Shrinking the Term Structure*

Damir Filipović† Markus Pelger‡ Ye Ye§

This draft: August 6, 2024
First draft: September 23, 2022

Abstract

We propose a new framework to explain the factor structure in the full cross section of Treasury bond returns. Our method unifies non-parametric curve estimation with cross-sectional factor modeling. We identify smoothness as a fundamental principle of the term structure of returns. Our approach implies investable factors, which correspond to the optimal spanning basis functions in decreasing order of smoothness. Our factors explain the slope and curvature shapes frequently encountered in PCA. In a comprehensive empirical study, we show that the first four factors explain the time-series variation and risk premia of the term structure of excess returns. Cash flows are covariances as the exposure of bonds to factors is fully explained by cash flow information. We identify a state-dependent complexity premium. The fourth factor, which captures complex shapes of the term structure premium, substantially reduces pricing errors and pays off during recessions.

Keywords: Term structure of interest rates, bond returns, factor space, U.S. Treasury securities, non-parametric method, principal components, machine learning in finance, reproducing kernel Hilbert space
JEL classification: C14, C38, C55, G12

*We thank Robert Anderson, Michael Bauer, Xiaohong Chen, Richard Crump, Darrell Duffie, Peter Feldhütter, Kay Giesecke, Lisa Goldberg, Nikolay Gospodinov, Tze Lai, David Lando, Martin Lettau, Philippe Mueller (discussant), Dino Palazzo, Aaron Sidford, Anders Trolle, Irina Zviadadze (discussant), and seminar and conference participants at Stanford, UC Berkeley, Copenhagen Business School, University of Cincinnati, Korean Advanced Institute of Science & Technology, Korean Business School, Balyasny Asset Management, SFS Cavalcade, Annual Meeting of the European Finance Association, Conference on Computational and Financial Econometrics for helpful comments. 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 and NBER, Email: mpelger@stanford.edu.
§Stanford University, Department of Management Science & Engineering, Email: yeye@stanford.edu.

1 Introduction

The term structure of interest rates is a fundamental economic quantity for financial and macroeconomic research and applications. The term structure is spanned by the discount bonds of all maturities, that is default-free bonds that pay one dollar in the future for given maturities. Discount bