(b) shows the 10-year forward rates for a one year investment over time. We use forward rates instead of yields as long maturity yields are the average of forward rates and hence the effects for the long maturity spectrum are partly averaged out in the yields. The estimates of FB deviate excessively and suggest that this method is not reliable for long maturities. This is further confirmed in the out-of-sample yield errors in Figure A.13, which spike during the times when the FB estimates deviate from the other methods.

The KR and LW estimates of the 10-year forward rate are close most of the time, but LW is not that stable and has several instances where it overshoots. For a better comparison, Figure 12(c) shows the forward rates for only these two methods. There are visible differences for some days, and a close look at Figure A.13 reveals that the out-of-sample yield errors of LW also exceed those of KR on those days. In order to better understand these results, we inspect December 2008, which has a visible spike for LW relative to KR. The yield curve plots in Figure 13 illustrate the

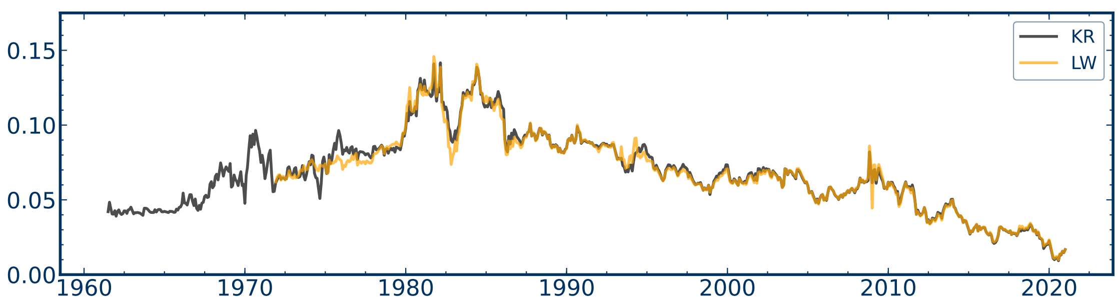
Figure 12: Short and long maturity rate estimates over time



(a) 30-Day yield estimates



(b) 10-Year forward rate estimates

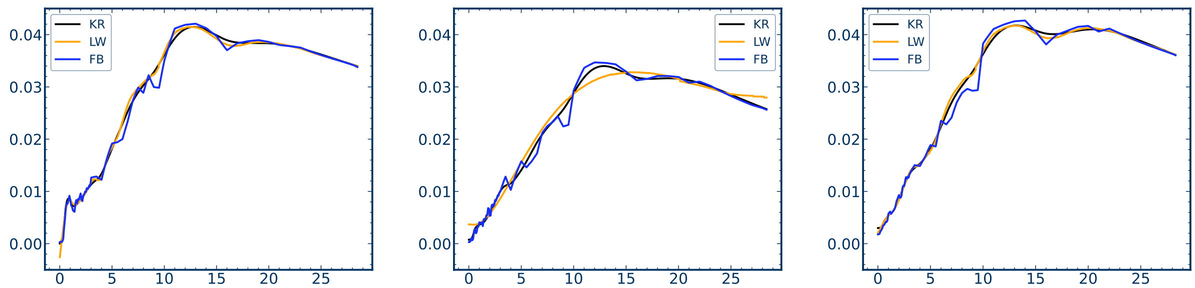


(c) 10-Year forward rate estimates (only KR and LW)

The plot shows the time-series of short and long maturity rate estimates for different methods over time. Subplot (a) displays the annualized 30-day yield estimates. Subplot (b) shows the 10-year forward rate for a one-year horizon, that is the the time t forward rate for a one-year investment from $t+10$ to $t+11$ locked in at time t. Subplot (c) is the same plot as (b) but only for the KR and LW estimates. The time-series of FB ends in December 2013, while the other four time-series are available until December 2020. Prior to the 1970s, the 10-year forward rate needs to be obtained via extrapolation which we do for the KR, GSW and NSS methods. Since we do not extrapolate LW, their estimates starts after the 1970s.

dynamic instability of LW for this time period. It shows the estimated yield curves for the end of month of 11/2008, 12/2008 and 1/2009. For both, 11/2008 and 1/2009, all three non-parametric methods estimate a nontrivial shape for longer maturity yields. In fact, the LW estimates have even more curvature than the corresponding KR estimates. However, while the shape of the yield curve estimates of KR (and FB) stays roughly the same on the in-between month 12/2008, the LW estimate substantially changes its shape. This new shape is not only inconsistent with the other methods and over time, but also leads to a spike in pricing errors. This arguably erroneous shape estimate distorts the dynamics of the LW forward rate for that date.[27] This is the consequence of the instability of a non-parametric method that is not sufficiently regularized. We conclude that our KR method provides the most reliable and stable dynamics for interest rates.

Figure 13: Yield curve estimates of non-parametric methods on dates around December 2008



(a) 2008-11-28 **(b)** 2008-12-31 **(c)** 2009-01-30

This figure shows the yield curve estimates for the three days: 2008-11-28 (left panel), 2008-12-31 (mid panel), and 2009-01-30 (right panel).

Given the robustness of the KR method, we already obtain stable time-series for all maturities. We want to note that our method can be easily modified to obtain even smoother temporal behavior. Instead of "anchoring" the discount curves at the vector of 1's, we can use the previous day estimates as a prior. Hence, we would non-parametrically model the change relative to the previous day discount curve. This idea is explained in more detail in Appendix A. This is another attractive feature of our approach.

3.8 Basis Functions

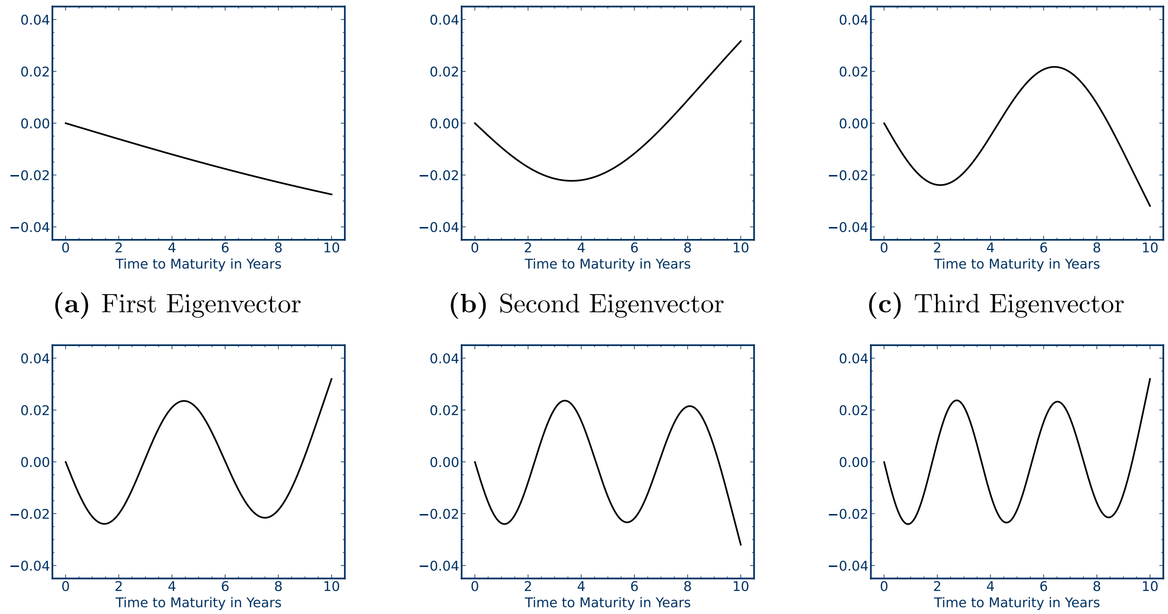
What are the basis functions that lead to the superior fit of our method? As a starting point, Figure A.17 in the appendix plots some kernel basis functions of our baseline model.[28] As predicted by the closed-form expressions in Theorem 2, these are monotonically increasing functions with different slopes and by themselves offer limited additional insight. However, what matters is the space spanned by the kernel basis functions. As our KR estimator simplifies to a simple ridge regression, the smoothness reward will put most weight on the leading principal components of the kernel space. Hence, the eigenvectors of the dominant eigenvalues of the kernel space are the most important basis functions. In the following we restrict our analysis to the first 10 years of maturity as we relate it to the PCA of a panel of estimated discount curves, which without excessive extrapolation is only available for this shorter maturity.[29]

[27] These example days also illustrate the instability of the LW method for the very short end, where the estimated yields are negative.

[28] As stated in Theorem 2, these kernel basis functions correspond to the columns of the kernel matrix $\boldsymbol{K}$.

[29] In the main text we show the results for the panel from November 1971 to December 2013, as this is the subset for which we have estimates of discount bonds up to 10 years of maturity for FB and LW. In the appendix in Figure A.18 we show the corresponding results for the longer panel from June 1961 to December 2013 with up to five years of maturity. This is the highest maturity and longest panel for which we do not require extrapolation for FB and LW. We have confirmed on the panel starting after 1982 that all qualitative findings generalize to maturities longer

Figure 14: Eigenvectors of the KR kernel matrix



(a) First Eigenvector **(b)** Second Eigenvector **(c)** Third Eigenvector

(d) Fourth Eigenvector **(e)** Fifth Eigenvector **(f)** Sixth Eigenvector

This figure plots the eigenvectors of the six largest eigenvalues of the KR kernel matrix $\boldsymbol{K}_{ij} = k(x_i, x_j)$ for the baseline model $\lambda = 1$, $\alpha = 0.05$ and $\delta = 0$.

Figure 14 depicts the eigenvectors of the six largest eigenvalues of the kernel matrix up to 10 years of maturity. The shape of the first two eigenvectors can be recognized as a slope and curvature pattern. More generally, the eigenvectors relate to polynomials of increasing order. The first eigenvector appears to be a polynomial of order one, the second eigenvector is a polynomial of order two, etc. A smaller smoothness parameter λ implies a higher weight on higher order polynomials. Intuitively, a very smooth curve with large λ enforces an approximation with a lower order polynomial function. Recall, that these basis functions are not ad-hoc choices, but the solution of a problem with parameters that are optimally selected from the data.

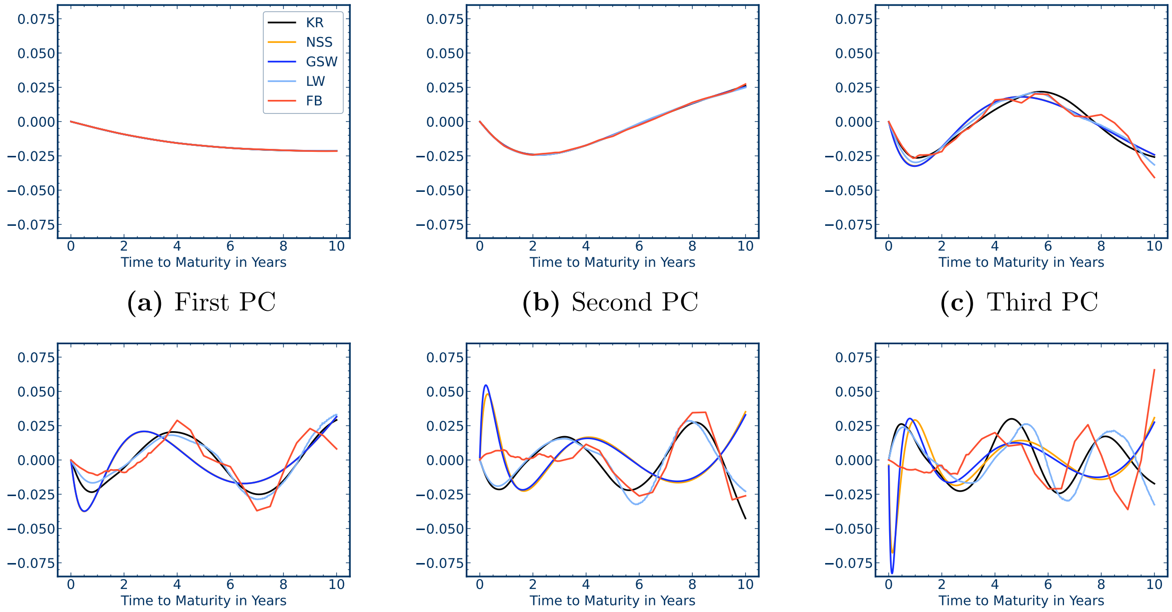
The eigenvectors of the kernel matrix can be interpreted as portfolio weights for discount bonds. As shown in Equations (10) and (11) our KR estimator maps into a portfolio composed of the observed traded bonds. In the hypothetical case, where the traded bonds form a complete set of discount bonds and hence the cash flow matrix C simplifies to the identity matrix, these eigenvectors correspond to the portfolio weights on the discount bonds. By increasing λ, we essentially approximate the full cross-section of discount bonds with a small number of portfolios of discount bonds, whose portfolio weights are given by the eigenvectors in Figure 14.

We recognize the same shapes in the PC estimates in Figure 15. This figure shows the eigenvectors of the six largest eigenvalues of the covariance matrix based on the panels of estimated discount curves for different methods.[30] First, the PC estimates of all methods have the shape of

than 10 years.

[30]While using yields or returns of discount bonds for the PCA estimation will affect the magnitude of the eigenvalues,

Figure 15: Principal Component Analysis of panel of discount bonds



(a) First PC **(b)** Second PC **(c)** Third PC

(d) Fourth PC **(e)** Fifth PC **(f)** Sixth PC

This figure shows the first 6 principal components (PCs) estimated from the panel of discount curves for the five methods KR, GSW, NSS, LW, and FB. The PCs correspond to the eigenvectors of the largest eigenvalues of the covariance matrix of discount bond prices. The panel are the estimated discount bond prices up to 10-year maturity for the sample from November 1971 to December 2013.

polynomials with increasing order. Note that the sign of the eigenvectors is not identified and that some of the eigenvectors need to be flipped. Second, we observe that all methods provide very similar estimates for the first two PCs. FB, GSW and NSS start to differ from KR and LW for the third and higher PCs. There are visible differences between KR and LW for the fourth to sixth PC.

The non-parametric basis functions of the KR method are the same basis functions that explain the variation in a panel of discount curves. The differences in estimated yield curves is reflected in different estimates of higher order PCs. The estimates of NSS, GSW and FB lead to particularly misspecified fifth and sixth PCs. The occurrence of level, slope and curvature type patterns in a panel of discount curves is not a coincidence. It is rather the consequence that these are the type of basis functions that best explain the functional form of the yield curve. As an illustrative example consider the case where the KR estimator would be limited to using only the first two PCs of the kernel matrix. By construction, these PCs would appear as slope and curvature "factors" in the panel of estimated discount bonds.

Filipović, Pelger, and Ye (2022) explore the implications of these basis functions for the term structure premium and factor structure of the cross-section of bond returns. In more detail, they

it has a negligible effect on the shapes of the eigenvectors. This aspect is studied in depth in Filipović, Pelger, and Ye (2022).

explain the relationship between fitting a non-parametric model with few basis functions and the implication for a low dimensional factor model for a panel. This allows us to unify the term structure asset pricing literature with the fundamental non-parametric estimation problem of the discount curve.

4 Economic Importance

The estimated yield curve is a key factor for many research areas in economics and finance. This includes, among others, understanding the term structure effects, bond term premia, return forecasting, exchange rates, monetary policy and broadly asset pricing and derivatives. A more precise and robust estimation of the yield curve benefits the downstream applications. The economic implications of the choice of discount curve estimation method depend on the applications, and on which methods are compared with each other.

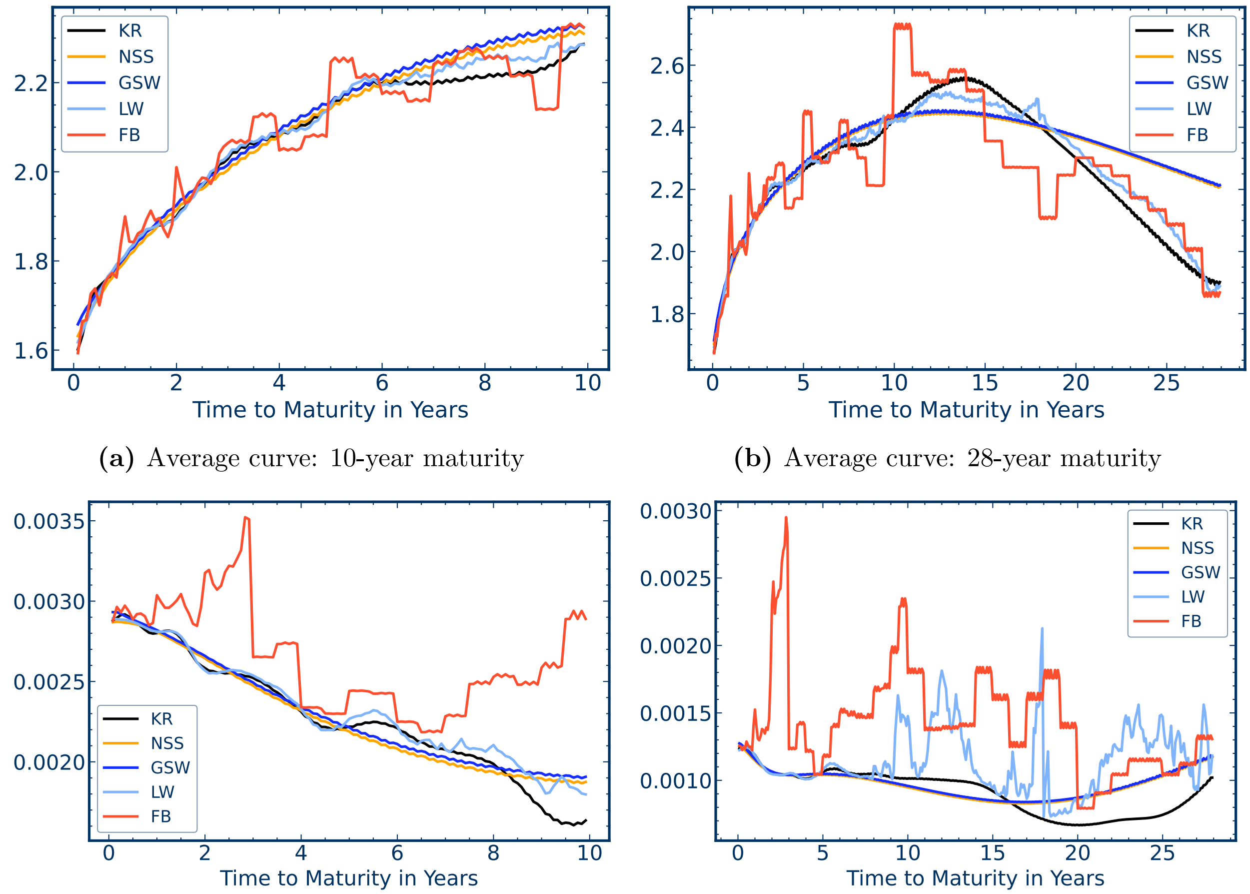
4.1 Bias and Instabilities of Forward Rate Time-Series

We examine forward rate curves over time to understand whether there are systematic biases in the time-series and how rates vary for different estimators. Our analysis focuses on forward rates instead of yields as they isolate the effects for different maturities. Yields are essentially averages of forward rates over the various maturities. Hence, long maturity yields average over the effects on short and long maturity forward rates, while forward rates provide a more direct measure. We obtain the one-month forward rate curve based on the yield estimates of the KR, NSS, GSW, FB and LW methods at daily frequency. Our analysis studies the time-series of the full panel of daily observations that are available for all methods. Hence the panel with maximum maturity of 10 years can take advantage of the longer sample from 1971-11-10 to 2013-12-30, while the the 28-year maturity sample uses the shorter full panel from 1985-11-25 to 2000-01-01.[31]

Imprecise methods with higher out-of-sample errors suffer from a systematic bias. The top panel of Figure 16 shows the average forward rate curve for different estimators. The time-series average shows that there is a systematic difference in forward rates. The KR estimate are on average lower for long maturities. In more detail, the less precise methods GSW and NSS, and LW have an upward bias for the range of maturities between 6 to 10 years and for maturities larger than 18 years. In addition, there is a downward bias for imprecise methods that systematically underestimate forward rates for the range of 11 to 17 years. The functional form for GSW and NSS is distorted for long maturity estimates and deviates substantially from the estimates of the non-parametric estimates. The FB estimates result in an extremely unstable time-series, which is reflected by the kinks and spikes in the average forward rate curves. The systematic bias for the imprecise methods can impact downstream applications and can result in a potential bias for parameter estimates that are based on the GSW, NSS, LW or FB time-series.

[31] We have also conducted the analysis for alternative panels including a full panel with 23 years maximum maturity from 1985-11-25 to 2013-12-30. The results are qualitatively the same as for the two cases presented in the main text and available upon request.

Figure 16: Forward rate curves over time



(a) Average curve: 10-year maturity

(b) Average curve: 28-year maturity

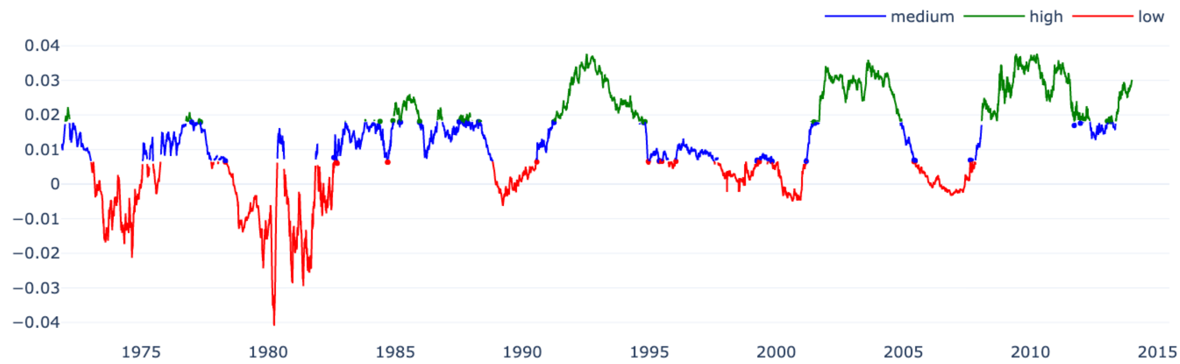
(c) Standard deviation: 10-year maturity

(d) Standard deviation: 28-year maturity

This figure shows the average 1-month forward rate curve and its standard deviation over time for different estimators. We obtain the 1-month forward rate curve based on the yield estimates of the KR, NSS, GSW, FB and LW methods at daily frequency. The top panel shows the average forward rates over time, while the bottom panel displays the standard deviation over time. The 10-year maturity sample uses the full panel of daily observations from 1971-11-10 to 2013-12-30, while the 28-year maturity sample uses the shorter full panel from 1985-11-25 to 2000-01-01. Rates are annualized.

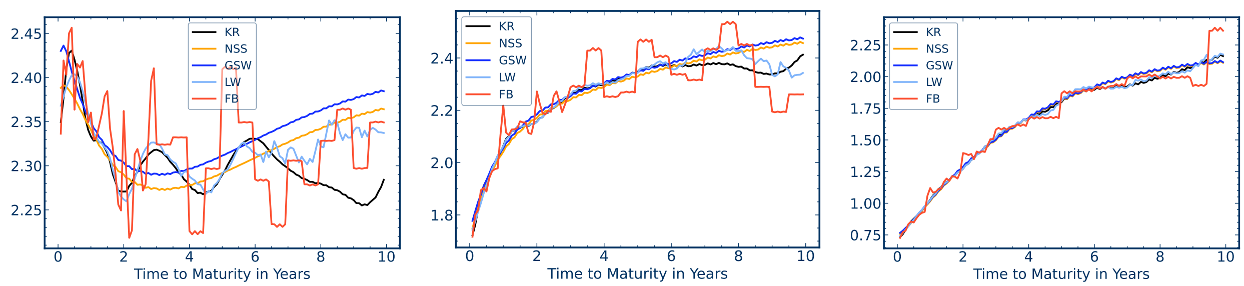
The imprecise methods with higher out-of-sample errors suffer from a systematically higher variation and instability in their estimated time-series. The bottom panel of Figure 16 shows the time-series standard deviation of the forward rates for different estimators. Obviously, we do not assume that forward rate curves or their moving averages are constant over time and we expect variations around their unconditional average. However, spikes in the standard deviation curves for nearby maturities and an excessive variation relative to benchmark estimators are evidence for instabilities in the time-series. The FB time-series stand out by their extreme instability. The spikes in the standard deviation and mean of the time-series indicate outliers in the time-series. The LW estimates result in the second most unstable time-series as measured by the standard deviation. The visible spikes, as seen for example for a maturity of 18 years, occur because for some days the LW approach overfits outliers that significantly distort the forward rate curves. The

Figure 17: Time-series of 10Y-1Y yield spread



The plot shows the time-series of the yield spread between 10 years and 1 year of maturity. The yields are estimated with the KR method. We obtain three yield spread terciles (low, medium, high) based on the quantiles of the full time-series.

Figure 18: Forward rates conditional on yield spread



(a) low yield spread **(b)** medium yield spread **(c)** high yield spread

This figure shows the average 1-month forward rates conditional on yield spread terciles. We obtain the 1-month forward rates up to 10 years of maturity based on the yield estimates of the KR, NSS, GSW, FB and LW methods at daily frequency. The 10-year maturity sample uses the full panel of daily observations from 1971-11-10 to 2013-12-30. From the left to the right, the conditional averages are taken over the time periods of low, medium and high yield spreads respectively. Rates are annualized.

GSW and NSS estimates are very close and show little variation over time. This comes at the cost of the extreme bias as displayed in the distorted shape for the average rates. KR has the most stable time-series. This suggests that KR provides the best trade-off between bias and variance among all methods. The stability and lower variation of the KR time-series can have important implications for downstream applications, such as regressions or GMM estimates that use a panel of interest rates as input. Parameter estimates with the KR time-series are expected to have smaller confidence intervals, and are less distorted by time-series outliers.

The precision of the KR estimates is even more important for a conditional analysis. So far, we have presented results for the unconditional average of forward rate curves. Next, we want to understand how forward rate curves differ depending on economic conditions. A conditional analysis includes regressions of yields and forward rates on a set of covariates. There are obviously many conditioning variables that have been studied in the term structure literature. Here, we focus on a transparent example and consider the yield spread. The yield spread is directly related to macroeconomic conditions and allows us to isolate the effect for time periods with inverted yield curves. Figure 17 shows the times-series of the yield spread between 10 years and 1 year of maturity. The yields are estimated with the KR method. We define the three states of low, medium and high yield spreads based on the terciles of the full yield spread time-series.

Figure 18 shows forward rates conditional on the yield spread. The forward rate curves are averaged over low, medium and high yield spread tercile periods. Our analysis focuses on the panel of up to 10 years of maturity, but the qualitative results are similar for longer maturities. First, the systematic overestimation bias of long maturities for imprecise methods continues to hold in different spread time periods. Second, the relative bias in the curves is even stronger for times of low yield spread, which include episodes of inverted yield curves. Hence, the precision of the KR method in terms of lower bias and variance is even more important for a conditional analysis.

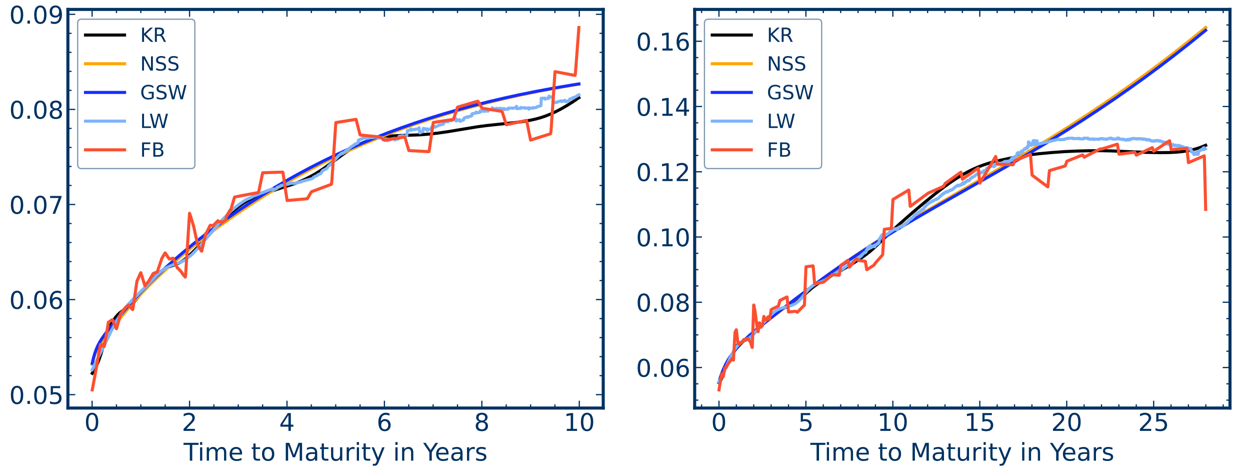
4.2 Investment Implications

The term structure of interest rates directly impacts investment decisions in numerous ways. First, understanding the returns of discount bonds directly maps into investment strategies. Second, factors that explain the term structure are of interest for asset pricing and for earning the term structure risk premia. Third, many trading strategies and derivatives depend on risk-free rates of various maturities. As a first step, we study the returns of discount bonds. We calculate the returns from the estimated discount bond prices of the different methods. Importantly, only KR discount bonds are based on tradable portfolios of Treasury bonds. The discount bond prices of GSW, NSS, FB and LW are "artificial numbers", and do not map directly into investable portfolios of traded assets.

The precision of KR matters for the investment analysis of discount bond returns. Figure 19 shows the average monthly returns of discount bonds. After subtracting the short term risk-free rate, the average return corresponds to the unconditional term structure risk premia for different maturities. The returns mirror the same systematic biases and instabilities that we have already found for forward rates. The returns of FB are extremely unstable. Hence, portfolio strategies, for example mean-variance efficient portfolios, based on FB returns can be heavily distorted by overfitting in-sample outliers. The less precise methods GSW and NSS, and LW systematically overestimate the average returns for 6 to 10 years of maturity and for maturities larger than 18 years. The instability in the price estimates of FB and LW results in more volatile estimated return time-series. We conclude that the estimation of the term structure premium and portfolio strategies on discount bonds can benefit from the precision of KR.

Figure 20 shows the discount bond returns conditional on yield spreads. This is an illustration of portfolio strategies that time the macroeconomic conditions. We plot the average one-day returns of discount bonds conditional on yield spread terciles. We study the longer panel of up to 10 years of maturity, but the results are similar for longer maturities. First, we observe that the return

Figure 19: Average returns of discount bonds

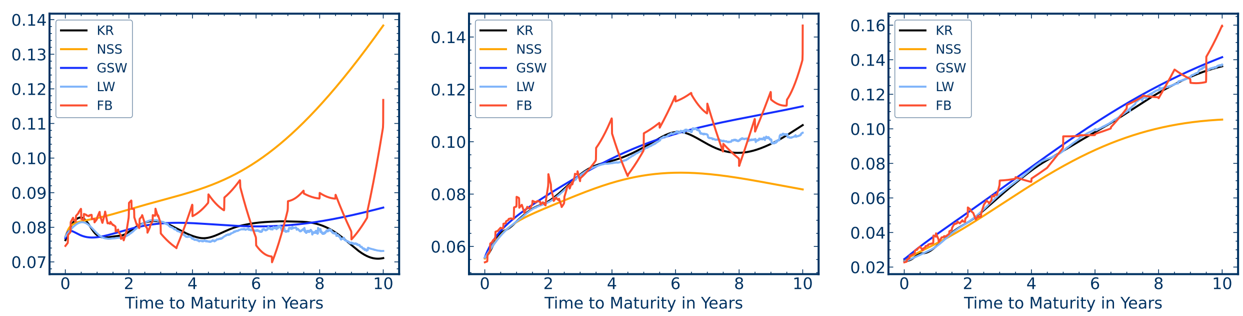


(a) 10-year maturity

(b) 28-year maturity

This figure shows the average monthly returns of discount bonds for different maturities implied by the discount bond price estimates of the KR, NSS, GSW, FB and LW methods. The 10-year maturity sample uses the full panel from 1971-11-10 to 2013-12-30, while the 28-year maturity sample uses the shorter full panel from 1985-11-25 to 2000-01-01. Returns are annualized.

Figure 20: Discount bond returns conditional on yield spread



(a) low yield spread

(b) medium yield spread

(c) high yield spread

This figure shows the average daily discount bond returns conditional on yield spread terciles. We obtain the daily returns up to 10 years of maturity from by the discount bond price estimates of the KR, NSS, GSW, FB and LW methods at daily frequency. The 10-year maturity sample uses the full panel of daily observations from 1971-11-10 to 2013-12-30. From the left to the right, the conditional averages are taken over the time periods of low, medium and high yield spreads respectively. Returns are annualized.

distortion for FB, NSS and GSW is more extreme for times of low and medium yield spreads. These biases would for example affect how fixed income investments would be evaluated during times of inverted yield curves. We conclude that FB, NSS and GWS cannot reliably be used to study the discount bond returns in times of low and medium yield spreads. As seen throughout the paper, NSS and GSW provide essentially the same yield and forward rate curves. However, returns can magnify smaller discrepancies between price time-series as they correspond to percentage changes. The NSS estimates are affected by a few outliers in the time-series, which seems to distort the average return curve in the conditional analysis.[32] LW continues to be upward biased and less stable. As before, KR provides the most stable time-series. We conclude that the precision of KR matters even more for a conditional investment analysis.

The factors that explain the cross-section of discount bond prices depend on the discount curve estimates. The principal component analysis of the previous section illustrates that the differences between KR and the alternative LW, FB, GSW and NSS methods are more pronounced in the fourth and fifth principal components. Filipović, Pelger, and Ye (2022) study the investment implications of term structure factors estimated with a KR approach. More specifically, Filipović, Pelger, and Ye (2022) study the term structure risk premium and the connection between the non-parametric estimate and the bond risk factors. They can link bond risk factors to higher order PCs and show that they carry a substantial risk premium. In particular, they show how the precise estimation of the yield curve can result in profitable and exploitable trading strategies.

4.3 Broader Economic Implications

As we have seen above, a more precise and robust estimation of the yield curve can benefit the downstream finance applications, for which it is used as an input. The broader economic implications depend on the type of applications, and on which methods are compared with each other. We now highlight some areas that are the most likely to be affected by a specific choice of estimator.

Economic studies that are based on aggregated quantities depend less on precise yield estimates. More specifically, the differences between the estimated yields might affect less aggregated averages or if the parameter of interest is robust to noisy yield estimates. A particular example is testing the expectation hypothesis with the Campbell and Shiller (1991) regression. All methods, KR, LW, FB, GSW and NSS, reject the expectation hypothesis and have similar levels of significance.[33]

The economic implications of using non-parametric versus parametric estimators are the most pronounced and visible in many applications. Parametric estimates such as GSW differ the most from the non-parametric KR and LW estimates. We have shown that GSW and NSS estimates have the strongest bias due to the misspecified functional form. The short maturities rates are extremely unstable and have large out-of-sample errors. The higher-order PCA factors based on GSW and NSS panels are distorted. Essentially any empirical asset pricing study requires a risk-free rate, which is usually taken as the 30-day Treasury yield. The GSW short-term rates are not suitable for any of those applications and can lead to non-negligible spurious effects. Cochrane and Piazzesi (2009) and Gürkaynak, Sack, and Wright (2010) made the point that the GSW yields reduce predictability because of their overly simplistic form. The estimates of KR and LW are not identical, but relatively close and, more importantly, deviate in a similar way from the GSW

[32]The time-series of NSS discount bond returns has visible outliers in low yield spread time periods. A closer analysis of these days shows that the fit of the NSS curves is distorted by a small number of outlier bond prices. The GSW estimates of the NSS model are less affected as they apply stricter filters, which seem to remove some of these outlier bond prices. The effect of the distorted NSS estimates is visible in discount bond returns but not forward rates, as relative changes are magnified for discount bonds with low prices.

[33]The results are available upon request.

curves. Therefore, we conjecture that the findings of Liu and Wu (2021) with respect to the work of Cochrane and Piazzesi (2005) and Giglio and Kelly (2017) carry over to our setting. Hence, the choice of yield curve has economic implications in terms of forecasting regressions and excess volatility of long-term bond prices among others.

Within the class of non-parametric estimators, the FB method is the most problematic. The FB panels of yields are the least stable in terms of variation and non-smoothness of curves over time. The returns of FB discount bonds are distorted due to overfitting outliers. We expect regression results with FB rates or returns to depend more on the choice of sample as temporal outliers strongly impact the fitting of FB curves. We also expect the confidence intervals in such regressions to be larger. This sensitivity is particularly relevant in a conditional analysis that studies the relationship between the term structure and other covariates in specific market regimes. However, for very short-maturity interest rates, like the 30-day Treasury yield, the FB estimates perform similar to KR and LW. Hence, in a conventional reduced-form asset pricing application, the short-term risk free rates of KR, LW or FB for obtaining returns in excess of the risk-free rate would give essentially the same results.

The differences between non-parametric methods such as KR and LW are more subtle. LW is on average the closest to KR and is the second most precise method in terms of out-of-sample errors. However, LW has systematic biases in forward rate and return curves, but those are less pronounced than for GSW and NSS. More importantly, the time-series of LW are less stable and more prone to outliers. Hence, applications that require precise time-series of yields for all maturities are the most likely to be affected. This is particularly relevant for a conditional analysis, for example, when studying yields or discount bond returns for all maturities under specific market conditions. A local smoothing estimator like LW is less stable and precise for maturity ranges with sparse or noisy observations. Hence, in particular the maturity ranges around 10 years and larger than 18 years can be more affected by instabilities and overfitting in their time-series. The principal component analysis of the previous section illustrates that the cross-sectional differences between the KR and LW methods appear in the fourth and fifth principal components. Hence, the economic implications can differ for applications that depend on a precise estimation of yield factors. For example, Filipović, Pelger, and Ye (2022) study the term structure risk premium and show the importance of the precise estimation of these higher order term structure factors. Last but not least, another important area is the pricing and hedging of interest rate derivatives, where smaller errors and the effect of outliers can be magnified.

A large part of the term structure literature focuses on the U.S. Treasury market, which is among the most liquid government bond markets. We expect the benefits of our robust and flexible KR estimator, which provides precise discount curve estimates even with sparse and noisy data, to be even larger for government bonds of other countries. For this reason we are going to create a publicly available comprehensive yield data library of KR estimates for the major government bond markets.

In summary, because KR provides uniformly more precise and robust estimates than the benchmarks, there is no argument for using any other existing parametric or non-parametric methods. The economic implications of the choice of yield curve estimate depend very much on the application.

5 Conclusion

The precise and robust estimation of the yield curve is of fundamental importance for economic researchers and practitioners. This paper develops a robust, flexible and easy-to-implement method to estimate the yield curve. Our approach is a machine learning solution that is tailored to the economic problem. It imposes minimal assumptions on the true underlying yield curve, using only the core elements that define the estimation problem. We trade off pricing errors against an economically motivated smoothness reward of the discount curve. This uniquely determines the optimal basis functions that span the discount curve in a reproducing kernel Hilbert space. We show that most existing models for estimating the discount curve are nested within our general framework by imposing additional ad-hoc assumptions. We provide a closed-form solution of our machine learning estimator as a simple kernel ridge regression, which is straightforward to implement.

Our estimates set the new standard for yield curve estimation. We show in an extensive empirical study on U.S. Treasury securities that our method uniformly dominates all parametric and non-parametric benchmarks. Our method achieves smaller out-of-sample yield and pricing errors, while being robust to outliers. We provide a publicly available and regularly updated new benchmark dataset for daily zero-coupon Treasury yields based on our estimates. Our benchmark dataset provides the most precise zero-coupon Treasury yield estimates for all maturity ranges, while being robust to data selection choices.

Besides superior estimates, our novel perspective also provides new insights for spanning the term structure. The non-parametric basis functions that best explain the discount curve on any specific day are closely related to the basis functions that explain the variation in a panel of discount curves. The occurrence of level, slope and curvature type patterns in such a panel is not a coincidence, but a consequence of our finding that these type of basis functions best explain the functional form of the yield curve. So far, the literature has largely separated the two problems of yield curve estimation and explaining the term structure premium for a cross-section of bond returns. Our findings lay the foundation for a new direction that can connect these two problems by unifying term structure asset pricing with non-parametric yield curve estimation.

References

Andersen, L. (2007): "Discount curve construction with tension splines," *Review of Derivatives Research*, 10(3), 227–267.

Bliss, R. R. (1996): "Testing term structure estimation methods," Working Paper 96-12a, Federal Reserve Bank of Atlanta, Atlanta, GA.

Brezis, H. (2011): *Functional analysis, Sobolev spaces and partial differential equations*, Universitext. Springer, New York.

Cochrane, J. H., and M. Piazzesi (2005): "Bond Risk Premia," *American Economic Review*, 95(1), 138–160.

——— (2009): "Decomposing the yield curve," *Working paper*.

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

Cucker, F., and D.-X. Zhou (2007): *Learning theory: an approximation theory viewpoint*, vol. 24 of *Cambridge Monographs on Applied and Computational Mathematics*. Cambridge University Press, Cambridge, With a foreword by Stephen Smale.

Dybvig, P. H., J. E. Ingersoll, and S. A. Ross (1996): "Long Forward and Zero-Coupon Rates Can Never Fall," *The Journal of Business*, 69(1), 1–25.

EIOPA (2020): "Technical documentation of the methodology to derive EIOPA's risk-free interest rate term structures," Discussion paper, European Insurance and Occupational Pensions Authority.

Fama, E. F., and R. R. Bliss (1987): "The Information in Long-Maturity Forward Rates," *The American Economic Review*, 77(4), 680–692.

Filipović, D. (2001): *Consistency problems for Heath-Jarrow-Morton interest rate models*, vol. 1760 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin.

——— (2009): *Term-structure models*, Springer Finance. Springer-Verlag, Berlin, A graduate course.

Filipović, D., M. Larsson, and A. B. Trolle (2017): "Linear-Rational Term Structure Models," *The Journal of Finance*, 72(2), 655–704.

Filipović, D., M. Pelger, and Y. Ye (2022): "Shrinking the Term Structure," *Working paper*.

Filipović, D., and S. Willems (2018): "Exact Smooth Term-Structure Estimation," *SIAM J. Financial Math.*, 9(3), 907–929.

Fisher, M., D. Nychka, and D. Zervos (1995): "Fitting the term structure of interest rates with smoothing splines," Finance and Economics Discussion Series 95-1, Board of Governors of the Federal Reserve System (U.S.).

Giglio, S., and B. Kelly (2017): "Excess Volatility: Beyond Discount Rates*," *The Quarterly Journal of Economics*, 133(1), 71–127.

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

Gürkaynak, R. S., B. Sack, and J. H. Wright (2010): "The TIPS Yield Curve and Inflation Compensation," *American Economic Journal: Macroeconomics*, 2(1), 70–92.

Hastie, T., R. Tibshirani, and J. Friedman (2009): *The elements of statistical learning*, Springer Series in Statistics. Springer, New York, second edn., Data mining, inference, and prediction.

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

Hull, J., and A. White (1990): "Pricing Interest-Rate-Derivative Securities," *The Review of Financial Studies*, 3(4), 573–592.

Jacod, J., and A. N. Shiryaev (2003): *Limit theorems for stochastic processes*, vol. 288 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, second edn.

Jeffrey, A., O. Linton, and T. Nguyen (2006): "Flexible term structure estimation: Which method is preferred?," *Metrika*, 63(1), 99 – 122.

Lagerås, A., and M. Lindholm (2016): "Issues with the Smith–Wilson method," *Insurance: Mathematics and Economics*, 71, 93–102.

Linton, O., E. Mammen, J. P. Nielsen, and C. Tanggaard (2001): "Yield curve estimation by kernel smoothing methods," *Journal of Econometrics*, 105(1), 185–223, Forecasting and empirical methods in finance and macroeconomics.

Liu, Y., and J. C. Wu (2021): "Reconstructing the yield curve," *Journal of Financial Economics*, 142(3), 1395–1425.

McCulloch, J. H. (1971): "Measuring the Term Structure of Interest Rates," *The Journal of Business*, 44(1), 19–31.

Nelson, C. R., and A. F. Siegel (1987): "Parsimonious Modeling of Yield Curves," *Journal of Business*, 60(4), 473–489.

Paulsen, V. I., and M. Raghupathi (2016): *An Introduction to the Theory of Reproducing Kernel Hilbert Spaces*, Cambridge Studies in Advanced Mathematics. Cambridge University Press.

Rasmussen, C. E., and C. K. I. Williams (2006): *Gaussian Processes for Machine Learning*, Adaptive Computation and Machine Learning. MIT Press.

Schölkopf, B., and A. J. Smola (2018): *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. The MIT Press.

Smith, A., and T. Wilson (2001): "Fitting yield curves with long term constraints," *Working paper.*

Svensson, L. E. (1994): "Estimating and Interpreting Forward Interest Rates: Sweden 1992 - 1994," Discussion paper, National Bureau of Economic Research, Cambridge, MA.

Tanggaard, C. (1997): "Nonparametric Smoothing of Yield Curves," *Review of Quantitative Finance and Accounting*, 9(3), 251–267.

Teschl, G. (2012): *Ordinary differential equations and dynamical systems*, vol. 140 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI.

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

Vasicek, O. A., and H. G. Fong (1982): "Term Structure Modeling Using Exponential Splines," *The Journal of Finance*, 37(2), 339–348.

Viehmann, T. (2019): "Variants of the Smith-Wilson method with a view towards applications," *Working paper.*

A Theory

In this appendix we provide the theoretical background and proofs of our main results.

A.1 Proof of Theorem 1

As $g(x)$ is finite by assumption, we can define the x-forward measure $\mathbb{Q}^x \sim \mathbb{Q}$, given by its Radon–Nikodym derivative $\frac{d\mathbb{Q}^x}{d\mathbb{Q}} = g(x)^{-1} \mathrm{e}^{-\int_0^x r_u\,du}$. Hence the integrand in (3) is equal to $g(t)\mathbb{E}_{\mathbb{Q}^t}\left[r_t^2 + |\mu_t|\right]$. Jensen's inequality now implies that $\int_0^x \mathbb{E}_{\mathbb{Q}}\left[\mathrm{e}^{-\int_0^t r_u\,du}\,|r_t|\right] dt < \infty$, for any finite x.

Elementary stochastic calculus shows

$$\mathrm{e}^{-\int_0^x r_u du} = 1 - \int_0^x \mathrm{e}^{-\int_0^t r_u du}\, r_t\, dt. \tag{20}$$

In view of the above, we can change the order of integration and obtain

$$g(x) = 1 - \int_0^x \mathbb{E}_{\mathbb{Q}}\left[\mathrm{e}^{-\int_0^t r_u\,du}\, r_t\right] dt. \tag{21}$$

Decomposition (21) reveals that the discount bond price equals its face value one minus the present value of the continuous cash flow of short (overnight) rates. It also shows that $g(x)$ is differentiable in x with derivative $g'(x)$ given in (6), which can also be written as $g'(x) = -g(x)\mathbb{E}_{\mathbb{Q}^x}[r_x]$. This expression is in line with the well-known fact that the instantaneous forward rate $f(x) = -\frac{g'(x)}{g(x)}$ equals the expected value of the future short rate r_x under $\mathbb{Q}^x$.

Further stochastic calculus shows that the inner integrand in (21) can be expanded as

$$\mathrm{e}^{-\int_0^t r_u\,du}\, r_t = r_0 - \int_0^t \mathrm{e}^{-\int_0^s r_u du}\,(r_s^2 - \mu_s)\, ds + \int_0^t \mathrm{e}^{-\int_0^s r_u du}\, dM_s. \tag{22}$$

In view of assumptions (3) and (4) we can change the order of integration in (21), and the last term in (22) is a martingale, see (Jacod and Shiryaev, 2003, Theorem I.4.40). We obtain

$$g(x) = 1 - r_0\, x + \int_0^x \int_0^t \mathbb{E}_{\mathbb{Q}}\left[\mathrm{e}^{-\int_0^s r_u du}\,(r_s^2 - \mu_s)\right] ds\, dt. \tag{23}$$

Decomposition (23) shows that $g(x)$ is twice weakly differentiable in x with second derivative given in (7). This completes the proof of Theorem 1.

A.2 General Functional Analytic Perspective

We recap a fundamental notion in statistical machine learning, namely that of a reproducing kernel Hilbert space. For more background and applications we refer the reader to, e.g., Rasmussen and Williams (2006); Cucker and Zhou (2007); Hastie, Tibshirani, and Friedman (2009); Paulsen and Raghupathi (2016); Schölkopf and Smola (2018). Let E be an arbitrary set and $\mathcal{H}$ a Hilbert space of functions $h : E \to \mathbb{R}$. $\mathcal{H}$ is called a *reproducing kernel Hilbert space (RKHS)* if, for any $x \in E$, there exists a function $k_x \in \mathcal{H}$ such that the scalar product $\langle h, k_x \rangle_{\mathcal{H}} = h(x)$ acts as evaluation at x for all $h \in \mathcal{H}$. The function $k : E \times E \to \mathbb{R}$ induced by $k(x, y) = \langle k_x, k_y \rangle_{\mathcal{H}} = k_y(x)$ is called the *reproducing kernel* of $\mathcal{H}$. It has the property that for any finite selection of points $x_1, \ldots, x_n \in E$ the $n \times n$ matrix with elements $k(x_i, x_j)$ is symmetric and positive semi-definite. Thanks to the powerful property called the representer theorem, many kernel-based machine learning problems boil down to finite-dimensional standard convex optimization. Our following results are a variant thereof.

For our purpose of learning the discount curve g, we work on the right-open interval $E = [0, \tau)$ for a time horizon $\tau \leq \infty$. We also consider extensions to the closure $E = [0, \infty]$, see Section A.7 below. As $g(0) = 1$, it is convenient to model the discount curve as

$$g = p + h \tag{24}$$

for some exogenous prior curve $p : E \to \mathbb{R}$ with $p(0) = 1$, and a hypothesis function h optimally chosen from a RKHS $\mathcal{H}$ consisting of functions $h : E \to \mathbb{R}$ with zero initial value $h(0) = 0$.[34] We denote by $k : E \times E \to \mathbb{R}$ the reproducing kernel of $\mathcal{H}$, so that $k(0, 0) = 0$. A feasible example of the prior curve is $p(x) = \mathrm{e}^{-rx}$ for some constant prior interest rate r. A special case is the constant $p = 1$, which actually is our choice in the main text. We derive the main results in this section for the present general setup so that researchers can draw on it while using their preferred hypothesis spaces.

As in Section 2.1, we let $P = (P_1, \ldots, P_M)^\top$ denote the observed prices of M fixed income securities with cash flows on dates $0 < x_1 < \cdots < x_N \in E$ summarized in the $M \times N$ cash flow matrix C. We write $C_i = (C_{i1}, \ldots, C_{iN})$ for the i-th row of C. We also write $\boldsymbol{x} = (x_1, \ldots, x_N)^\top$ and, accordingly, $f(\boldsymbol{x}) = (f(x_1), \ldots, f(x_N))^\top$ for any function f. The fundamental value of security i is thus $P_i^g = C_i(p(\boldsymbol{x}) + h(\boldsymbol{x}))$. We then learn the ground truth discount curve (24) from the data as solution of the optimization problem

$$\min_{h \in \mathcal{H}} \left\{ \sum_{i=1}^{M} \omega_i (P_i - C_i(p(\boldsymbol{x}) + h(\boldsymbol{x})))^2 + \lambda \|h\|_{\mathcal{H}}^2 \right\} \tag{25}$$

for some exogenous weights $0 < \omega_i \leq \infty$ and a regularization parameter $\lambda > 0$. This nests and generalizes (9) in three ways. First, we model the discount curve g by (24). Second, we measure regularity of g by the $\mathcal{H}$-norm $\|h\|_{\mathcal{H}}$. And, third, we now explicitly allow for infinite weights, $\omega_i = \infty$, which corresponds to an exact pricing condition, $P_i = C_i(p(\boldsymbol{x}) + h(\boldsymbol{x}))$. Accordingly, we define the index sets $\mathcal{I}_1 = \{i \mid \omega_i = \infty\}$ and $\mathcal{I}_0 = \{1, \ldots, M\} \setminus \mathcal{I}_1$. Problem (25) is then to be read as the constrained optimization problem

$$\begin{aligned} &\min_{h \in \mathcal{H}} \left\{ \sum_{i \in \mathcal{I}_0} \omega_i (P_i - C_i(p(\boldsymbol{x}) + h(\boldsymbol{x})))^2 + \lambda \|h\|_{\mathcal{H}}^2 \right\} \\ &\text{subject to } P_i - C_i(p(\boldsymbol{x}) + h(\boldsymbol{x})) = 0, \quad i \in \mathcal{I}_1. \end{aligned} \tag{26}$$

This problem is well-posed and admits a closed-form solution, as the following kernel representer

[34] Model (24) is in line with the additive structure of the linear-rational term structure models, see (Filipović, Larsson, and Trolle, 2017, Eqn. (21)).

theorem shows. Thereto, we define the positive semi-definite $N \times N$ kernel matrix $\boldsymbol{K} = k(\boldsymbol{x}, \boldsymbol{x}^\top)$, that is, with components $\boldsymbol{K}_{ij} = k(x_i, x_j)$. Moreover, for any index set $\mathcal{I}$, of size $|\mathcal{I}|$, we denote by $C_\mathcal{I}$ the $|\mathcal{I}| \times N$-matrix consisting of the rows C_i, $i \in \mathcal{I}$.

Theorem A.1 (General kernel ridge regression solution)
Assume that either $\mathcal{I}_1 = \emptyset$ or $C_{\mathcal{I}_1}\boldsymbol{K}C_{\mathcal{I}_1}^\top$ is invertible. Then the $M \times M$-matrix $C\boldsymbol{K}C^\top + \Lambda$ is invertible, where $\Lambda = \operatorname{diag}(\lambda/\omega_1, \ldots, \lambda/\omega_M)$ and we define $\lambda/\infty = 0$. Moreover, there exists a unique solution $h = \hat{h} \in \mathcal{H}$ of (25), and it is given by

$$\hat{h} = k(\cdot, \boldsymbol{x})^\top \beta \tag{27}$$

where β is given by

$$\beta = C^\top (C\boldsymbol{K}C^\top + \Lambda)^{-1}(P - Cp(\boldsymbol{x})). \tag{28}$$

The corresponding discount curve is $\hat{g} = p + \hat{h}$.

Proof. Define the linear sampling operator $S : \mathcal{H} \to \mathbb{R}^N$ by $Sh = h(\boldsymbol{x})$ for $h \in \mathcal{H}$. Its adjoint is given by $S^*\beta = k(\cdot, \boldsymbol{x})^\top \beta$ for $\beta \in \mathbb{R}^N$, and $SS^* : \mathbb{R}^N \to \mathbb{R}^N$ has matrix representation $\boldsymbol{K}$. Hence we can rewrite (26) in operator form

$$\begin{aligned} &\min_{h \in \mathcal{H}} \left\{ \sum_{i \in \mathcal{I}_0} \omega_i (P_i - C_i p(\boldsymbol{x}) - C_i S h)^2 + \lambda \|h\|_\mathcal{H}^2 \right\} \\ &\text{subject to } P_i - C_i p(\boldsymbol{x}) - C_i S h = 0, \quad i \in \mathcal{I}_1. \end{aligned} \tag{29}$$

Existence for (29) follows from well-known results on the existence of minima of convex functions on Hilbert spaces, see, e.g., (Brezis, 2011, Corollary 3.23). Indeed, by the presence of the penalty term $\lambda\|h\|_\mathcal{H}^2$, the objective function, say $\Phi(h)$, in (29) is continuous, strictly convex in h, and coercive, $\Phi(h) \to \infty$ as $\|h\|_\mathcal{H} \to \infty$. Uniqueness follows from the strict convexity of Φ.[35] By the same token, it follows that the solution of (29) must lie in the orthogonal complement of the null space of CS. That is, $h = S^* C^\top q$, for some $q \in \mathbb{R}^M$. Plugging this in (29) leads to the convex optimization problem in $\mathbb{R}^M$

$$\begin{aligned} &\min_{q \in \mathbb{R}^M} \left\{ \sum_{i \in \mathcal{I}_0} \omega_i (P_i - C_i p(\boldsymbol{x}) - C_i \boldsymbol{K} C^\top q)^2 + \lambda q^\top C\boldsymbol{K}C^\top q \right\} \\ &\text{subject to } P_i - C_i p(\boldsymbol{x}) - C_i \boldsymbol{K} C^\top q = 0, \quad i \in \mathcal{I}_1. \end{aligned} \tag{30}$$

The Lagrangian $L : \mathbb{R}^M \times \mathbb{R}^{|\mathcal{I}_1|} \to \mathbb{R}$ of (30) is given by

$$L(q, \nu) = \sum_{i \in \mathcal{I}_0} \omega_i (P_i - C_i p(\boldsymbol{x}) - C_i \boldsymbol{K} C^\top q)^2 + \lambda q^\top C\boldsymbol{K}C^\top q + 2 \sum_{j \in \mathcal{I}_1} \nu_j (P_j - C_j p(\boldsymbol{x}) - C_j \boldsymbol{K} C^\top q).$$

[35] Note that while the solution to (29) is unique in $\mathcal{H}$, a solution to (30) may not be unique in $\mathbb{R}^M$. Indeed, we do not assume that $\ker S^* C^\top$ is zero.

The first order conditions, $D_q L = 0$ and $D_\nu L = 0$, read

$$\sum_{i\in\mathcal{I}_0}\Big(\omega_i(P_i - C_i p(\boldsymbol{x}) - C_i \boldsymbol{K} C^\top q) - \lambda q_i\Big)\, C\boldsymbol{K}C_i^\top + \sum_{j\in\mathcal{I}_1}(\nu_j - \lambda q_j)\, C\boldsymbol{K}C_j^\top = 0, \tag{31}$$

$$P_{\mathcal{I}_1} - C_{\mathcal{I}_1} p(\boldsymbol{x}) - C_{\mathcal{I}_1}\boldsymbol{K}C^\top q = 0. \tag{32}$$

A particular solution (q,ν) to (31)–(32) is given by setting $\nu = \lambda q_{\mathcal{I}_1}$ where q solves

$$(C\boldsymbol{K}C^\top + \Lambda)q = P - Cp(\boldsymbol{x}). \tag{33}$$

Indeed, (33) admits a unique solution q, as $C\boldsymbol{K}C^\top + \Lambda$ is invertible by assumption. This completes the proof. $\square$

A.2.1 Proof of Theorem 2

The first part of Theorem 2 follows from Theorem A.1, for $\tau = \infty$ and the constant prior curve $p = 1$ in (24), noting that

$$\mathcal{G}_{\alpha,\delta} = \{g = 1 + h \mid h \in \mathcal{H}_{\alpha,\delta}\} \tag{34}$$

for the RKHS $\mathcal{H} = \mathcal{H}_{\alpha,\delta}$ defined in Section A.4 below. The expressions for the reproducing kernels, and the last statement, follow accordingly from Lemmas 6 and 7. This completes the proof of Theorem 2.

A.3 A Workable Hypothesis Space

We propose a workable hypothesis space that comes with minimal and reasonable structural assumptions. Specifically, we fix a time horizon $\tau \le \infty$. We study in detail the RKHS $\mathcal{H} = \mathcal{H}_{w,\delta}$ consisting of differentiable functions $h : [0,\tau) \to \mathbb{R}$ of the form $h(x) = \int_0^x h'(t)\,dt$ with absolutely continuous derivatives, $h'(x) = h'(0) + \int_0^x h''(t)\,dt$, for locally integrable h'', satisfying the boundary condition

$$\lim_{x\to\tau} h'(x) = 0, \tag{35}$$

and with finite norm given by

$$\|h\|_{w,\delta} = \left(\int_0^\tau \left(\delta h'(x)^2 + (1-\delta)h''(x)^2\right) w(x)\,dx\right)^{\frac{1}{2}}, \tag{36}$$

for some measurable weight function $w : [0,\tau) \to [1,\infty)$ and shape parameter $\delta \in (0,1)$. The RKHS $\mathcal{H}_{w,\delta}$ is a weighted Sobolev type space.[36] We extend the setup to include the boundary cases $\delta \in \{0,1\}$, as follows.

[36] In particular, $h \in \mathcal{H}_{w,\delta}$ implies that $h' \in H^1(I)$ for the standard Sobolev space $H^1(I)$ on $I = (0,\tau)$ in (Brezis, 2011, Chapter 8), for any $\delta \in (0,1)$. Note that h' and h'' in (36) are representatives of their equivalence classes with respect to dx-a.s. equality.

(i) For $\delta = 0$, we assume in addition to the above that

$$C_{w,0} = \int_0^\tau w(x)^{-1}\,dx < \infty. \tag{37}$$

(ii) For $\delta = 1$, we only assume that functions h in $\mathcal{H}_{w,1}$ are absolutely continuous of the form $h(x) = \int_0^x h'(t)\,dt$, for locally integrable h', and with finite norm (36).

Remark 1

The theory in this section is developed under minimal assumptions on the weight function w and time horizon $\tau \le \infty$. In particular, some assumptions are only effective for $\tau = \infty$, while they hold trivially if $\tau < \infty$, such as (37) *or* (43) *below because $w \ge 1$. Throughout, the following examples serve as illustration:*

(i) *Exponential weight function $w(x) = \mathrm{e}^{\alpha x}$ for some $\alpha \ge 0$. Conditions* (37)*, and* (39) *and* (43) *below are satisfied if $\alpha > 0$ or $\tau < \infty$. This is the choice in the main text.*

(ii) *Power weight function $w(x) = (1+x)^\alpha$ for some $\alpha \ge 0$. Condition* (37) *is equivalent to $\alpha > 1$ or $\tau < \infty$. The stronger condition* (43) *is equivalent to $\alpha > 3$ or $\tau < \infty$, whereas condition* (39) *is never satisfied for $\tau = \infty$.*

For $\tau = \infty$, the boundary condition (35) arises naturally from representing the discount curve $g(x) = \exp(-\int_0^x f(t)\,dt)$ in terms of the forward curve f. With prior curve $p = 1$, we have $h'(x) = g'(x) = -f(x)g(x)$, and thus (35) is tantamount to $\lim_{x\to\infty} g(x) = 0$ and bounded forward rates, $\limsup_{x\to\infty} |f(x)| < \infty$. This obviously generalizes to any differentiable prior curve p with $\lim_{x\to\infty} p'(x) = 0$. For $\tau < \infty$, the economic interpretation of (35) is less obvious. We provide an alternative to (35) in Section A.5 below.

For $\delta \in (0,1)$ and $\tau = \infty$, condition (35) holds a fortiori for $h \in \mathcal{H}_{w,\delta}$ due to the finiteness of the integral (36). This follows similarly as in (Brezis, 2011, Corollary 8.9). For $\delta = 0$, without assuming the boundary condition (35), the right hand side of (36) would not define a bona fide norm, as it is zero for the linear function $h(x) = x$.

Remark 2

Assumption (37) *implies, by the Cauchy–Schwarz inequality, that functions $h \in \mathcal{H}_{w,\delta}$ satisfy*

$$\int_0^\tau |h''(x)|\,dx = \int_0^\tau w(x)^{-\frac{1}{2}} |h''(x)| w(x)^{\frac{1}{2}}\,dx \le C_{w,0}^{\frac{1}{2}} (1-\delta)^{-\frac{1}{2}} \|h\|_{w,\delta} < \infty, \tag{38}$$

for all $\delta \in [0,1)$.[37] *As a consequence, $h'(x)$ converges to a finite limit for $x \to \tau$. On the other hand, assumption* (37) *cannot be relaxed for $\delta = 0$ and $\tau = \infty$. This can be seen by the counter-example of a constant weight function $w = 1$ and h with $h''(x) = (1+x)^{-\frac{2}{3}}$, which satisfies $\|h\|_{w,0} < \infty$ but $h'(x) = h'(0) + \frac{1}{3}(1+x)^{\frac{1}{3}}$ is not bounded. The latter contradicts condition* (35).

[37] In a similar vein, it can be shown that $\int_0^\tau (1+x) w(x)^{-1}\,dx < \infty$ implies that $\int_0^\tau h'(x)^2\,dx < \infty$, and thus $h' \in H^1(I)$, also for $h \in \mathcal{H}_{w,0}$. See also footnote 36.

The following lemma relates the function sets $\mathcal{H}_{w,\delta}$ and the norms (36) for varying w and δ. For fixed weight function w, varying the shape parameter δ does not affect $\mathcal{H}_{w,\delta}$ as a set, and the norms (36) are equivalent, except for the boundary case $\delta = 1$, and for $\delta = 0$ under some technical condition, see (39).

Lemma 1

The following hold:

(i) $\mathcal{H}_{w,\delta} \subseteq \mathcal{H}_{\tilde{w},\delta}$ *and* $\|h\|_{\tilde{w},\delta} \le \|h\|_{w,\delta}$, *for all* $\delta \in [0,1]$, *if* $\tilde{w} \le w$

(ii) $\mathcal{H}_{w,\delta} = \mathcal{H}_{w,\tilde{\delta}}$ *and* $\|h\|_{w,\tilde{\delta}} \le \max\{\frac{\tilde{\delta}}{\delta}, \frac{1-\tilde{\delta}}{1-\delta}\}^{\frac{1}{2}} \|h\|_{w,\delta}$, *for all* $\delta, \tilde{\delta} \in (0,1)$

(iii) $\mathcal{H}_{w,\delta} \subset \mathcal{H}_{w,1}$ *and* $\|h\|_{w,1} \le \delta^{-\frac{1}{2}} \|h\|_{w,\delta}$, *for all* $\delta \in (0,1)$

(iv) $\mathcal{H}_{w,\delta} \subseteq \mathcal{H}_{w,0}$ *and* $\|h\|_{w,0} \le (1-\delta)^{-\frac{1}{2}} \|h\|_{w,\delta}$, *for all* $\delta \in [0,1)$

(v) $\mathcal{H}_{w,0} \subseteq \mathcal{H}_{w,\delta}$ *and* $\|h\|_{w,\delta} \le (\delta C_{w,1} C_{w,2} + 1 - \delta)^{\frac{1}{2}} \|h\|_{w,0}$, *for all* $\delta \in [0,1]$, *if*

$$\begin{aligned} C_{w,1} &= \sup_{x \in [0,\tau)} w(x)^{\frac{1}{2}} \int_x^\tau w(t)^{-\frac{1}{2}}\, dt < \infty, \quad \text{and} \\ C_{w,2} &= \sup_{x \in [0,\tau)} w(x)^{-\frac{1}{2}} \int_0^x w(t)^{\frac{1}{2}}\, dt < \infty. \end{aligned} \tag{39}$$

Proof. (i)–(iv) follow directly from the definition of $\mathcal{H}_{w,\delta}$ and the norm (36).

Now assume that (39) holds. Let $h \in \mathcal{H}_{w,0}$, so that $h'(x) = -\int_x^\tau h''(t)\, dt$, by (35) and (38). Then

$$\begin{aligned} h'(x)^2 &= \left(\int_x^\tau w(t)^{-\frac{1}{4}} w(t)^{\frac{1}{4}} h''(t)\, dt \right)^2 \le \left(\int_x^\tau w(t)^{-\frac{1}{2}}\, dt \right) \left(\int_x^\tau w(t)^{\frac{1}{2}} h''(t)^2\, dt \right) \\ &\le C_{w,1} w(x)^{-\frac{1}{2}} \int_x^\tau w(t)^{\frac{1}{2}} h''(t)^2\, dt. \end{aligned}$$

Hence

$$\begin{aligned} \int_0^\tau h'(x)^2 w(x)\, dx &\le C_{w,1} \int_0^\tau \int_x^\tau w(x)^{\frac{1}{2}} w(t)^{\frac{1}{2}} h''(t)^2\, dt\, dx \\ &= C_{w,1} \int_0^\tau \left(\int_0^t w(x)^{\frac{1}{2}}\, dx \right) w(t)^{-\frac{1}{2}} h''(t)^2 w(t)\, dt \\ &\le C_{w,1} C_{w,2} \|h\|_{w,0}^2, \end{aligned}$$

and thus $h \in \mathcal{H}_{w,\delta}$, for any $\delta \in [0,1]$, which proves (v). □

Remark 3

Lemma 1(ii) *states that the sets* $\mathcal{H}_{w,\delta}$ *are identical, and the norms* $\|\cdot\|_{w,\delta}$ *equivalent, for a fixed* w *and varying* $\delta \in (0,1)$. *Lemma 1*(iv) *and* (v) *imply that this extends to* $\delta \in [0,1)$ *if* (39) *holds. Condition* (39) *holds for any continuous weight function* w *if* $\tau < \infty$. *If* $\tau = \infty$ *then* (39) *is* *satisfied for the exponential weight function* $w(x) = \mathrm{e}^{\alpha x}$ *for any* $\alpha > 0$*. Indeed, in this case,* $C_{w,1} = C_{w,2} = \frac{2}{\alpha}$*. Condition* (39) *cannot be relaxed if* $\tau = \infty$ *in general, as seen by the counter-example* $w(x) = (1+x)^\alpha$*, which satisfies* (37) *but not* (39) *for any* $\alpha > 1$*, and* $h \in \mathcal{H}_{w,0}$ *given by* $h'(x) = (1+x)^{-\frac{1+\alpha}{2}}$*, for which we obtain* $\|h\|_{w,\delta} = \infty$ *and thus* $h \notin \mathcal{H}_{w,\delta}$ *for any* $\delta \in (0,1]$*.*

We can characterize the growth rate of functions in $\mathcal{H}_{w,\delta}$ as follows.

Lemma 2
Any function $h \in \mathcal{H}_{w,\delta}$ *is point-wise dominated by*

$$|h(x)| \le \begin{cases} \delta^{-\frac{1}{2}} \|h\|_{w,\delta}\, x^{\frac{1}{2}}, & \text{if } \delta \in (0,1], \\ C_{w,0}^{\frac{1}{2}} \|h\|_{w,0}\, x, & \text{if } \delta = 0. \end{cases}$$

Proof. Assume first that $\delta \in (0,1]$. Using the Cauchy–Schwarz inequality and that $w \ge 1$, we obtain $|h(x)| \le \int_0^x |h'(t)| \sqrt{w(t)} \cdot 1\, dt \le \delta^{-\frac{1}{2}} \|h\|_{w,\delta}\, x^{\frac{1}{2}}$, as claimed. If $\delta = 0$, then (38) implies that $|h'(x)| \le C_{w,0}^{\frac{1}{2}} \|h\|_{w,0}$ for all x, which proves the claim. □

Lemma 2 implies that the point-wise evaluation $h \mapsto h(x)$ acts as a bounded linear functional on $\mathcal{H}_{w,\delta}$, which confirms that $\mathcal{H}_{w,\delta}$ is a RKHS. Its reproducing kernel can be determined as follows.

Lemma 3
Assume that, for any fixed $y \in [0,\tau)$*, there exists a solution* ϕ *of the linear differential equation*

$$\delta \phi w - (1-\delta)(\phi' w)' = 1_{[0,y]}, \tag{40}$$

along with the boundary condition $\phi'(0) = 0$ *if* $\delta \in [0,1)$*, and such that* $\psi \in \mathcal{H}_{w,\delta}$ *for* $\psi(x) = \int_0^x \phi(t)\, dt$*. Then the reproducing kernel* k *of* $\mathcal{H}_{w,\delta}$ *satisfies*

$$k(\cdot, y) = \psi. \tag{41}$$

In particular, for the boundary cases $\delta \in \{0,1\}$*, we have*

$$k(x,y) = \begin{cases} \int_0^\tau (t \wedge x)(t \wedge y) w(t)^{-1}\, dt, & \text{if } \delta = 0, \\ \int_0^{x \wedge y} w(t)^{-1}\, dt, & \text{if } \delta = 1. \end{cases} \tag{42}$$

Proof. Fix $y \in [0,\tau)$, and let $h \in \mathcal{H}_{w,\delta}$. If $\delta \in [0,1)$ and $\tau < \infty$, we have $h'(\tau) = 0$ by assumption (35). If $\tau = \infty$, we assume that $h'(x) = 0$ for $x > n$ for some finite n. In either case, we obtain, via integration by parts of the second term if $\delta \in [0,1)$,

$$\begin{aligned} \langle \psi, h \rangle_{w,\delta} &= \int_0^\tau (\delta \psi'(x) h'(x) + (1-\delta)\psi''(x) h''(x)) w(x)\, dx \\ &= \int_0^\tau (\delta \psi'(x) w(x) - (1-\delta)(\psi'' w)'(x)) h'(x)\, dx = \int_0^\tau 1_{[0,y]}(x) h'(x)\, dx = h(y), \end{aligned}$$

where we used that $\delta\psi' w - (1-\delta)(\psi'' w)' = 1_{[0,y]}$ in view of (40). By Lemma 4, we conclude that $\langle \psi, h\rangle_{w,\delta} = h(y)$ for all $h \in \mathcal{H}_{w,\delta}$, which is the reproducing kernel property and proves (41).

For the last statement, it follows by inspection that $\phi(t) = \int_t^\tau (s\wedge y) w(s)^{-1}\, ds$ and $\phi = 1_{[0,y]} w^{-1}$ satisfy the assumptions for $\delta = 0$ and $\delta = 1$, respectively. Plugging these in (41), and changing the order of integration for $\delta = 0$, proves (42). □

The following lemma is in the spirit of (Brezis, 2011, Theorem 8.7), of independent interest, and used in the proof of Lemma 3.

Lemma 4
Assume $\tau = \infty$. Then for any $h \in \mathcal{H}_{w,\delta}$ there exists a sequence of functions $h_n \in \mathcal{H}_{w,\delta}$ such that $h_n'(x) = 0$ for $x > n$ and $h_n \to h$ in $\mathcal{H}_{w,\delta}$, as $n \to \infty$.

Proof. Let ζ be a smooth function on $\mathbb{R}$ such that $0 \le \zeta(t) \le 1$, $\zeta(t) = 1$ for $t \le 1/2$, and $\zeta(t) = 0$ for $t > 1$. Assume first that $\delta \in (0,1]$. Let $h \in \mathcal{H}_{w,\delta}$ and define $h_n(x) = \int_0^x h'(t)\zeta(t/n)\, dt$. Hence $h_n'(x) = h'(x)\zeta(x/n)$, which is zero for $x > n$. If h' is differentiable, we further have $h_n''(x) = h''(x)\zeta(x/n) + h'(x)\frac{1}{n}\zeta'(x/n)$. As ζ and ζ' are bounded by some constant C_ζ, it is easy to see that $h_n \in \mathcal{H}_{w,\delta}$. Further, we have

$$\begin{aligned}\|h-h_n\|_{w,\delta}^2 &= \int_0^\infty \left(\delta h'(x)^2(1-\zeta(x/n))^2 + (1-\delta)(h''(x)(1-\zeta(x/n)) - h'(x)\frac{1}{n}\zeta'(x/n))^2\right) w(x)\, dx \\ &\le 2\int_{n/2}^\infty \left(\delta h'(x)^2 + (1-\delta)h''(x)^2\right) w(x)\, dx + \frac{2(1-\delta)C_\zeta^2}{n^2}\int_0^\infty h'(x)^2 w(x)\, dx.\end{aligned}$$

Because the last factor $\int_0^\infty h'(x)^2 w(x)\, dx \le \frac{1}{\delta}\|h\|_{w,\delta}^2$, we conclude that $\|h-h_n\|_{w,\delta} \to 0$, as $n \to \infty$.

Now assume that $\delta = 0$. Let $h \in \mathcal{H}_{w,0}$ and this time define $h_n \in \mathcal{H}_{w,0}$ by $h_n'(x) = -\int_x^\infty h''(t)\zeta(t/n)\, dt$. Then $h_n''(x) = h''(x)\zeta(x/n)$ and $h_n'(x)$ are zero for $x > n$. We further have

$$\|h-h_n\|_{w,0}^2 = \int_0^\infty h''(x)^2(1-\zeta(x/n))^2 w(x)\, dx \le \int_{n/2}^\infty h''(x)^2 w(x)\, dx,$$

which converges to zero, as $n \to \infty$. □

We end this section with a lemma that gives conditions on w such that the functions in $\mathcal{H}_{w,\delta}$ be continuously extendable to $[0,\tau]$. This will be used in Section A.7 below.

Lemma 5
Assume that either $\delta \in (0,1]$ and (37) holds, or that $\delta = 0$ and

$$C_{w,3} = \int_0^\tau (1+x^2) w(x)^{-1} dx < \infty. \tag{43}$$

Then any function $h \in \mathcal{H}_{w,\delta}$ satisfies $\int_0^\tau |h'(x)|\, dx < \infty$, and thus admits a finite limit $h(\tau) = \lim_{x\to\tau} h(x)$. In particular, h is bounded.

Proof. Assume first that $\delta \in (0,1]$. Using the Cauchy–Schwarz inequality, we derive $\int_0^\tau |h'(x)|dx = \int_0^\tau w(x)^{-\frac{1}{2}}|h'(x)|w(x)^{\frac{1}{2}}\,dx \le C_{w,0}^{\frac{1}{2}}\delta^{-\frac{1}{2}}\|h\|_{w,\delta} < \infty$ by (37). If $\delta = 0$, we use the stronger assumption (43) and obtain $\int_0^\tau |h'(x)|dx \le \int_0^\tau \int_x^\tau |h''(t)|\,dt\,dx = \int_0^\tau tw(t)^{-\frac{1}{2}}|h''(t)|w(t)^{\frac{1}{2}}\,dt \le C_{w,3}^{\frac{1}{2}}\|h\|_{w,0} < \infty$. In either case, we obtain that $h(x) = \int_0^x h'(t)dt$ converges to a finite limit as $x \to \tau$, as claimed. □

A.4 Exponential Weight Function

We now provide closed-form expressions for the reproducing kernel for the important case of an exponential weight function $w(x) = \mathrm{e}^{\alpha x}$ for some $\alpha \ge 0$, which satisfies (37), (39) and (43) if $\alpha > 0$. We denote by $\mathcal{H}_{\alpha,\delta}$ the corresponding RKHS and by $\|\cdot\|_{\alpha,\delta}$ the norm (36), in accordance with (8).

We henceforth focus on, and derive results for, the infinite time horizon $\tau = \infty$, which is the choice in the main text. Most of the results are carry over to $\tau < \infty$ in a straightforward way. See Section A.4.5.

Lemma 6
Let $\tau = \infty$ and $w(x) = \mathrm{e}^{\alpha x}$ for some $\alpha > 0$. The reproducing kernel of $\mathcal{H}_{\alpha,\delta}$ is given by (14)–(16) *according to Cases* (iii)–(v) *in Theorem 2.*

Proof. Expressions (14) and (16) follow directly integrating the right hand side of (42).

It remains to prove (15). Thereto fix $y \ge 0$. Equation (40) becomes a non-homogeneous linear differential equation with constant coefficients for ϕ,

$$-\frac{\delta}{1-\delta}\phi(t) + \alpha\phi'(t) + \phi''(t) = -\frac{1}{1-\delta}1_{[0,y]}(t)\mathrm{e}^{-\alpha t}. \tag{44}$$

This can be solved in closed form, see, e.g., (Teschl, 2012, Section 3.3). The characteristic equation, $-\delta/(1-\delta) + \alpha t + t^2 = 0$, has roots $t = -\ell_1, -\ell_2$. Hence the general solution to the homogeneous equation is $\phi_h(t) = c_1\mathrm{e}^{-\ell_1 t} + c_2\mathrm{e}^{-\ell_2 t}$, for constants c_1, c_2 to be determined by the boundary conditions. Plugging this in (Teschl, 2012, Equation (3.57)), along with the boundary conditions stated in Lemma 3, that is, $\phi'(0) = 0$ and $\lim_{t\to\infty}\phi(t) = 0$, yields

$$\phi(t) = \frac{1}{(1-\delta)\sqrt{D}}\left(-\frac{\ell_1}{\ell_2}\mathrm{e}^{-\ell_2 t}\int_0^y \mathrm{e}^{-\ell_2 s}\,ds + \mathrm{e}^{-\ell_2 t}\int_0^t 1_{[0,y]}(s)\mathrm{e}^{-\ell_1 s}\,ds + \mathrm{e}^{-\ell_1 t}\int_t^\infty 1_{[0,y]}(s)\mathrm{e}^{-\ell_2 s}\,ds\right)$$

where we use the identities $\alpha = \ell_1 + \ell_2$ and $\sqrt{D} = \ell_2 - \ell_1$. In view of (41), using the identity

$\delta/(1-\delta) = -\ell_1\ell_2$ and integration by parts, we derive

$$
\begin{aligned}
\delta\sqrt{D}k(x,y) &= \delta\sqrt{D}\int_0^x \phi(t)\,dt \\
&= \frac{\ell_1^2}{\ell_2}\left(1-\mathrm{e}^{-\ell_2 x}\right)\int_0^y \mathrm{e}^{-\ell_2 s}\,ds \\
&\quad + \left[\mathrm{e}^{-\ell_2 t}\int_0^t 1_{[0,y]}(s)\ell_1\mathrm{e}^{-\ell_1 s}\,ds\right]_0^x - \int_0^x \mathrm{e}^{-\ell_2 t}1_{[0,y]}(t)\ell_1\mathrm{e}^{-\ell_1 t}dt \\
&\quad + \left[\mathrm{e}^{-\ell_1 t}\int_t^\infty 1_{[0,y]}(s)\ell_2\mathrm{e}^{-\ell_2 s}\,ds\right]_0^x + \int_0^x \mathrm{e}^{-\ell_1 t}1_{[0,y]}(t)\ell_2\mathrm{e}^{-\ell_2 t}dt \\
&= \frac{\ell_1^2}{\ell_2^2}\left(1-\mathrm{e}^{-\ell_2 x}\right)\left(1-\mathrm{e}^{-\ell_2 y}\right) \qquad (45)\\
&\quad + \mathrm{e}^{-\ell_2 x}\left(1-\mathrm{e}^{-\ell_1(x\wedge y)}\right) - \frac{\ell_1}{\alpha}\left(1-\mathrm{e}^{-\alpha(x\wedge y)}\right) \\
&\quad + \mathrm{e}^{-\ell_1 x}\left(\mathrm{e}^{-\ell_2(x\wedge y)}-\mathrm{e}^{-\ell_2 y}\right) - \left(1-\mathrm{e}^{-\ell_2 y}\right) + \frac{\ell_2}{\alpha}\left(1-\mathrm{e}^{-\alpha(x\wedge y)}\right) \\
&= \frac{\ell_1^2}{\ell_2^2}\left(1-\mathrm{e}^{-\ell_2 x}\right)\left(1-\mathrm{e}^{-\ell_2 y}\right) \\
&\quad + \mathrm{e}^{-\ell_2 x} + \mathrm{e}^{-\ell_2 y} - 1 + \frac{\sqrt{D}}{\alpha}\left(1-\mathrm{e}^{-\alpha(x\wedge y)}\right) - \mathrm{e}^{-\ell_1(x\wedge y)-\ell_2(x\vee y)},
\end{aligned}
$$

where we used the identity $-\mathrm{e}^{-\ell_2 x}\mathrm{e}^{-\ell_1(x\wedge y)} + \mathrm{e}^{-\ell_1 x}\left(\mathrm{e}^{-\ell_2(x\wedge y)}-\mathrm{e}^{-\ell_2 y}\right) = -\mathrm{e}^{-\ell_1(x\wedge y)-\ell_2(x\vee y)}$ in the last equality. Dividing both sides of (45) by $\delta\sqrt{D}$ and some algebraic simplifications prove (15). This completes the proof. □

A.4.1 Constant Weight Function

We next study the special case of constant weight function $w = 1$, which amounts to set $\alpha = 0$. Accordingly, we denote by $\mathcal{H}_{0,\delta}$ the corresponding RKHS.

Remark 4
If $\tau = \infty$, then $\mathcal{H}_{0,\delta}$ is only a well-defined RKHS for $\delta \in (0,1]$, which we henceforth assume whenever we deal with $\mathcal{H}_{0,\delta}$ and $\tau = \infty$. This is because the constant $w = 1$ does not satisfy (37)*, as explained in Remark 2. This is also reflected by the property that $k(x,x) \to \infty$ as $\alpha \to 0$ for the kernel* (14)*, as can easily be verified.*

Lemma 7
Let $\tau = \infty$ and $w = 1$ be constant. The reproducing kernel of $\mathcal{H}_{0,\delta}$ is given by (12)–(13) *according to Cases* (i)–(ii) *in Theorem 2. If $\delta = 0$ then $\mathcal{H}_{0,\delta}$ is not a RKHS and there exists no reproducing kernel.*

Proof. We obtain the expressions on the right hand side of (12) and (13) by letting $\alpha \to 0$ in (15) and (16), respectively. It is then easily verified by Lemma 3 that these are indeed the desired reproducing kernels. The last statement follows from Remark 4. □

A.4.2 Proof of Theorem 3

Theorem 3 follows from (34) and the structural properties of $\mathcal{H}_{\alpha,\delta}$ proved in Lemma 8 below. More specifically, Case (i) Fama–Bliss follows from Lemma 8(ii). Case (ii) NSS follows from Lemma 8(i). Case (iii): let $\tilde{g}'$ be any C^1-extension of g' on $[0,\infty)$, e.g., by setting $\tilde{g}' = g'$ on $[0, x_N]$ and $\tilde{g}'(x) = g'(x_N) + g''(x_N)(x - x_N)$ for $x > x_N$. Let ψ be a C^1-function with $\psi = 1$ on $[0, x_N]$ and $\psi(x) = 0$ for $x > 2x_N$. Then $g(x) = 1 + \int_0^x \psi(t)\tilde{g}'(t)\,dt$ is C^2 on $[0,\infty)$ with $g'(x) = g''(x) = 0$ for $x > 2x_N$, and hence $g \in \mathcal{G}_{\alpha,\delta}$ as desired. Note that Case (iii) also holds if g is only once weakly differentiable for $\delta = 1$.

The following lemma is used in the proof above.

Lemma 8
Let $\alpha, \alpha' \geq 0$, $\gamma > 0$ and $\delta \in [0,1]$. The following hold:

(i) *If $g(x) = \mathrm{e}^{-\int_0^x f(z)\,dz}$ for some bounded and absolutely continuous function f such that $\lim_{x\to\infty} \frac{1}{x}\int_0^x f(z)\,dz = \gamma$ and f' is either bounded or satisfies $\int_0^\infty f'(x)^2 \mathrm{e}^{\alpha x}\,dx < \infty$. Then $h = g - 1 \in \mathcal{H}_{\alpha,\delta}$ if $\alpha < 2\gamma$.*

(ii) *If $\delta = 1$ then property* (i) *holds under the weaker assumption that f is bounded measurable and $\lim_{x\to\infty} \frac{1}{x}\int_0^x f(z)\,dz = \gamma$.*

(iii) *Let $h \in \mathcal{H}_{\alpha,\delta}$ and define $h_\gamma(x) = \mathrm{e}^{-\gamma x} h(x)$. Then $h_\gamma \in \mathcal{H}_{\alpha',\delta}$ if $\alpha' < \alpha + 2\gamma$.*

Proof. (i): Differentiation gives $h' = g' = -fg$ and $h'' = (-f' + f^2)g$. By assumption, for any $\epsilon > 0$ there exists some finite x_ϵ such that $\frac{1}{x}\int_0^x f(z)\,dz > \gamma - \epsilon/2$, and hence $g(x)^2 \mathrm{e}^{\alpha x} < \mathrm{e}^{-(2\gamma-\epsilon-\alpha)x}$, for all $x > x_\epsilon$. In particular, g is bounded. Hence, under either assumption on f', we find that $\|h\|_{\alpha,\delta} < \infty$ if $\alpha < 2\gamma - \epsilon$. As $\epsilon > 0$ was arbitrary, this proves the claim.

(ii): This follows from the same arguments as in item (i).

(iii): Differentiation gives $h_\gamma'(x) = \mathrm{e}^{-\gamma x}(h'(x) - \gamma h(x))$ and $h_\gamma''(x) = \mathrm{e}^{-\gamma x}(h''(x) - 2\gamma h'(x) + \gamma^2 h(x))$. In view of Lemma 2, the function h grows at most like x (if $\alpha > 0$) or grows at most like $x^{\frac{1}{2}}$ (if $\alpha = 0$ and thus $\delta \in (0,1]$ in view of Remark 4), respectively.[38] Hence in either case we obtain that $\|h_\gamma\|_{\alpha',\delta} < \infty$ if $\alpha' < \alpha + 2\gamma$. □

A.4.3 Another Special Case: Smith–Wilson

Smith and Wilson (2001) has been the insurance industry standard in Europe for constructing the discount curve used in the regulatory Solvency II framework. See the technical documentation of the European Insurance and Occupational Pensions Authority EIOPA (2020), and also Lagerås and Lindholm (2016) and Viehmann (2019). Smith–Wilson consider discount curves of the form $g_{SW}(x) = \mathrm{e}^{-y_\infty x} g_0(x)$, for some $g_0 \in \mathcal{G}_{0,\delta_0}$ with $\delta_0 \in (0,1)$, and $y_\infty = \log(1 + UFR) > 0$, for the

[38] In view of Lemma 5, functions in $\mathcal{H}_{\alpha,\delta}$ are actually bounded, for $\alpha > 0$.

so-called ultimate forward rate $UFR > 0$.[39] Lemma 8(i) and (iii), and Lemma 1(iv) then imply that g_{SW} lies in $\mathcal{G}_{\alpha,\delta}$ for any $\alpha \in [0, 2y_\infty)$ and $\delta \in [0,1]$.

The Smith–Wilson method assumes exact pricing of all bonds up to a certain maturity $x_N < \infty$, which is also called the last liquid point, and disregards all bonds with larger maturity. This is just a special case of our main Theorem 2 with vanishing smoothness parameter, $\Lambda = 0$. See Theorem A.1 for the technical details.

A.4.4 Consistency with Arbitrage-Free Dynamic Term Structure Models

Our framework contains all discount curves that are generated by stochastic models of the Heath, Jarrow, and Morton (1992) type. More specifically, (Filipović, 2001, Section 5.2) provides technical assumptions on the forward curve volatility asserting that the forward curve f_t prevailing at any time $t \geq 0$ satisfies $f_t - f_t(0) \in \mathcal{H}_{\alpha,1}$, whenever the initial forward curve f_0 does so, $f_0 - f_0(0) \in \mathcal{H}_{\alpha,1}$, for a given $\alpha > 0$.[40] Moreover, the infinite-maturity forward rates do not move, $f_t(\infty) = f_0(\infty)$, which is consistent with a celebrated theorem of Dybvig, Ingersoll, and Ross (1996) stating that long forward rates can never fall.[41] Lemma 8(i) and (34) then imply that the corresponding discount curves g_t lie in $\mathcal{G}_{\alpha,\delta}$ for all $t \geq 0$ if $\alpha < 2f_0(\infty)$. Examples include the short rate models of Vasicek (1977), Cox, Ingersoll, and Ross (1985), and their extensions in Hull and White (1990), see (Filipović, 2001, Sections 7.4.1–7.4.2).

A.4.5 Bounded Maturity Domain

It is straightforward to modify Lemmas 6 and 7 for a finite time horizon $\tau < \infty$. In particular, the right hand side of (42) for $\delta = 1$ reveals that the functional form of the kernel does not depend on τ. Furthermore, the case $(\alpha, \delta) = (0, 0)$ is now well specified.[42] More specifically, integrating the right hand side of (42) for $\delta = 0$ gives the closed form expressions for the kernel in the two cases:

(i) $\alpha = 0$, $\delta = 0$:

$$k(x,y) = -\frac{(x \wedge y)^3}{6} - \frac{(x \wedge y)(x \vee y)^2}{2} + xy\tau; \tag{46}$$

(ii) $\alpha > 0$, $\delta = 0$:

$$k(x,y) = -\frac{x \wedge y}{\alpha^2} \mathrm{e}^{-\alpha(x \wedge y)} + \frac{2}{\alpha^3}\left(1 - \mathrm{e}^{-\alpha(x \wedge y)}\right) - \frac{x \wedge y}{\alpha^2} \mathrm{e}^{-\alpha(x \vee y)} - \frac{xy}{\alpha} \mathrm{e}^{-\alpha\tau}. \tag{47}$$

Expression (46) corresponds exactly to the smoothing splines of Fisher, Nychka, and Zervos (1995) and Tanggaard (1997). Note that this kernel function $x \mapsto k(x,y)$ is linear-quadratic in x

[39] Smith and Wilson (2001) define the kernel $W(x,y) = \mathrm{e}^{-y_\infty(x+y)}\delta\rho k(x,y)$, which is also known as Wilson function, where $k(x,y)$ is the kernel given in (12). In view of Lemma 2, such discount curves extend well to infinity, $\lim_{x\to\infty} g_{SW}(x) = 0$. We discuss this property in more generality in Appendix A.7.

[40] In view of Lemma 5, that means that f_t absolutely continuous, admits a finite limit $f_t(\infty) = \lim_{x\to\infty} f_t(x)$, and f_t' satisfies $\int_0^\infty f_t'(x)^2 \mathrm{e}^{\alpha x}\, dx < \infty$.

[41] See also (Filipović, 2009, Lemma 7.3).

[42] In contrast to the case in Remark 4, it can be verified, e.g., by a third order Taylor expansion of $\alpha^3 k(x,y)$ in α around $\alpha = 0$, that $k(x,y)$ in (47) converges to (46) as $\alpha \to 0$.

for $x > y$, which does not seem to be a suitable shape for basis functions to span the discount curve. This could partly explain the bad performance of these cubic splines reported in the literature.

A better choice might be the weighted kernel (47), for $\alpha > 0$. But also this kernel function $x \mapsto k(x, y)$ contains a linear term, which by the way is just the difference to expression (14). Again, this term does not seem to be a suitable component for spanning the discount curve. This lets us prefer the infinite time horizon, $\tau = \infty$, which is the choice in the main text.

The expressions for the kernel in the cases where $\delta \in (0, 1)$ follow by similar calculations solving the linear differential equation (44) as in the proof of Lemma 6. We leave this to the reader.

A.5 An Alternative RKHS on Bounded Maturity Domain

As mentioned in Section A.3, the economic interpretation of the boundary condition (35) is not so obvious for a finite time horizon $\tau < \infty$. We thus propose here an alternative RKHS $\mathcal{H} = \mathcal{H}_{w,0,\eta}$, which generalizes $\mathcal{H}_{w,\delta}$ for $\delta = 0$ by relaxing (35). It consists of differentiable functions $h : [0, \tau) \to \mathbb{R}$ of the form $h(x) = \int_0^x h'(t)\, dt$ with absolutely continuous derivatives, $h'(x) = h'(0) + \int_0^x h''(t)\, dt$, for locally integrable h'', with finite norm given by

$$\|h\|_{w,0,\eta} = \left(\eta \int_0^\tau h''(x)^2 w(x)\, dx + (1-\eta) h'(\tau)^2 \right)^{\frac{1}{2}}, \tag{48}$$

for some measurable weight function $w : [0, \tau) \to [1, \infty)$, and shape parameter $\eta \in (0, 1)$. The second term in (48) replaces the boundary condition (35) such that (48) defines a bona fide norm.

It follows by definition that $\mathcal{H}_{w,0} \subset \mathcal{H}_{w,0,\eta}$ and $\|h\|_{w,0} = \eta^{-\frac{1}{2}} \|h\|_{w,0,\eta}$ for all $h \in \mathcal{H}_{w,0}$. A similar argument as in Lemma 2 shows that $\mathcal{H}_{w,0,\eta}$ is a RKHS. Its reproducing kernel can be determined similarly as in Lemma 3. More specifically, we fix $y \in [0, \tau)$, denote $\psi = k(\cdot, y) \in \mathcal{H}_{w,0,\eta}$, and let $h \in \mathcal{H}_{w,0,\eta}$. By the reproducing kernel property, ψ satisfies

$$\langle \psi, h \rangle_{w,0,\eta} = h(y) = \int_0^\tau 1_{[0,y]}(x) h'(x)\, dx. \tag{49}$$

On the other hand, integration by parts shows that

$$\begin{aligned}
\langle \psi, h \rangle_{w,0,\eta} &= \eta \int_0^\tau \psi''(x) h''(x) w(x)\, dx + (1-\eta)\psi'(\tau) h'(\tau) \\
&= \eta \psi''(\tau) w(\tau) h'(\tau) - \eta \psi''(0) w(0) h'(0) - \eta \int_0^\tau (\psi'' w)'(x) h'(x)\, dx + (1-\eta)\psi'(\tau) h'(\tau).
\end{aligned}$$

We see that (49) holds for all $h \in \mathcal{H}_{w,0,\eta}$ if ψ satisfies the linear differential equation with boundary conditions

$$
\begin{aligned}
-(\psi'' w)' &= 1_{[0,y]}, \\
\psi''(0) &= 0, \\
\eta\psi''(\tau)w(\tau) + (1-\eta)\psi'(\tau) &= 0, \\
\psi(0) &= 0,
\end{aligned}
$$

which corresponds to (40). It can be easily verified that the solution $\psi(x) = k(x,y)$ is given by

$$
k(x,y) = \frac{\eta}{1-\eta} xy + \int_0^\tau (t \wedge x)(t \wedge y) w(t)^{-1}\, dt, \tag{50}
$$

which obviously generalizes (42) for $\delta = 0$. Indeed, (42) is a boundary case of (50) for $\eta \to 0$.

Expression (50) can now be combined with (46) and (47) to obtain the closed form expressions for the reproducing kernel of the RKHS $\mathcal{H}_{w,0,\eta}$ for the exponential weight function $w(x) = \mathrm{e}^{\alpha x}$ in the two cases:

(i) $\alpha = 0$:

$$
k(x,y) = -\frac{(x\wedge y)^3}{6} - \frac{(x \wedge y)(x \vee y)^2}{2} + xy\left(\frac{\eta}{1-\eta} + \tau\right); \tag{51}
$$

(ii) $\alpha > 0$:

$$
k(x,y) = -\frac{x\wedge y}{\alpha^2}\mathrm{e}^{-\alpha(x\wedge y)} + \frac{2}{\alpha^3}\left(1 - \mathrm{e}^{-\alpha(x\wedge y)}\right) - \frac{x \wedge y}{\alpha^2}\mathrm{e}^{-\alpha(x\vee y)} + xy\left(\frac{\eta}{1-\eta} - \frac{\mathrm{e}^{-\alpha\tau}}{\alpha}\right). \tag{52}
$$

The functional shapes of (51) and (52) are essentially the same as those of (46) and (47), respectively, so the same criticism as in Section A.4.5 applies here. There is one intriguing exception, namely when $\frac{\eta}{1-\eta} = \frac{\mathrm{e}^{-\alpha\tau}}{\alpha}$. In this particular case, (52) is exactly the restriction to $[0,\tau)$ of the reproducing kernel (14) of the RKHS $\mathcal{H}_{\alpha,0}$ on $[0,\infty)$.

A.6 Gaussian Process Perspective

As in Section 2.4, we assume here that the discount curve $g : [0,\infty) \to \mathbb{R}$ is a Gaussian process with mean function m and covariance kernel k. For more background and applications of Gaussian processes we refer to, e.g., Rasmussen and Williams (2006).

As $g(0) = 1$, the mean function and kernel must satisfy $m(0) = 1$ and $k(0,0) = 0$. The mean function m can be interpreted as a prior for the discount curve. E.g., m could be the discount curve estimated on a previous business day. An alternative prior could be $m(x) = \mathrm{e}^{-rx}$ for some constant prior interest rate r.

We also assume that the errors ϵ_i in the pricing equation (1) are modeled as independent centered Gaussian random variables with variance parameters $\sigma_i^2 \geq 0$, that is $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \Sigma^\epsilon)$ with $\Sigma^\epsilon = \mathrm{diag}(\sigma_1^2, \dots, \sigma_M^2)$. This is intimately related to the kernel ridge regression problem (25) with variance weights $\omega_i = \lambda/\sigma_i^2$, as we shall see now. In particular, an exact pricing of security i amounts to a zero variance of ϵ_i, that is, $\sigma_i^2 = 0$.

Indeed, for n arbitrary cash flow dates $\boldsymbol{z} = (z_1, \dots, z_n)^\top$, we have that $g(\boldsymbol{z})$ and P are jointly Gaussian distributed as

$$\begin{pmatrix} g(\boldsymbol{z}) \\ P \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} m(\boldsymbol{z}) \\ Cm(\boldsymbol{x}) \end{pmatrix}, \begin{pmatrix} k(\boldsymbol{z}, \boldsymbol{z}^\top) & k(\boldsymbol{z}, \boldsymbol{x}^\top)C^\top \\ Ck(\boldsymbol{x}, \boldsymbol{z}^\top) & C\boldsymbol{K}C^\top + \Sigma^\epsilon \end{pmatrix} \right). \tag{53}$$

Here, as before, we write $k(\boldsymbol{z}, \boldsymbol{x}^\top)$ for the $n \times N$ matrix with entries $k(z_i, x_j)$, and similarly $k(\boldsymbol{z}, \boldsymbol{z}^\top)$, $m(\boldsymbol{z})$, so that $\boldsymbol{K} = k(\boldsymbol{x}, \boldsymbol{x}^\top)$. As in Theorem A.1, we henceforth assume that either $\mathcal{I}_1 = \emptyset$ or $C_{\mathcal{I}_1}\boldsymbol{K}C_{\mathcal{I}_1}^\top$ is invertible, so that the $M \times M$-matrix $C\boldsymbol{K}C^\top + \Sigma^\epsilon$ is invertible.

Bayesian updating in (53) implies that the conditional distribution of g, given the observed prices P, is then still Gaussian with posterior mean function (18) and posterior variance given by the posterior kernel (19). We thus have proved the following lemma, as desired.

Lemma 9

Suppose the kernel k and the prior mean function $m = p$ are as in Section A.2, and $\Sigma = \Lambda$ as given in Theorem A.1. Then the posterior mean function in Equation (18) *coincides with the kernel ridge regression estimator $\hat{g}$ in Theorem A.1.*

A.6.1 Proof of Theorem 4

Theorem 4 now follows from the above discussion and Lemma 9, for the constant prior curve $p = 1$ in (24), and noting (34).

A.6.2 Scaling Invariance of the Posterior Mean Function

We end this section with an interesting and useful observation. Inspection shows that the posterior mean function (18) is invariant with respect to scaling the kernel k and the pricing error variance Σ by any common factor $s > 0$. That is, if we replace k and Σ by $\tilde{k} = sk$ and $\tilde{\Sigma} = s\Sigma$, respectively, the scalings on the right hand side of (18) offset. On the other hand, the scaling obviously impacts the prior and posterior variance of the Gaussian process g. From (53) we obtain the prior log-likelihood function of s and all else equal, given the observed prices P,

$$\mathcal{L}(s) = -q_2 \frac{1}{s} - \frac{M}{2} \log(s) - q_1$$

for $q_2 = \frac{1}{2}(P - Cm(\boldsymbol{x}))^\top (C\boldsymbol{K}C^\top + \Sigma)^{-1}(P - Cm(\boldsymbol{x}))$, $q_1 = \frac{1}{2}\log|C\boldsymbol{K}C^\top + \Sigma| + \frac{M}{2}\log(2\pi)$. We can now calibrate the scaling factor s by maximizing the log-likelihood. Remarkably, this is given in closed form. Indeed, by differentiation, it is readily verified that $\mathcal{L}(s)$ attains its maximum at

$$\hat{s} = \frac{2q_2}{M}. \tag{54}$$

Note that in terms of the kernel ridge regression problem (25), the scaling amounts to replace k by $\tilde{k} = sk$, as above, and λ by $\tilde{\lambda} = s\lambda$. Indeed, the RKHS $\tilde{\mathcal{H}}$ corresponding to the kernel $\tilde{k}$ coincides with $\mathcal{H}$ as a set, $\tilde{\mathcal{H}} = \mathcal{H}$, whereas the squared norms are related by $\|h\|^2_{\tilde{\mathcal{H}}} = \frac{1}{s}\|h\|^2_{\mathcal{H}}$ for any function $h \in \mathcal{H}$.[43]

A.7 Infinite-Maturity Yield

In this section, we assume that the RKHS $\mathcal{H}$ consists of functions $x \mapsto h(x)$ that are defined on the closure $[0,\infty]$ of $[0,\infty)$, including the point $x = \infty$. This is tantamount to assuming that the reproducing kernel k is can be extended to $[0,\infty] \times [0,\infty]$, see (Paulsen and Raghupathi, 2016, Corollary 5.8). We also assume that the prior curve $x \mapsto p(x)$ is defined on $[0,\infty]$. In this case, all results of the previous sections literally carry over, as we never assumed that the last cash flow date x_N is finite. In this setting, we now study the infinite-maturity yield. We do so first for the general case and then derive more explicit results for the weighted Sobolev type space with exponential weight function.

A.7.1 General Case

We let $\hat{g} = p + \hat{h} : [0,\infty] \to \mathbb{R}$ be the optimal discount curve given in Theorem A.1, and denote by $\hat{y}(x) = -\frac{1}{x}\log \hat{g}(x)$ the corresponding zero-coupon yield curve. The next lemma gives general sufficient conditions for the existence and characterization of the infinite-maturity yield.

Lemma 10
Assume that there exists a positive function q and parameter $r > 0$ such that the following limits exist,

$$\lim_{x\to\infty} \frac{1}{x}\log q(x) = 0, \tag{55}$$

$$\lim_{x\to\infty} (p(x) - p(\infty))q(x)\mathrm{e}^{rx} = \gamma_0, \tag{56}$$

$$\lim_{x\to\infty} (k(x,x_j) - k(\infty,x_j))q(x)\mathrm{e}^{rx} = \gamma_j, \quad j = 1,\dots,N, \tag{57}$$

for some real $\gamma_0,\dots,\gamma_N$ such that $\gamma_0 + \sum_{j=1}^N \beta_j\gamma_j > 0$. Then $\lim_{x\to\infty}\hat{y}(x) = r$ if and only if $\hat{g}(\infty) = 0$.

Proof. By assumption (55), we have $q(x)\mathrm{e}^{rx} \to \infty$ as $x \to \infty$. Hence, (56) and (57) imply $\lim_{x\to\infty}\hat{g}(x) = \hat{g}(\infty)$. Hence $\lim_{x\to\infty}\hat{y}(x) = r > 0$ only if $\hat{g}(\infty) = 0$, which proves the sufficiency of the statement.

To prove necessity, we now assume that $\hat{g}(\infty) = 0$. Then we can write $\hat{g}(x) = \hat{g}(x) - \hat{g}(\infty) = p(x) - p(\infty) + \sum_{j=1}^N (k(x,x_j) - k(\infty,x_j))\beta_j$, and thus $\lim_{x\to\infty}\hat{g}(x)q(x)\mathrm{e}^{rx} = \gamma_0 + \sum_{j=1}^N \beta_j\gamma_j > 0$. We obtain

$$\hat{y}(x) = \frac{-\log(\hat{g}(x)q(x)\mathrm{e}^{rx})}{x} + \frac{\log q(x)}{x} + r,$$

[43] Indeed, this can easily be seen for functions of the form $h = \sum_{j=1}^n \beta_j k(\cdot,x_j)$.

which converges to r. This completes the proof of the lemma. $\square$

In order to identify the infinite-maturity yield, in view of Lemma 10, it is useful to impose the constraint $\hat{g}(\infty) = 0$ in Problem (25). This can be done by introducing a synthetic zero-coupon bond with infinite maturity, as explained as a special case of the next lemma. We denote by $\boldsymbol{K}_N = (k(x_1, x_N), \dots, k(x_N, x_N))$ the N-th row of $\boldsymbol{K}$.

Lemma 11
Assume $\omega_M = \infty$ and $C_M = (0, \dots, 0, 1)$, which corresponds to the exact pricing of the zero-coupon bond with maturity $x_N \leq \infty$ and face value 1, whose present value is P_M. Assume further that the first $M-1$ securities have no cash flow at x_N, i.e., $C_{iN} = 0$ for all $i < M$.

Then the last component of β in (28) is given by

$$\beta_N = \frac{P_M - \tilde{g}(x_N)}{s} \tag{58}$$

where $\tilde{g}(x_N) = p(x_N) + \boldsymbol{K}_N \tilde{C}^\top (\tilde{C}\boldsymbol{K}\tilde{C}^\top + \tilde{\Lambda})^{-1}(\tilde{P} - \tilde{C}p(\boldsymbol{x}))$ is the value at x_N of the discount curve $\tilde{g}$ estimated based on the first $M-1$ securities with prices $\tilde{P} = (P_1, \dots, P_{M-1})^\top$ and $(M-1) \times N$-cash flow matrix $\tilde{C}$ given by $\tilde{C}^\top = (C_1^\top, \dots, C_{M-1}^\top)$, for $\tilde{\Lambda} = \operatorname{diag}(\lambda/\omega_1, \dots, \lambda/\omega_{M-1})$, and where $s > 0$ is given in (59).

If $x_N < \infty$, this corresponds to fixing an exogenous target zero-coupon yield y_N for maturity x_N, which is encoded by setting $P_M = \mathrm{e}^{-x_N y_N}$, so that $\hat{g}(x_N) = \mathrm{e}^{-x_N y_N}$. If $x_N = \infty$, we set $P_M = 0$, so that $\hat{g}(\infty) = 0$.

Proof. Using blockwise inversion of $\mathcal{M} = C\boldsymbol{K}C^\top + \Lambda = \begin{pmatrix} \tilde{C}\boldsymbol{K}\tilde{C}^\top + \tilde{\Lambda} & \tilde{C}\boldsymbol{K}_N^\top \\ \boldsymbol{K}_N\tilde{C}^\top & k(x_N, x_N) \end{pmatrix}$ in (28) gives the M-th row of $\mathcal{M}^{-1}$ by $(-s^{-1}\boldsymbol{K}_N\tilde{C}^\top(\tilde{C}\boldsymbol{K}\tilde{C}^\top + \tilde{\Lambda})^{-1}, s^{-1})$, for the Schur complement

$$s = k(x_N, x_N) - \boldsymbol{K}_N\tilde{C}^\top(\tilde{C}\boldsymbol{K}\tilde{C}^\top + \tilde{\Lambda})^{-1}\tilde{C}\boldsymbol{K}_N^\top, \tag{59}$$

which is positive, $s > 0$, as $\mathcal{M}$ is invertible. Hence $\beta_N = -s^{-1}\boldsymbol{K}_N\tilde{C}^\top(\tilde{C}\boldsymbol{K}\tilde{C}^\top + \tilde{\Lambda})^{-1}(\tilde{P} - \tilde{C}p(\boldsymbol{x})) + s^{-1}(P_M - p(x_N))$, which proves the claim. $\square$

A.7.2 Weighted Sovolev Type Space with Exponential Weight Function

The weighted Sobolev type space $\mathcal{H} = \mathcal{H}_{\alpha,\delta}$ with exponential weight function $w(x) = \mathrm{e}^{\alpha x}$ for any $\alpha > 0$ obviously satisfies (37) and (43), so that Lemma 5 applies and functions in $\mathcal{H}_{\alpha,\delta}$ extend continuously to $[0, \infty]$. As for the identification of the infinite-maturity yield in the case of an exponential weight function we have the following result.

Theorem A.2 (Identification of infinite-maturity yield and positivity of discount curve)
Let $w(x) = \mathrm{e}^{\alpha x}$ for some $\alpha > 0$, and include a synthetic zero-coupon bond with infinite maturity, $x_N = \infty$, which is exactly priced as in Lemma 11 so that $\hat{g}(\infty) = 0$. Then the following hold:

(i) *Let $r > 0$ be a parameter and assume that either*

(a) $r < \alpha$, and the limit (56) exists for $q = 1$ and is finite and positive, $\gamma_0 > 0$; or

(b) $r \geq \alpha$, and the limit (56) exists for

$$q(x) = \begin{cases} (1+x)^{-1}, & \text{if } \delta = 0, \\ 1, & \text{if } \delta \in (0, 1], \end{cases}$$

and is finite and non-negative, $\gamma_0 \geq 0$, and that $\beta_N < 0$.

Then the infinite-maturity yield exists and is equal to $\lim_{x\to\infty} \hat{y}(x) = \alpha \wedge r$.

(ii) *Assume that the prior curve is constant, $p = 1$, that $\beta_N \leq 0$, and that $\hat{g}(x_{N-1}) > 0$ for the largest finite maturity x_{N-1}. Then $\hat{g}(x) > 0$ for all finite $x > x_{N-1}$.*

Proof. Part (i) follows from combining Lemmas 10, 11 and 12. Part (ii) follows from the explicit expressions (62)–(64), noting that $\hat{g}(x) = \sum_{j=1}^{N}(k(x, x_j) - k(\infty, x_j))\beta_j$ by assumption. □

The assumptions in Theorem A.2 arguably are abstract. The following examples serve as illustration. Case (i)a is satisfied for the prior curve $p(x) = \mathrm{e}^{-rx}$. Case (i)b is satisfied for the constant prior curve, $p = 1$, or $p(x) = \mathrm{e}^{-\alpha x}$, while the sign of β_N has to be computed case by case.[44]

While Theorem A.2(i) implies that there exists some z_0 such that $\hat{g}(x) > 0$ for all finite $x > z_0$, it does not guarantee that $\hat{g}(x) > 0$ for all finite $x > x_{N-1}$, for the largest finite cash flow date x_{N-1}. A sufficient condition is therefore given in Theorem A.2(ii).

The following lemma is needed in the proof of Theorem A.2.

Lemma 12

Let $w(x) = \mathrm{e}^{\alpha x}$ for some $\alpha > 0$, and assume that $x_N = \infty$. Then the following hold:

(i) *If $\delta = 0$ then the kernel (14) satisfies (57) for the function $q(x) = (1 + x)^{-1}$,*

$$\lim_{x\to\infty} (k(x, x_j) - k(\infty, x_j))(1 + x)^{-1}\mathrm{e}^{\alpha x} = \begin{cases} 0, & j < N, \\ -\frac{1}{\alpha^2}, & j = N. \end{cases} \tag{60}$$

(ii) *If $\delta \in (0, 1]$ then the kernels (15) and (16) satisfy (57) for the constant function $q = 1$,*

$$\lim_{x\to\infty} (k(x, x_j) - k(\infty, x_j))\mathrm{e}^{\alpha x} = \begin{cases} 0, & j < N, \\ -\frac{1}{\alpha\delta}, & j = N. \end{cases} \tag{61}$$

[44] The sign of β_N coincides with the sign of $-\tilde{g}(x_N)$ in (58). But this is of theoretical rather than practical interest.

Proof. Fix $y \geq 0$. For the kernel (14), we derive

$$k(x,y) - k(\infty,y) = \begin{cases} -\frac{y}{\alpha^2}\mathrm{e}^{-\alpha x}, & \text{if } x > y, \\ -\left(\frac{x}{\alpha^2} + \frac{2}{\alpha^3}\right)\mathrm{e}^{-\alpha x}, & \text{if } y = \infty, \end{cases} \tag{62}$$

which proves (60). For the kernel (15), we derive

$$k(x,y) - k(\infty,y) = \begin{cases} \left(\frac{\alpha}{\delta \ell_2^2} + \frac{1}{\delta\sqrt{D}}\left(\frac{\ell_1^2}{\ell_2^2}\mathrm{e}^{-\ell_2 y} - \mathrm{e}^{-\ell_1 y}\right)\right)\mathrm{e}^{-\ell_2 x}, & \text{if } x > y, \\ \frac{\alpha}{\delta \ell_2^2}\mathrm{e}^{-\ell_2 x} - \frac{1}{\alpha\delta}\mathrm{e}^{-\alpha x}, & \text{if } y = \infty, \end{cases} \tag{63}$$

which proves (61) for $\delta \in (0,1)$. Similarly, for the kernel (16), we derive

$$k(x,y) - k(\infty,y) = \begin{cases} 0, & \text{if } x > y, \\ -\frac{1}{\alpha}\mathrm{e}^{-\alpha x}, & \text{if } y = \infty, \end{cases} \tag{64}$$

which proves (61) for $\delta = 1$, and thus completes the proof. □

Remark 5

For the same reason as mentioned in Remark 4, Lemma 5 does not apply for the Sobolev type space $\mathcal{H} = \mathcal{H}_{0,\delta}$ with constant weight function, $w = 1$, for any $\delta \in (0,1]$. Indeed, the unbounded function $h(x) = (1+x)^{\frac{1}{4}} - 1$ belongs to $\mathcal{H}_{0,\delta}$. While being consistent with Lemma 2, this example shows that the functions in $\mathcal{H}_{0,\delta}$ cannot be extended to $[0,\infty]$ in general. In particular, we cannot impose the constraint $\hat{g}(\infty) = 0$, and there is no way to identify the infinite-maturity yield as we did in Theorem A.2 for the exponential weight function with $\alpha > 0$.

A.8 Implementation

We provide the specification details of the implementation. This includes the choice of pricing error weights and a scale-normalization of the smoothness parameter.

A.8.1 Pricing Error Weights

Our objective function minimizes duration weighted pricing errors, which corresponds up to first order to yield pricing errors. This is the same weight as used among others by Gürkaynak, Sack, and Wright (2007). Here we provide further details. The yield to maturity (YTM) of bond i equals the root $Y_i = y$ of $P_i = \Pi_i(y)$, where we define the discounted cash flows for a given yield as $\Pi_i(y) = \sum_{j=1}^{N} C_{ij}\mathrm{e}^{-yx_j}$. The modified duration of bond i measures its sensitivity to YTM changes and is defined as

$$D_i = -\frac{1}{P_i}\Pi_i'(Y_i) = \frac{1}{P_i}\sum_{j=1}^{N} C_{ij}x_j\mathrm{e}^{-Y_i x_j}.$$

Note that both, Y_i and D_i, can be readily derived from market data. The YTM corresponding to the fundamental value of bond i is given as the root $y = Y_i^g$ of $\Pi_i(y) = P_i^g$. A Taylor expansion of $\Pi_i(y)$ at $y = Y_i$ gives, up to first order,

$$\underbrace{P_i - P_i^g}_{\text{pricing error } \epsilon_i} = -(\Pi_i(Y_i^g) - \Pi_i(Y_i)) = -\Pi_i'(Y_i)(Y_i^g - Y_i) + o(Y_i^g - Y_i) \approx D_i P_i \underbrace{(Y_i^g - Y_i)}_{\text{YTM error}}. \quad (65)$$

Dividing both sides of (65) by $D_i P_i$ suggests to use weights ω_i as shown in (17).

As an alternative, we also report the relative pricing error. This corresponds to the particular choice of weights $\omega_i = \frac{1}{M}\frac{1}{P_i^2}$, so that the weighted mean squared pricing error in the objective function (9) equals the relative pricing mean squared error,

$$\sum_{i=1}^{M} \omega_i (P_i - P_i^g)^2 = \frac{1}{M} \sum_{i=1}^{M} \left(\frac{P_i - P_i^g}{P_i} \right)^2 .$$

A.8.2 Scale-Normalization of λ

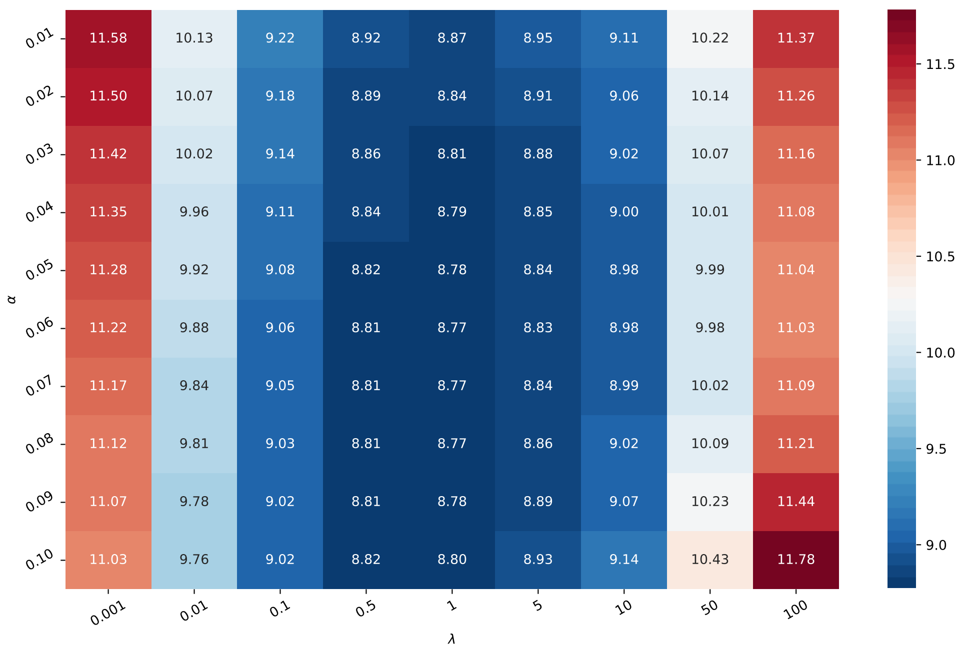
In order to have a meaningful scale and make the smoothness parameter λ comparable across the time-series of bonds, we normalize it by the largest maturity for a given day. More specifically, we replace λ in the objective function (9) by $\lambda/(365 x_N)$, where x_N is the largest maturity in years. The intuition behind is that on any given day the effective trade-off between pricing errors and curve smoothness takes place on the interval $[0, x_N]$. Beyond this time horizon, for $x > x_N$ the first and second derivatives of the discount curve, $g'(x)^2$ and $g''(x)^2$, are minimized without restrictions and thus close to zero. In other words, the effective integration domain is $[0, x_N]$ in the smoothness measure (8). The additional scaling of x_N by 365 is convenient, as it turns out to be essentially optimal such that our baseline model uses $\lambda = 1$. That is, the factor in front of the smoothness measure $\|g\|^2_{\alpha,\delta}$ is $1/(365 x_N)$.

B Empirical Results

In this appendix we collect additional empirical results.

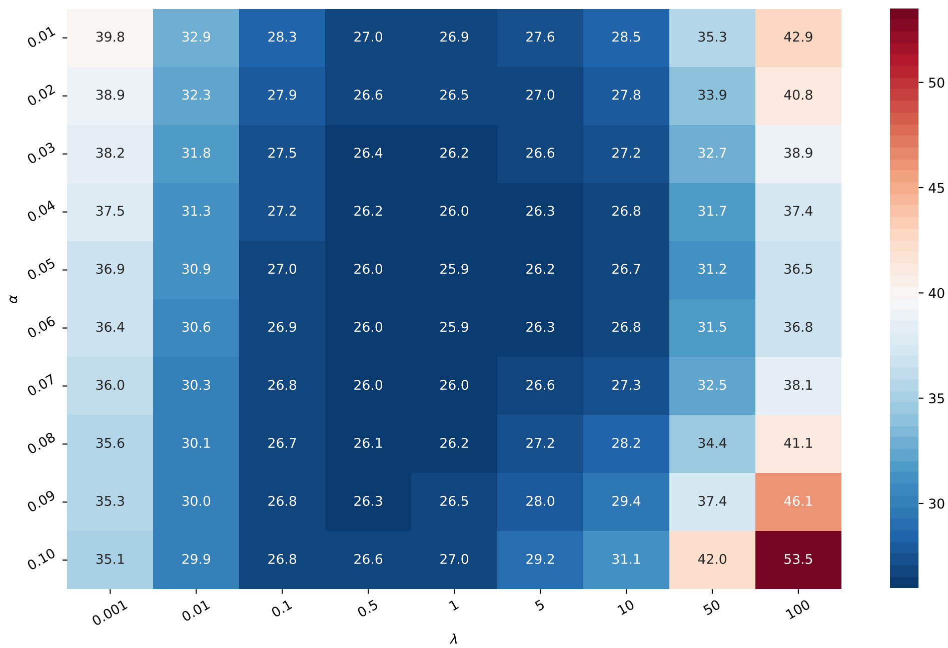
B.1 Parameter Selection

Figure A.1: Cross-validation duration weighted pricing RMSE for λ and α



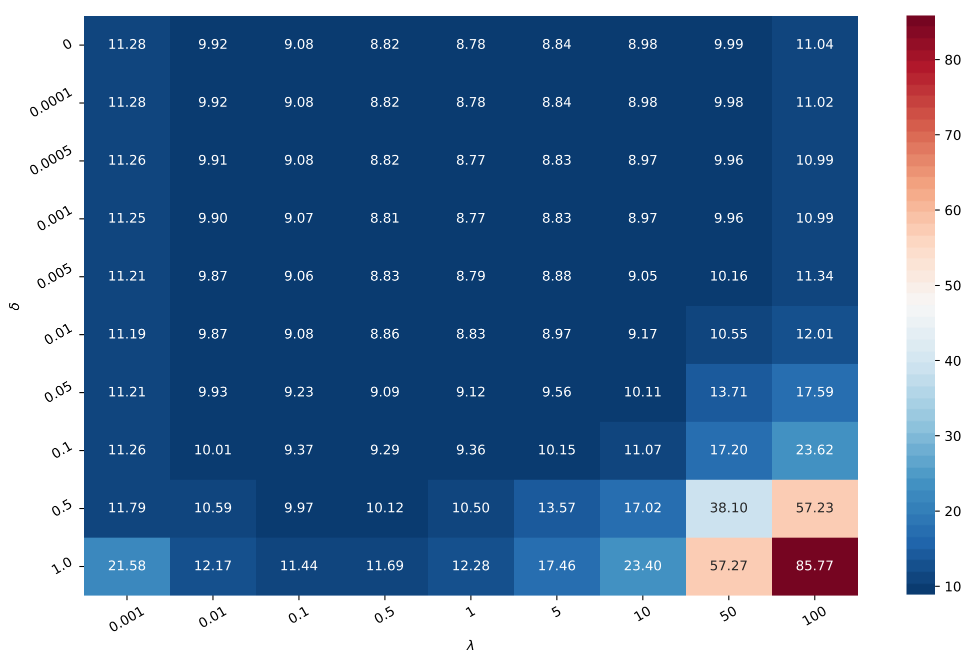
This figure shows the cross-validation duration weighted pricing RMSE in BPS for our KR method as a function of the smoothness parameter λ and maturity weight α. The tension parameter is set to $\delta = 0$. The results are calculated using quarterly data from June 1961 to December 2020.

Figure A.2: Cross-validation relative pricing RMSE for λ and α



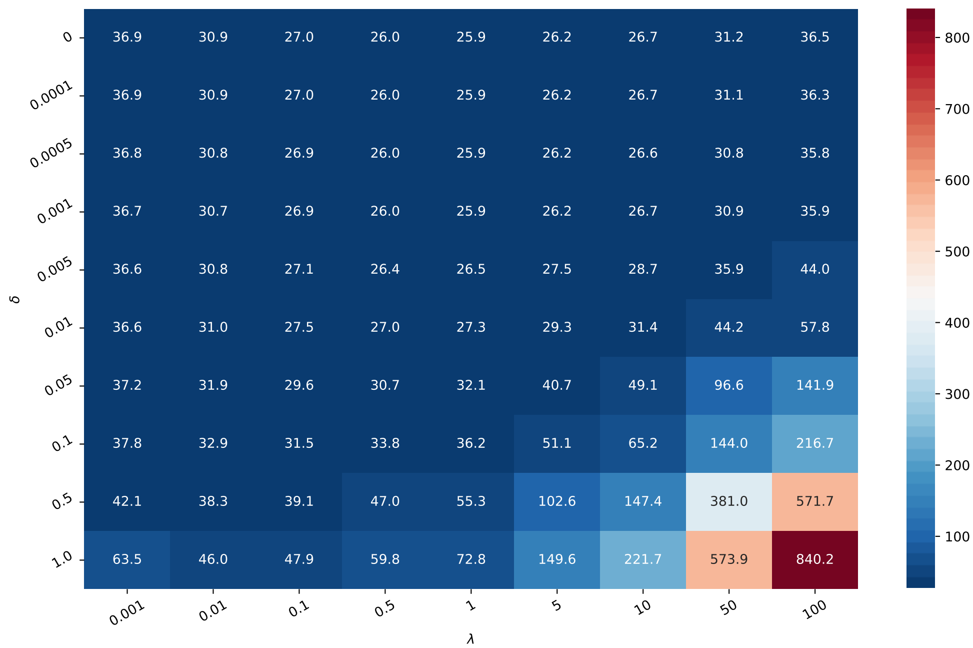
This figure shows the cross-validation relative pricing RMSE in BPS for our KR method as a function of the smoothness parameter λ and maturity weight α. The tension parameter is set to $\delta = 0$. The results are calculated using quarterly data from June 1961 to December 2020.

Figure A.3: Cross-validation duration weighted pricing RMSE for λ and δ



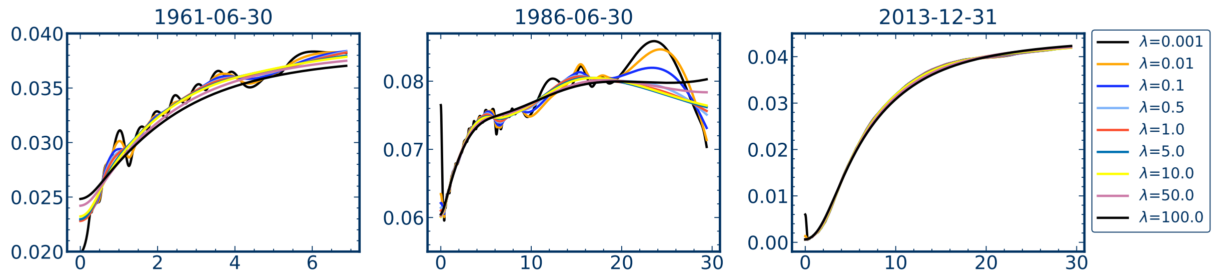
This figure shows the cross-validation duration weighted pricing RMSE in BPS for our KR method as a function of the smoothness parameter λ and tension parameter δ. The maturity weight is set to $\alpha = 0.05$. The results are calculated using quarterly data from June 1961 to December 2020.

Figure A.4: Cross-validation relative pricing RMSE for λ and δ

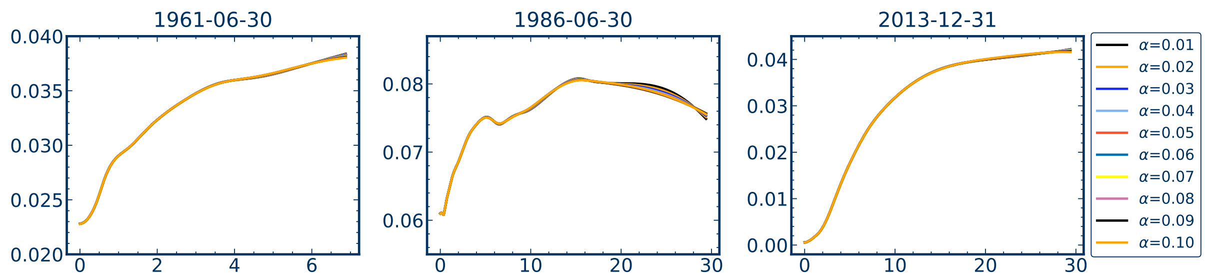


This figure shows the cross-validation relative pricing RMSE in BPS for our KR method as a function of the smoothness parameter λ and tension parameter δ. The maturity weight is set to $\alpha = 0.05$. The results are calculated using quarterly data from June 1961 to December 2020.

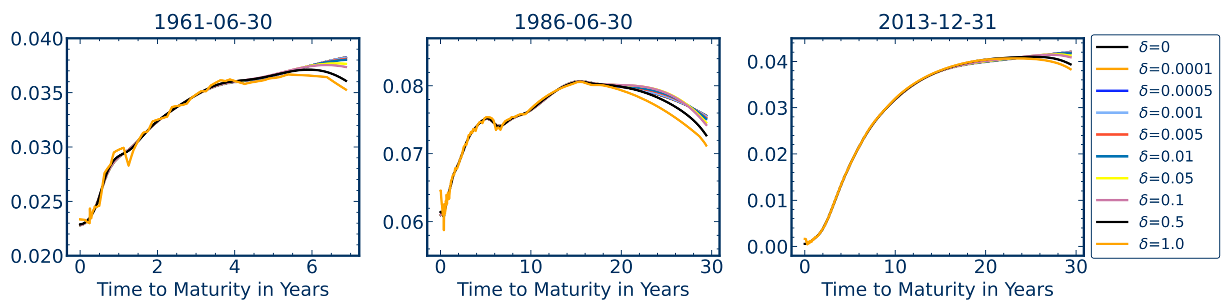
Figure A.5: KR yield curve estimates as a function of parameters



(a) Varying λ, fixed $\alpha = 0.05$ and $\delta = 0$



(b) Varying α, fixed $\lambda = 1$ and $\delta = 0$



(c) Varying δ, fixed $\alpha = 0.05$ and $\lambda = 1$

This figure shows yield curve estimates with KR for various combination of parameters on the representative example dates 1961-06-30 (top panel), 1986-06-30 (mid panel), and 2013-12-31 (bottom panel). The first column varies the smoothness parameter λ for fixed values $\alpha = 0.05$ and $\delta = 0$. The second column varies the maturity weight α for fixed values $\lambda = 1$ and $\delta = 0$. The third columns varies the tension parameter δ for fixed values $\lambda = 1$ and $\alpha = 0.05$.

B.2 Comparison Study

Table A.1: In-sample pricing comparison

	Price Dur.-w.	Price Relati.	YTM Aveg.	YTM Mat.-w.	YTM <3M	YTM 3M to 1Y	YTM 1Y to 2Y	YTM 2Y to 3Y	YTM 3Y to 4Y	YTM 4Y to 5Y	YTM 5Y to 7Y	YTM 7Y to 10Y	YTM 10Y to 20Y	YTM >20Y
Sample 6/1961-12/2013 (including FB)														
Full Data														
KR	**18.89**	**17.54**	**18.91**	**9.35**	43.60	**8.73**	**8.01**	**5.62**	**5.05**	**6.16**	**6.25**	**6.04**	**3.24**	0.78
NSS	30.99	27.56	31.02	14.41	68.26	13.94	10.27	7.54	6.86	8.23	8.21	10.12	7.19	3.45
GSW	35.43	28.05	35.47	15.49	76.55	16.88	10.21	7.72	6.65	7.93	7.95	10.21	7.29	3.48
LW	20.22	19.27	20.24	10.14	46.47	9.83	8.47	5.75	5.13	6.32	6.36	7.46	4.79	**0.77**
FB	18.91	28.68	18.95	11.35	**38.48**	10.26	10.40	9.99	9.09	8.49	7.99	8.17	9.42	1.21
3-Month Filter														
KR	**7.98**	**18.99**	**8.00**	**5.54**		**8.73**	**8.01**	**5.62**	**5.05**	**6.16**	**6.25**	**6.04**	**3.24**	0.78
NSS	11.52	29.71	11.54	8.42		13.94	10.27	7.54	6.86	8.23	8.21	10.12	7.19	3.45
GSW	13.22	29.90	13.26	8.70		16.88	10.21	7.72	6.65	7.93	7.95	10.21	7.29	3.48
LW	8.67	20.90	8.70	6.10		9.83	8.47	5.75	5.13	6.32	6.36	7.46	4.79	**0.77**
FB	10.75	31.10	10.80	8.33		10.26	10.40	9.99	9.09	8.49	7.99	8.17	9.42	1.21
KR Filter														
KR	**9.51**	**16.27**	**9.52**	**6.41**	18.11	**7.73**	**7.18**	**5.05**	**4.74**	**5.57**	**5.91**	**5.81**	**3.18**	0.78
NSS	22.61	26.51	22.62	11.75	45.91	13.09	9.31	6.91	6.50	7.56	7.82	9.88	7.13	3.45
GSW	27.13	26.93	27.17	12.84	54.43	16.16	9.27	7.03	6.25	7.20	7.49	9.93	7.21	3.48
LW	11.05	18.01	11.06	7.26	21.75	8.85	7.57	5.18	4.81	5.72	5.98	7.22	4.73	**0.77**
FB	10.44	27.52	10.47	8.37	**13.82**	9.07	9.28	9.30	8.74	7.84	7.39	7.76	9.32	1.21
NSS Filter														
KR	**10.31**	**16.59**	**10.31**	**6.73**	20.38	**8.05**	**7.40**	**5.06**	**4.78**	**5.58**	**6.12**	**5.93**	**3.22**	0.78
NSS	19.32	26.62	19.34	11.12	38.96	13.23	9.51	6.92	6.53	7.55	8.04	9.86	7.14	3.45
GSW	24.28	27.06	24.31	12.33	48.69	16.21	9.46	7.04	6.29	7.19	7.76	9.93	7.23	3.48
LW	11.67	18.25	11.68	7.52	23.55	9.12	7.81	5.19	4.84	5.72	6.21	7.22	4.76	**0.77**
FB	11.18	27.79	11.21	8.73	**16.14**	9.49	9.62	9.31	8.79	7.84	7.76	7.83	9.33	1.21
Sample 6/1961-12/2020 (excluding FB)														
Full Data														
KR	**17.27**	**16.44**	**17.28**	**8.58**	**40.66**	**8.16**	**7.27**	**5.11**	**4.57**	**5.55**	**5.56**	**5.22**	**2.85**	0.85
NSS	28.36	27.83	28.38	13.42	63.72	13.54	9.33	6.91	6.25	7.45	7.38	8.89	7.02	3.67
GSW	32.24	28.02	32.27	14.36	71.04	16.15	9.28	7.07	6.06	7.19	7.13	8.96	7.05	3.67
LW	18.44	18.15	18.46	9.28	43.19	9.14	7.68	5.22	4.64	5.69	5.66	6.46	4.29	**0.81**
3-Month Filter														
KR	**7.25**	**17.75**	**7.27**	**5.02**		**8.16**	**7.27**	**5.11**	**4.57**	**5.55**	**5.56**	**5.22**	**2.85**	0.85
NSS	10.75	29.98	10.78	7.83		13.54	9.33	6.91	6.25	7.45	7.38	8.89	7.02	3.67
GSW	12.23	29.88	12.26	8.06		16.15	9.28	7.07	6.06	7.19	7.13	8.96	7.05	3.67
LW	7.87	19.65	7.89	5.51		9.14	7.68	5.22	4.64	5.69	5.66	6.46	4.29	**0.81**
KR Filter														
KR	**8.65**	**15.34**	**8.66**	**5.84**	**16.72**	**7.23**	**6.52**	**4.60**	**4.30**	**5.03**	**5.26**	**5.03**	**2.81**	0.85
NSS	20.74	26.98	20.76	11.00	43.20	12.79	8.48	6.35	5.93	6.86	7.03	8.69	6.98	3.67
GSW	24.70	27.10	24.73	11.95	50.72	15.50	8.45	6.46	5.71	6.54	6.73	8.73	7.01	3.67
LW	10.01	17.07	10.02	6.59	19.95	8.23	6.88	4.71	4.36	5.16	5.33	6.26	4.25	**0.81**
NSS Filter														
KR	**9.40**	**15.61**	**9.41**	**6.16**	**19.04**	**7.54**	**6.73**	**4.62**	**4.33**	**5.04**	**5.44**	**5.13**	**2.83**	0.85
NSS	17.79	27.06	17.80	10.40	36.73	12.91	8.66	6.36	5.96	6.85	7.23	8.67	6.99	3.67
GSW	22.12	27.20	22.15	11.46	45.33	15.54	8.61	6.47	5.75	6.54	6.97	8.73	7.01	3.67
LW	10.62	17.28	10.63	6.86	21.89	8.50	7.10	4.73	4.39	5.16	5.53	6.26	4.27	**0.81**

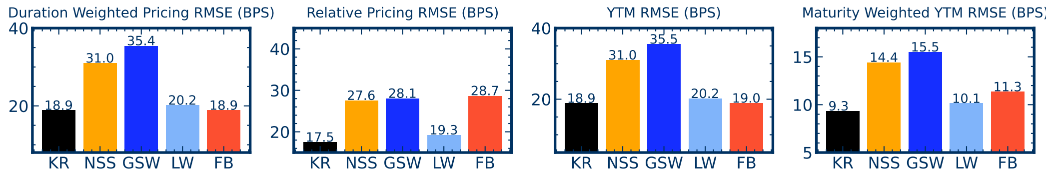
This table compares the in-sample pricing results for different methods. We report the duration weighted and relative pricing RMSE and average and maturity bucket weighted yield RMSE. In addition, we list the yield RMSE for ten different maturity buckets. The results are shown with and without filters for the short and the long sample. The best model performance is indicated in bold. All results are in basis points (BPS).

Table A.2: Out-of-sample pricing comparison

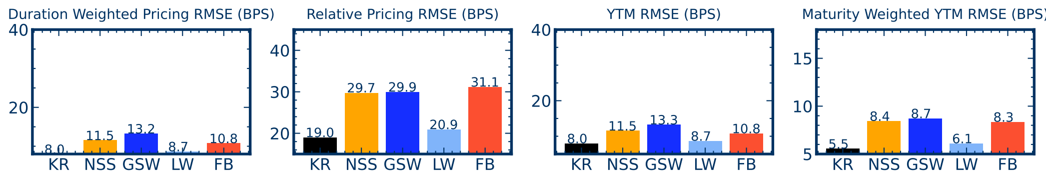
	Price Dur.-w.	Price Relati.	YTM Aveg.	YTM Mat.-w.	YTM <3M	YTM 3M to 1Y	YTM 1Y to 2Y	YTM 2Y to 3Y	YTM 3Y to 4Y	YTM 4Y to 5Y	YTM 5Y to 7Y	YTM 7Y to 10Y	YTM 10Y to 20Y	YTM >20Y
Sample 6/1961-12/2013 (including FB)														
Full Data														
KR	**17.55**	**29.53**	**17.56**	**10.95**	33.96	**10.24**	**10.23**	**8.41**	**8.05**	**8.98**	**9.03**	**8.87**	**6.85**	4.83
NSS	29.49	36.34	29.52	15.24	58.81	14.62	11.68	9.52	9.19	10.52	10.44	11.74	9.87	6.02
GSW	33.72	36.54	33.77	16.17	66.86	17.16	11.38	9.44	8.89	10.18	10.12	11.77	9.87	6.02
LW	18.70	30.89	18.72	11.60	36.53	11.11	10.55	8.51	8.09	9.09	9.14	10.04	8.09	**4.82**
FB	17.63	38.10	17.67	12.56	**29.52**	11.25	12.61	12.25	11.45	11.06	10.34	10.67	11.47	5.00
3-Month Filter														
KR	**10.25**	**31.82**	**10.27**	**8.39**		**10.24**	**10.23**	**8.41**	**8.05**	**8.98**	**9.03**	**8.87**	**6.85**	4.83
NSS	12.99	39.14	13.01	10.40		14.62	11.68	9.52	9.19	10.52	10.44	11.74	9.87	6.02
GSW	14.41	39.01	14.45	10.54		17.16	11.38	9.44	8.89	10.18	10.12	11.77	9.87	6.02
LW	10.79	33.33	10.81	8.83		11.11	10.55	8.51	8.09	9.09	9.14	10.04	8.09	**4.82**
FB	12.52	41.20	12.57	10.68		11.25	12.61	12.25	11.45	11.06	10.34	10.67	11.47	5.00
KR Filter														
KR	**11.25**	**28.57**	**11.25**	**8.91**	17.20	**9.46**	**9.48**	**7.84**	**7.74**	**8.36**	**8.73**	**8.69**	**6.74**	4.83
NSS	23.78	35.50	23.80	13.45	44.65	13.99	10.81	8.91	8.82	9.85	10.13	11.53	9.77	6.02
GSW	28.07	35.61	28.10	14.39	52.92	16.64	10.50	8.80	8.49	9.46	9.78	11.53	9.76	6.02
LW	12.51	29.95	12.52	9.60	20.25	10.34	9.75	7.94	7.78	8.47	8.84	9.87	7.99	**4.82**
FB	12.18	37.15	12.21	10.59	**14.24**	10.31	11.69	11.59	11.10	10.40	9.92	10.32	11.36	5.00
NSS Filter														
KR	**12.01**	**28.82**	**12.02**	**9.24**	19.57	**9.72**	**9.78**	**7.86**	**7.75**	**8.42**	**8.81**	**8.80**	**6.84**	4.83
NSS	20.55	35.53	20.57	12.83	38.00	14.10	11.12	8.92	8.83	9.87	10.18	11.44	9.77	6.02
GSW	25.33	35.65	25.36	13.89	47.57	16.66	10.83	8.80	8.50	9.49	9.85	11.45	9.77	6.02
LW	13.12	30.11	13.13	9.87	22.19	10.56	10.07	7.96	7.78	8.51	8.91	9.79	8.05	**4.82**
FB	12.92	37.32	12.95	10.93	**16.68**	10.67	12.07	11.59	11.10	10.45	10.09	10.33	11.30	5.00
Sample 6/1961-12/2020 (excluding FB)														
Full Data														
KR	**16.04**	**29.94**	**16.06**	**10.15**	**31.32**	**9.51**	**9.37**	**7.81**	**7.52**	**8.38**	**8.40**	**8.06**	**6.38**	4.79
NSS	27.01	37.90	27.03	14.31	54.83	14.13	10.71	8.80	8.54	9.78	9.69	10.66	9.65	6.27
GSW	30.71	37.87	30.75	15.12	61.95	16.38	10.45	8.74	8.28	9.48	9.41	10.68	9.60	6.24
LW	17.06	31.45	17.08	10.74	33.59	10.28	9.66	7.89	7.55	8.48	8.50	9.09	7.59	**4.78**
3-Month Filter														
KR	**9.49**	**32.11**	**9.50**	**7.80**		**9.51**	**9.37**	**7.81**	**7.52**	**8.38**	**8.40**	**8.06**	**6.38**	4.79
NSS	12.20	40.72	12.22	9.80		14.13	10.71	8.80	8.54	9.78	9.69	10.66	9.65	6.27
GSW	13.44	40.40	13.47	9.92		16.38	10.45	8.74	8.28	9.48	9.41	10.68	9.60	6.24
LW	9.98	33.81	10.00	8.20		10.28	9.66	7.89	7.55	8.48	8.50	9.09	7.59	**4.78**
KR Filter														
KR	**10.38**	**29.13**	**10.39**	**8.28**	**15.83**	**8.77**	**8.70**	**7.30**	**7.25**	**7.83**	**8.14**	**7.90**	**6.32**	4.79
NSS	21.90	37.25	21.91	12.70	42.10	13.55	9.94	8.26	8.22	9.18	9.42	10.49	9.59	6.27
GSW	25.64	37.16	25.67	13.53	49.40	15.90	9.67	8.17	7.93	8.84	9.12	10.48	9.53	6.24
LW	11.50	30.69	11.50	8.91	18.53	9.55	8.94	7.39	7.27	7.92	8.24	8.95	7.54	**4.78**
NSS Filter														
KR	**11.08**	**29.35**	**11.09**	**8.60**	**18.15**	**9.03**	**8.97**	**7.32**	**7.25**	**7.88**	**8.20**	**8.00**	**6.37**	4.79
NSS	18.98	37.11	18.99	12.10	35.84	13.66	10.22	8.27	8.23	9.20	9.46	10.40	9.46	6.27
GSW	23.16	37.02	23.19	13.04	44.31	15.92	9.97	8.17	7.93	8.87	9.18	10.40	9.41	6.24
LW	12.07	30.77	12.08	9.17	20.53	9.78	9.23	7.41	7.28	7.96	8.30	8.88	7.54	**4.78**
Cross-Sectional OOS														
KR	**8.87**	**27.49**	**8.89**	**6.73**		**9.18**	**8.63**	**6.06**	**5.55**	**6.69**	**7.15**	**8.18**	**7.69**	**1.48**
NSS	15.83	49.39	16.59	13.11		17.52	14.59	13.32	14.05	18.53	15.46	11.01	9.75	3.72
LW	9.28	28.97	9.30	7.07		10.10	8.77	6.30	5.76	6.91	7.20	8.50	8.26	1.82

This table compares the out-of-sample pricing results for different methods. We report the duration weighted and relative pricing RMSE and average and maturity bucket weighted yield RMSE. In addition, we list the yield RMSE for ten different maturity buckets. The results are shown with and without filters for the short and the long sample. The best model performance is indicated in bold. All results are in basis points (BPS). The main out-of-sample results are evaluated on the next business day after the estimation. We also include cross-sectional out-of-sample results based on 10-fold stratified sampling.

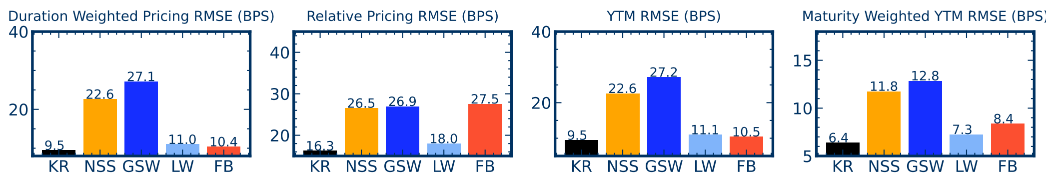
Figure A.6: In-sample results by evaluation metric for different filters



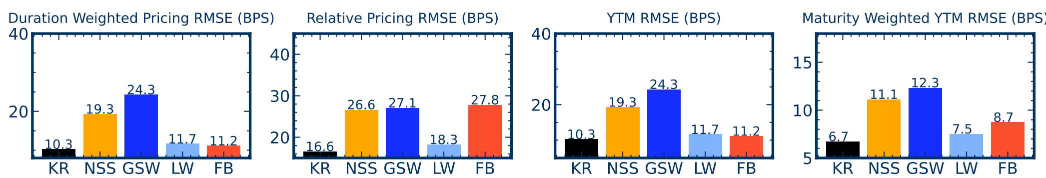
(a) Full data (no outlier removal)



(b) 3-Month Filter



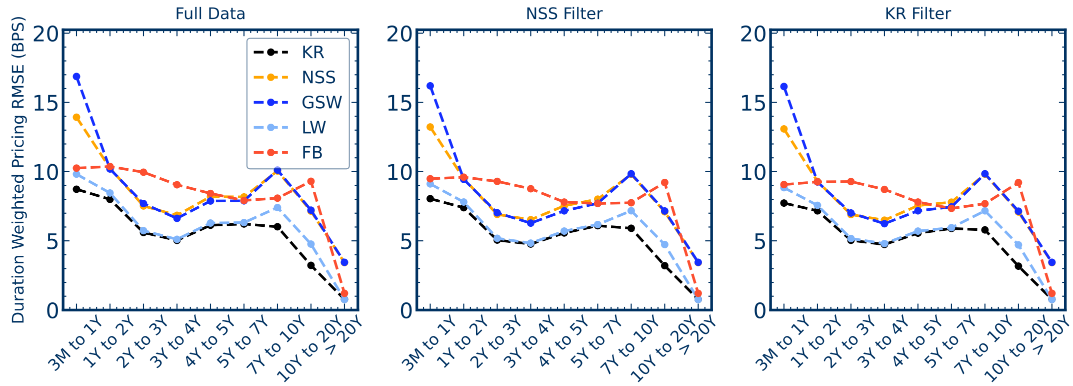
(c) KR Filter



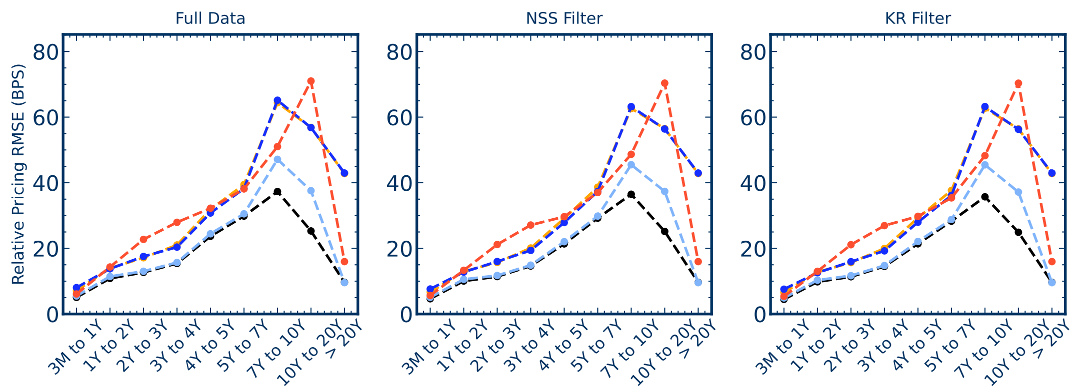
(d) NSS Filter

This plot shows aggregated evaluation metrics calculated in-sample on the sample from June 1961 to December 2013. Columns correspond to duration weighted pricing RMSE, relative pricing RMSE, YTM RMSE, and maturity-weighted YTM RMSE. All numbers are in basis points (BPS). The top panel correspond to results evaluated on the full data without filtering. The second panel shows results evaluated on data where securities maturing within three months are removed. In the third panel, results are evaluated on the sample for which an NSS filter is used to remove outlier securities, whose YTM fitting errors are at least three standard deviation away from the average YTM fitting error calculated using NSS curves in the same cross-section. The last panel is for results evaluated on the sample for which KR is used to remove outlier securities, and the rule is the same as that of the NSS filter. KR outperforms other methods in term of in-sample fitting quality according to all four evaluation metrics on datasets with and without outlier removal.

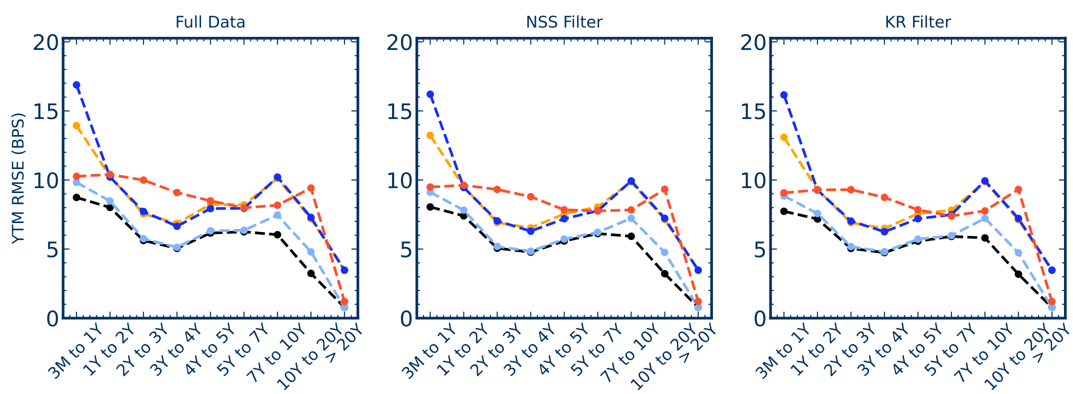
Figure A.7: In-sample pricing errors for different maturities and filters



(a) Duration Weighted Pricing RMSE



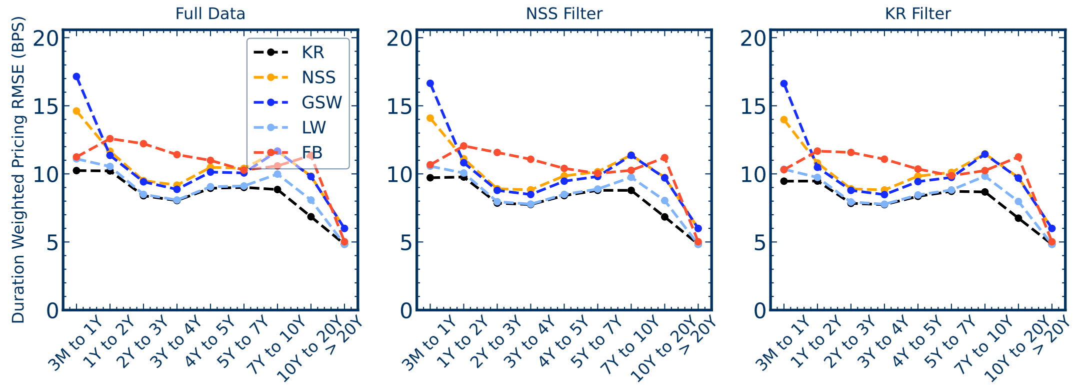
(b) Relative Pricing RMSE



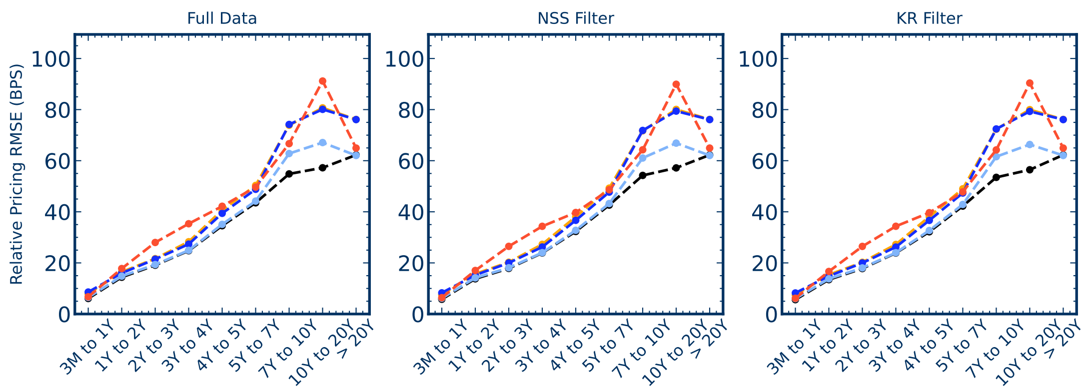
(c) YTM RMSE

This plots shows evaluation metrics calculated in-sample on the sample from June 1961 to December 2013. The top panel shows duration weighted pricing RMSE, the mid panel pricing RMSE, and the bottom panel the yield RMSE. All errors are in basis points (BPS). The first column corresponds to results evaluated on the full data without further filtering. The mid column shows the results evaluated on the sample for which an NSS filter is used to remove outlier securities, whose YTM fitting errors are at least three standard deviation away from the average cross-sectional YTM fitting error on the same day. The right columns shows the results evaluated on the sample for which outliers are removed with a three standard deviation filter based on KR estimates.

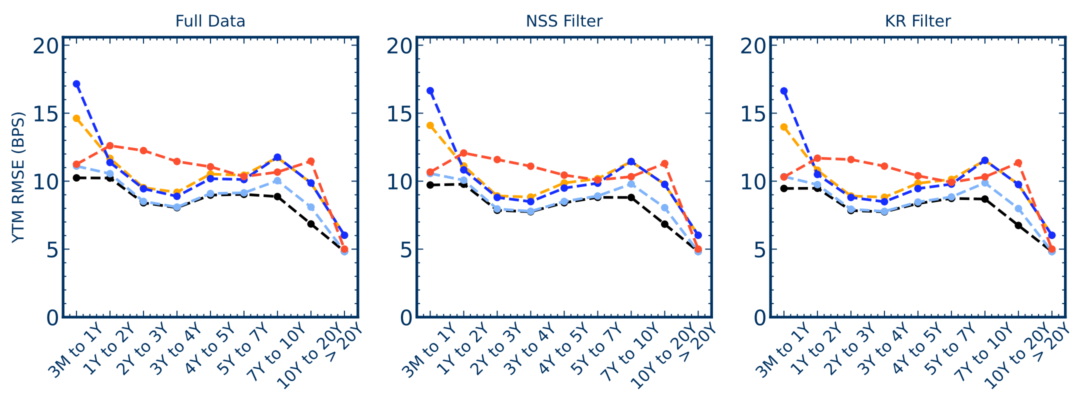
Figure A.8: Out-of-sample pricing errors for different maturities and filters



(a) Duration Weighted Pricing RMSE



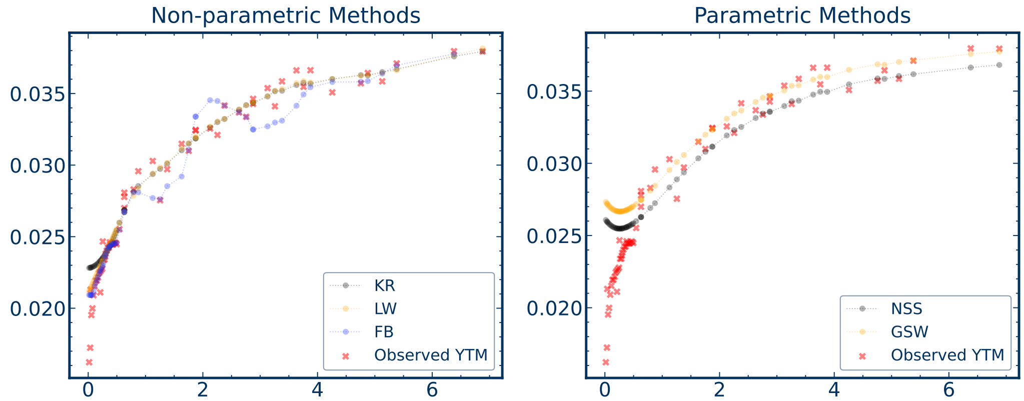
(b) Relative Pricing RMSE



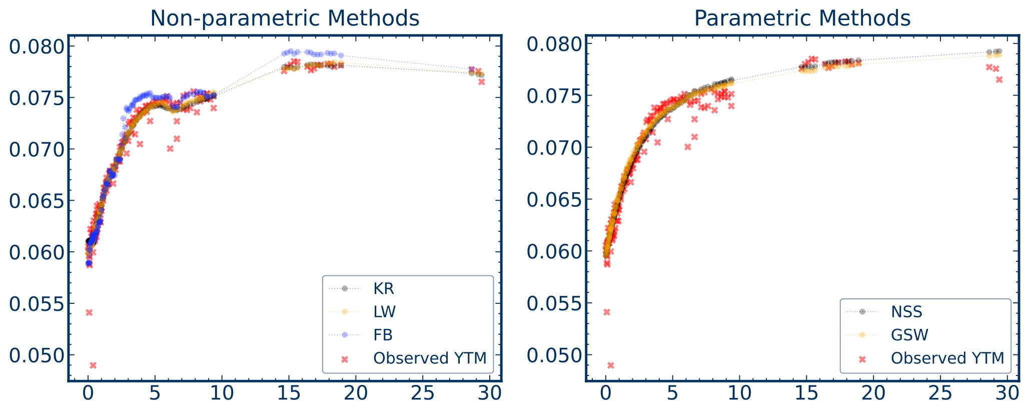
(c) YTM RMSE

This plots shows evaluation metrics calculated out-of-sample on the sample from June 1961 to December 2013. Out-of-sample errors are calculated using curves estimated at time t to price securities observed on the next business day. The top panel shows duration weighted pricing RMSE, the mid panel pricing RMSE, and the bottom panel the yield RMSE. All errors are in basis points (BPS). The first column corresponds to results evaluated on the full data without further filtering. The mid column shows the results evaluated on the sample for which an NSS filter is used to remove outlier securities, whose YTM fitting errors are at least three standard deviation away from the average cross-sectional YTM fitting error on the same day. The right columns shows the results evaluated on the sample for which outliers are removed with a three standard deviation filter based on KR estimates.

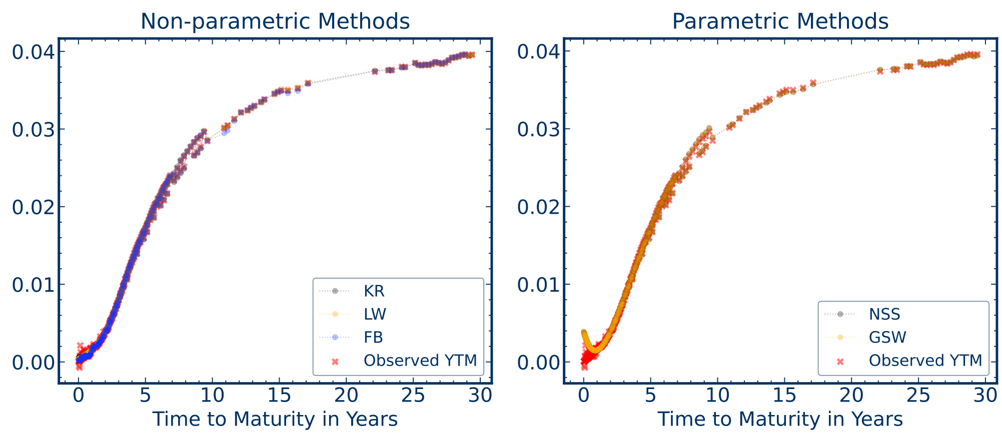
Figure A.9: Fitted yields of coupon bonds



(a) 1961-06-30



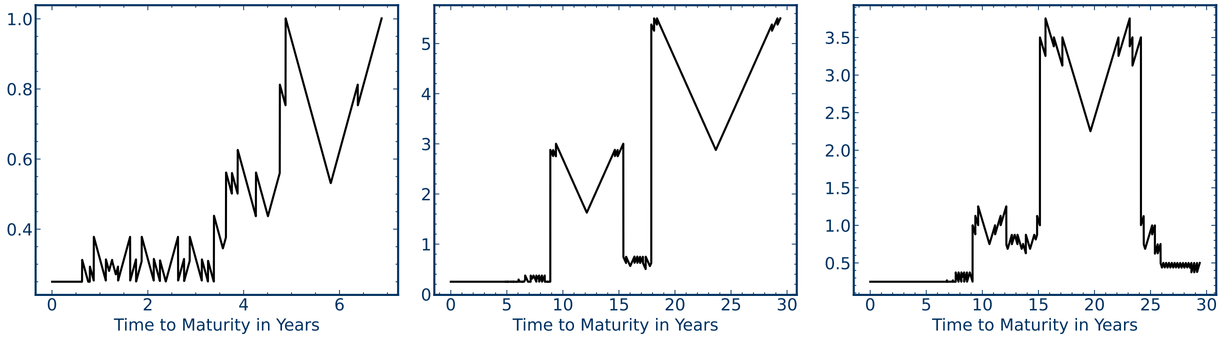
(b) 1986-06-30



(c) 2013-12-31

This figure shows the fitted YTM by different methods for the representative example dates: 1961-06-30 (top panel), 1986-06-30 (mid panel), and 2013-12-31 (bottom panel). The observed YTM is calculated using observed prices, which is compared against YTM given by model-implied prices. The left and right columns are separated by non-parametric and parametric methods.

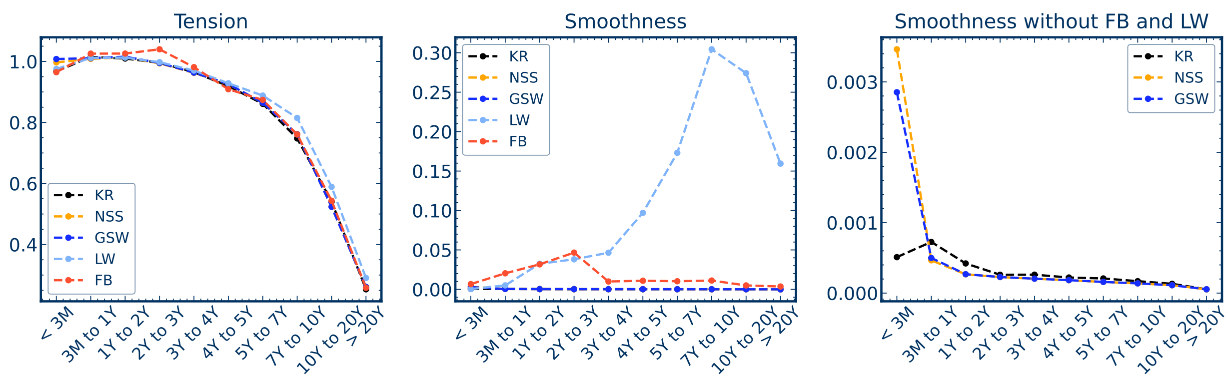
Figure A.10: LW bandwidth estimates



(a) 1961-06-30 **(b)** 1986-06-30 **(c)** 2013-12-31

This figure shows the LW bandwidth estimated for the example dates 1961-06-30, 1986-06-30, and 2013-12-31.

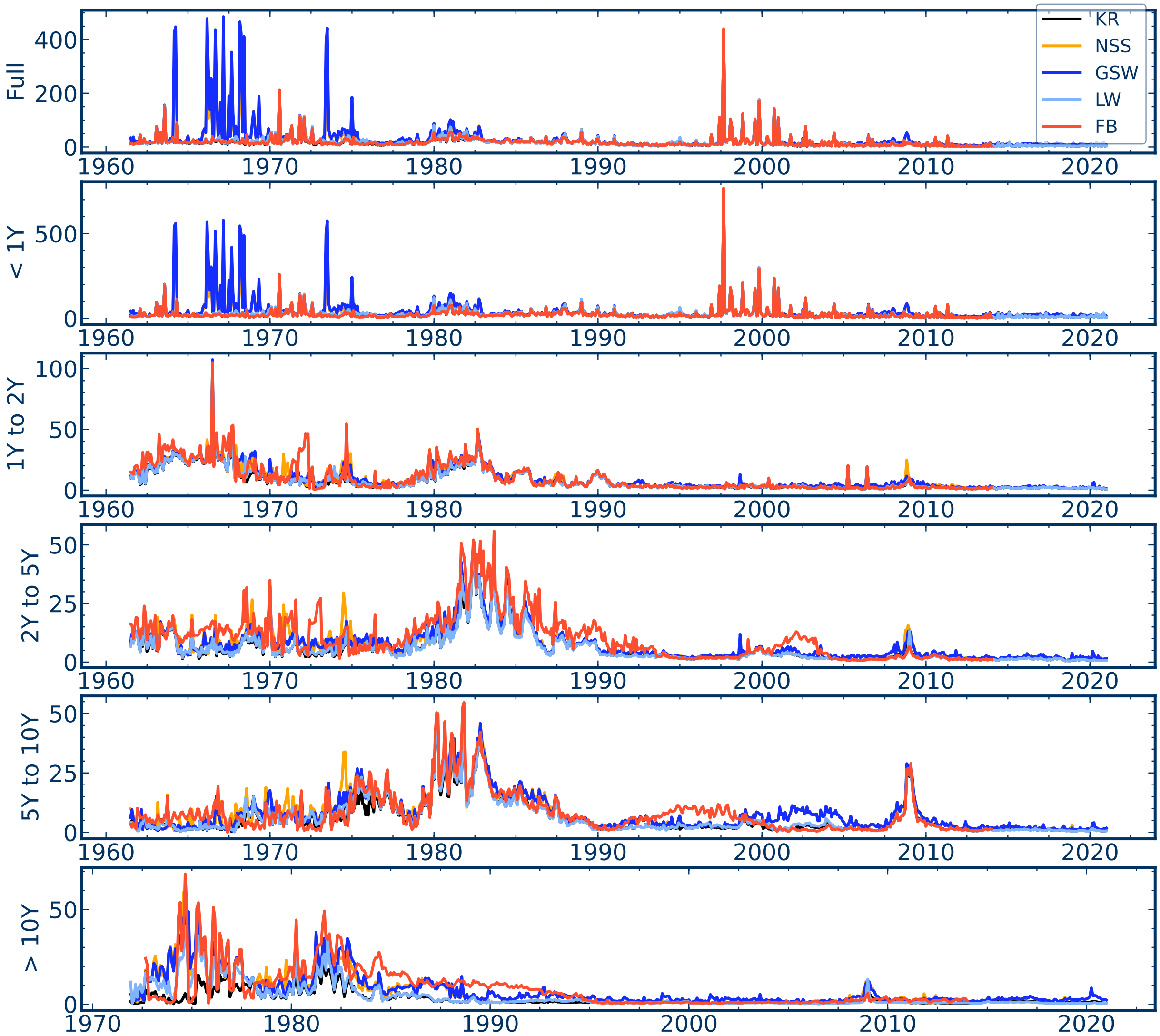
Figure A.11: Tension and curvature for different maturities with daily granularity



This plot shows the discretized measures for tension (left panel) and curvature (mid panel) for different maturity ranges. The right panel shows the curvature measure for KR, NSS, and GSW only. The discrete derivatives use daily granularity. Results are calculated on the sample from June 1961 to December 2013.

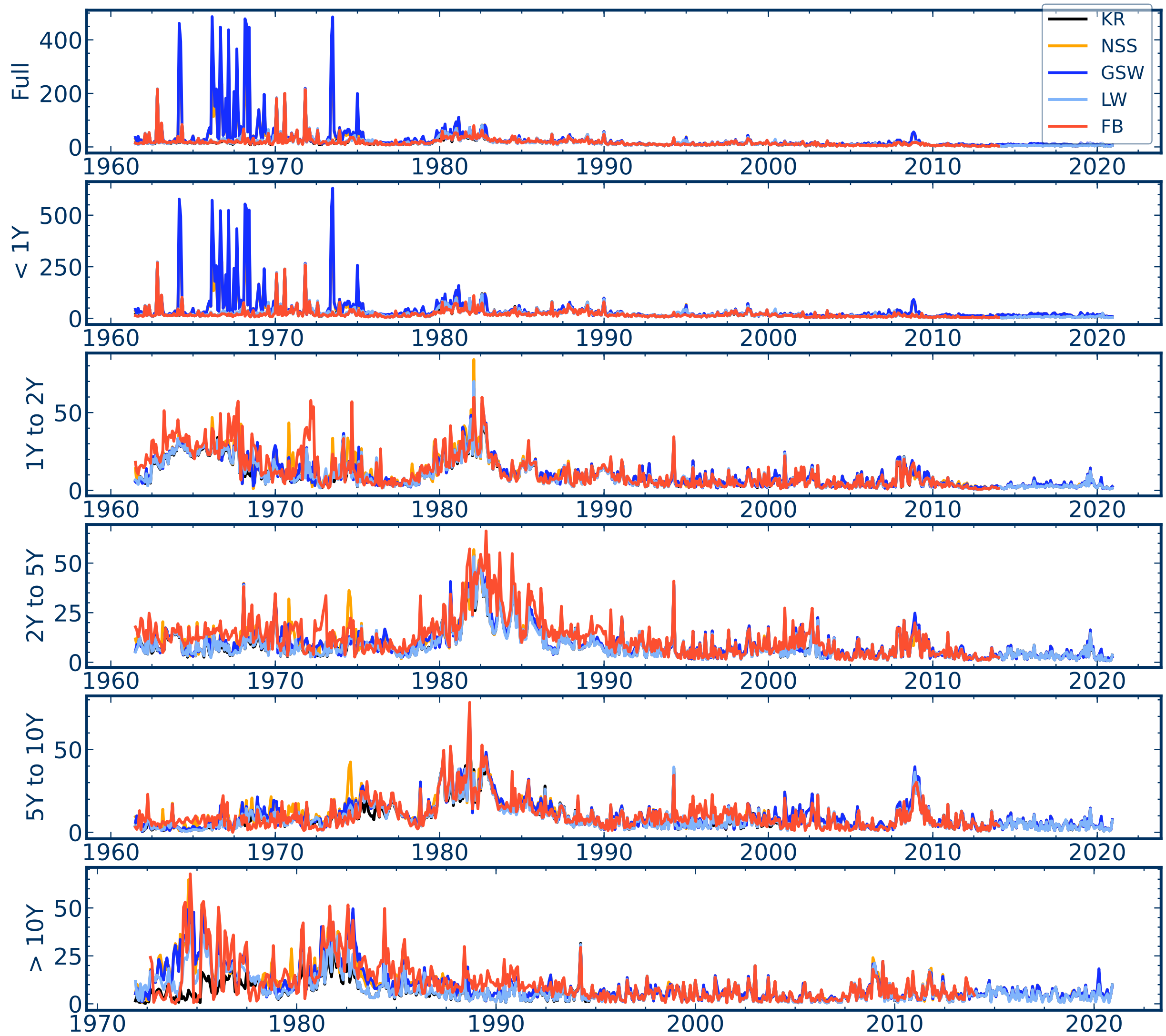
B.3 Results over Time

Figure A.12: In-sample yield fitting errors over time



The figure shows the in-sample YTM RMSE (BPS) over time for each last day of the month for five maturity buckets.

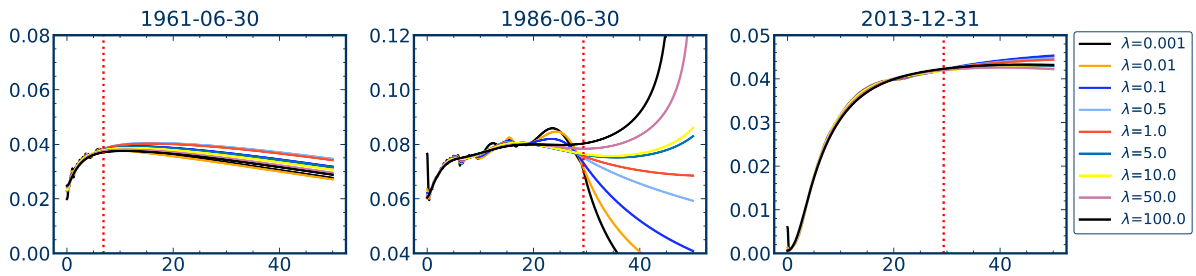
Figure A.13: Out-of-sample yield fitting error over time



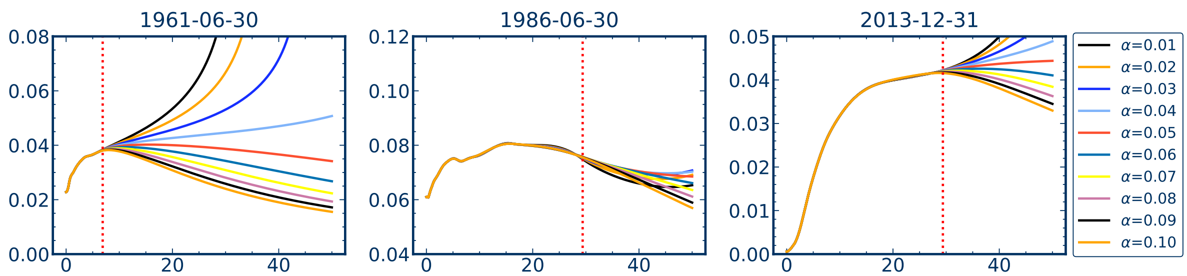
The figure shows the out-of-sample YTM RMSE (BPS) over time for each last day of the month for five maturity buckets. Out-of-sample YTM fitting errors are calculated using curves estimated at time t to price securities on the next business day $t+1$

B.4 Extrapolation

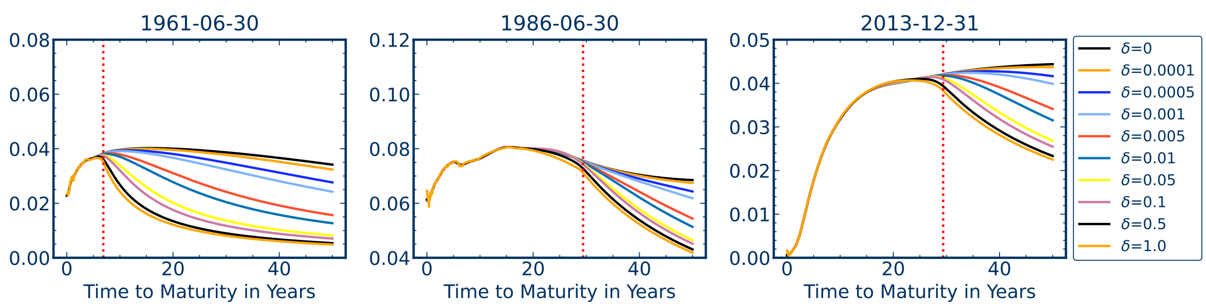
Figure A.14: Extrapolated KR yield curves



(a) Varying λ, fixed $\alpha = 0.05$ and $\delta = 0$



(b) Varying α, fixed $\lambda = 1$ and $\delta = 0$

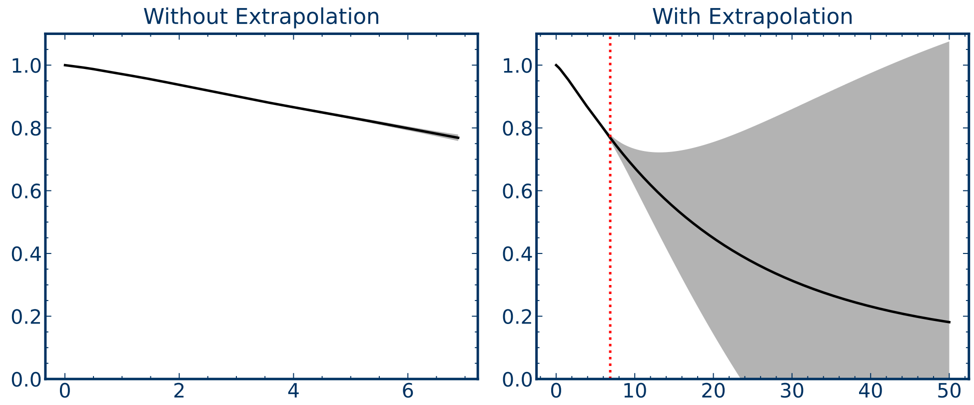


(c) Varying δ, fixed $\alpha = 0.05$ and $\lambda = 1$

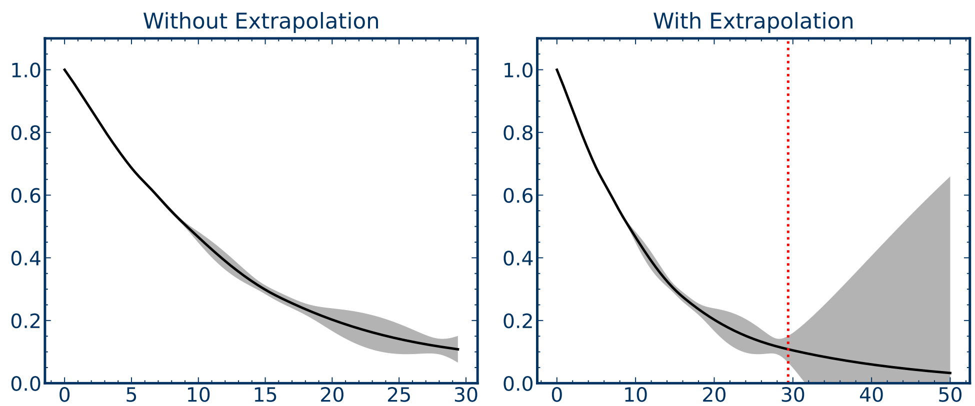
This figure shows yield curve estimates with extrapolation to 50-year maturity for KR as a function of parameters on the three representative example days 1961-06-30 (left column), 1986-06-30 (mid column), and 2013-12-31 (right column). The region to the right of the red dashed vertical line is the extrapolation region. Subfigure (a) varies the smoothness parameter λ for fixed values $\alpha = 0.05$ and $\delta = 0$. Subfigure (b) varies the maturity weight α for fixed values $\lambda = 1$ and $\delta = 0$. Subfigure (c) varies the tension parameter δ for fixed values $\lambda = 1$ and $\alpha = 0.05$.

B.5 Statistical Inference

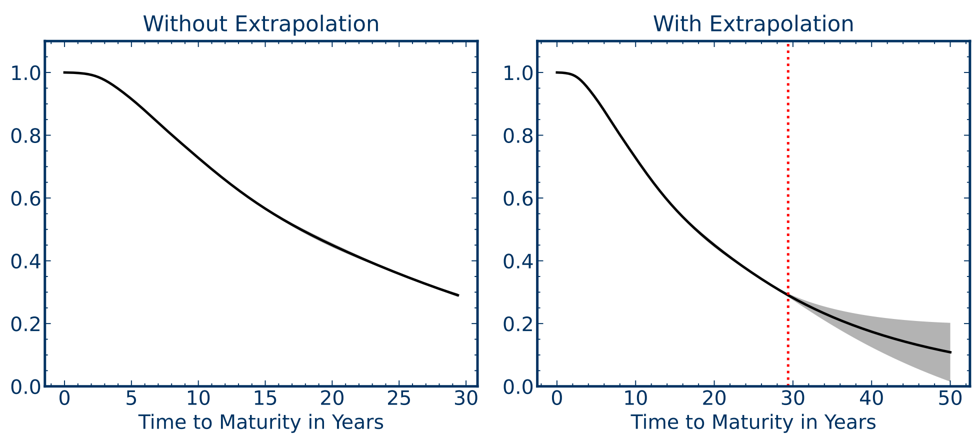
Figure A.15: KR discount curve confidence bands



(a) 1961-06-30



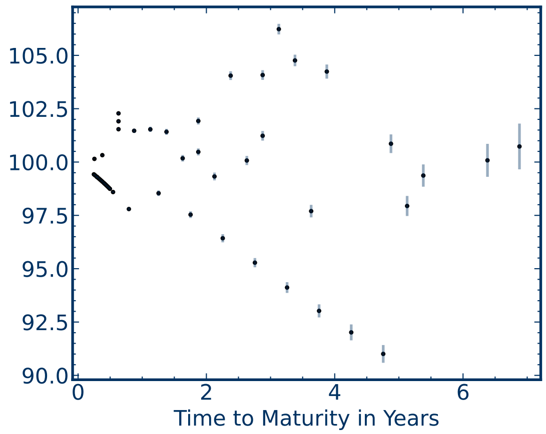
(b) 1986-06-30



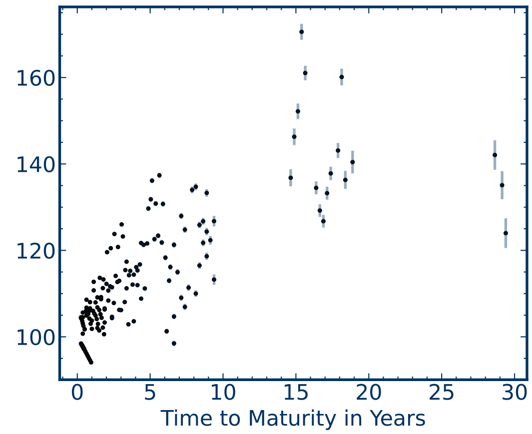
(c) 2013-12-31

The figure shows 3-standard-deviation confidence bands for discount curve estimates given by the KR model under the Gaussian process assumption. The panels correspond to the representative example dates 1961-06-30, 1986-06-30, and 2013-12-31. The left column shows results without extrapolation, and the right column includes extrapolation results for up to 50-year maturity.

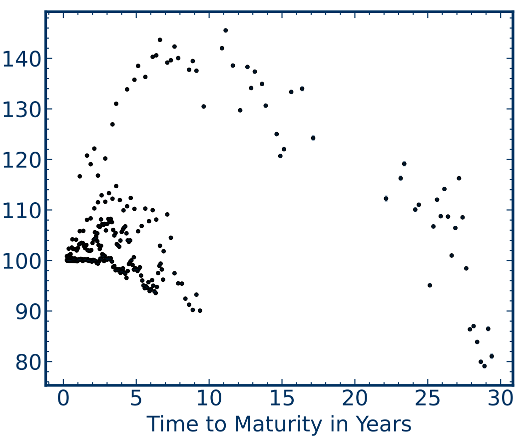
Figure A.16: KR confidence bands for prices



(a) 1961-06-30



(b) 1986-06-30

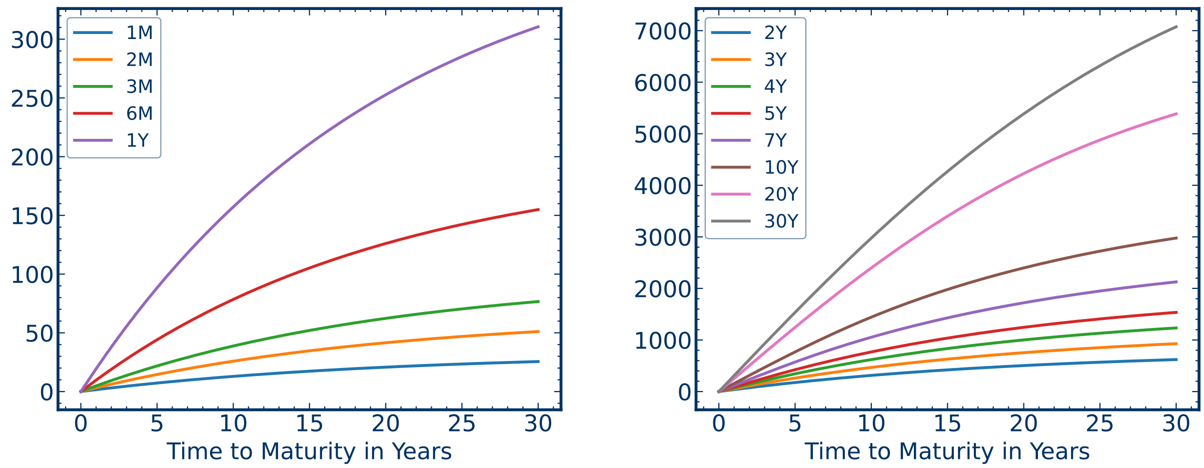


(c) 2013-12-31

The figure shows 3-standard-deviation confidence bands for fitted prices given by the KR model under the Gaussian process assumption. The panels correspond to the three representative example dates 1961-06-30, 1986-06-30, and 2013-12-31.

B.6 Basis Functions

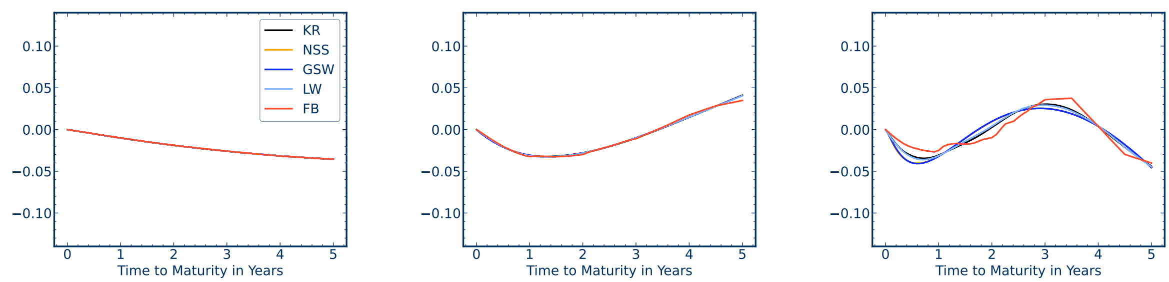
Figure A.17: KR kernel function



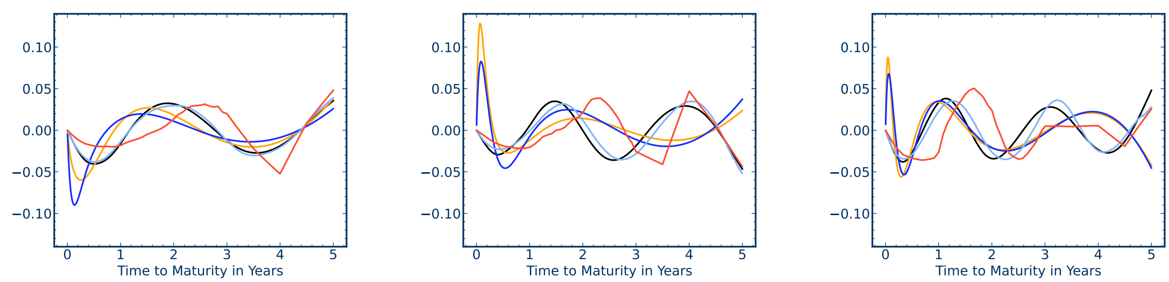
(a) $k(\cdot, x)$ for x no greater than one year **(b)** $k(\cdot, x)$ for x greater than one year

This figure visualizes the kernel function $k(\cdot, x)$ for different maturities x. It corresponds to the column vectors of the KR kernel matrix $\boldsymbol{K}_{ij} = k(x_i, x_j)$ for selected columns for the baseline model $\lambda = 1$, $\alpha = 0.05$ and $\delta = 0$.

Figure A.18: Principal Component Analysis of panel of discount bonds up to 5-year maturity



(a) First PC **(b)** Second PC **(c)** Third PC



(d) Fourth PC **(e)** Fifth PC **(f)** Sixth PC

This figure shows the first 6 principal components (PCs) estimated from the panel of discount curves for the five methods KR, GSW, NSS, LW, and FB. The PCs correspond to the eigenvectors of the largest eigenvalues of the covariance matrix of discount bond prices. The panel consists of the estimated discount bond prices up to 5-year maturity for the sample from June 1961 to December 2013.

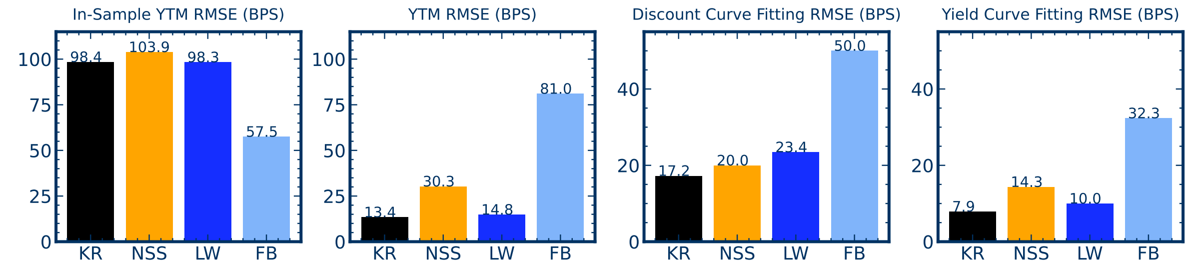
C Simulation

Our simulation confirms the main empirical finding that the KR method provides the best out-of-sample fit. For brevity, we include only an illustrative simulation, but the same finding holds for a variety of alternative simulation designs. We set the ground truth discount curve equal to the estimated KR and NSS discount curve estimated on the example day 2013-12-31. Hence, we have two empirically relevant discount curves based on a non-parametric and parametric model. The bond prices are observed with random noise. In more detail, we keep the maturity distribution of observed securities on the example day, and we assume that all securities are zero-coupon bonds for the ease of implementing FB. For each ground-truth discount curve, we obtain 10 sets of simulated noisy prices by adding independent Gaussian noise with mean zero and standard deviation one to the implied ground-truth prices, which are normalized to 100. We estimate the discount curve with the KR, NSS, LW, and FB method on each set of simulated noisy prices.

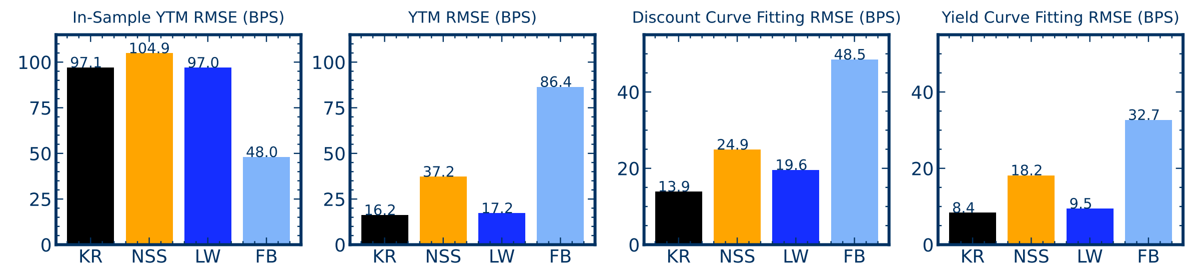
We report four evaluation metrics for the two different discount curves. The first metric is in-sample YTM RMSE, which measures how well the yield of the noisy bond prices is fitted. The second metric, YTM RMSE, can be interpreted as the out-of-sample yield error and evaluates the yield error implied by the noiseless prices. By comparing the first two metrics, we can evaluate the tendency of a method to overfit to noise in the data. The last two metrics are discount and yield curve fitting RMSE, which measure how well a method estimates the overall ground-truth discount and yield curves. Both can be interpreted as out-of-sample evaluations and measure the errors for the complete maturity spectrum. The results are averaged across the 10 simulation runs.

Figure A.19 reports the four evaluation metrics when either KR or NSS is the ground-truth discount curve. Comparing the first two panels, which show YTM RMSE evaluated on the noisy and the noiseless price data, we observe that while FB has the lowest in-sample YTM RMSE, it has the highest YTM RMSE evaluated on the noiseless data. This implies that FB overfits to noise in data. On the other hand, KR has the lowest out-of-sample RMSE, when comparing the estimated models with the yields for noiseless bonds and the ground-truth yield and discount curves. Hence, KR dominates the other methods in terms of capturing the underlying discount curve and robustness to noise.

Figure A.19: Simulation results



(a) KR discount curve



(b) NSS discount curve

This figure shows the simulation results when the ground-truth discount curve is set to the KR (top panel) or NSS (bottom panel) estimates on 2013-12-31. We keep the maturity distribution of observed securities on the example date, and we assume that all securities are zero-coupon bonds. We obtain 10 sets of simulated noisy prices by adding independent Gaussian noise with mean zero and standard deviation one to the implied ground-truth prices that are normalized to 100. We estimate KR, NSS, LW, and FB from each set of noisy prices, and report the four evaluation metrics: (1) In-sample YTM RMSE, which measures the yield error relative to noisy prices, (2) YTM RMSE, which corresponds to an out-of-sample evaluation of the yield error relative to noiseless prices, (3) discount curve fitting RMSE, which measures the error between the estimated and ground-truth discount curve, (4) yield curve fitting RMSE, which measures the error between the estimated and ground-truth yield curve. All metrics are reported in basis points. The results are averaged across 10 simulation runs.

Swiss Finance Institute

Swiss Finance Institute (SFI) is the national center for fundamental research, doctoral training, knowledge exchange, and continuing education in the fields of banking and finance. SFI's mission is to grow knowledge capital for the Swiss financial marketplace. Created in 2006 as a public–private partnership, SFI is a common initiative of the Swiss finance industry, leading Swiss universities, and the Swiss Confederation.

swiss:finance:institute

c/o University of Geneva, Bd. Du Pont d'Arve 42, CH-1211 Geneva 4
T +41 22 379 84 71, rps@sfi.ch, www.sfi.ch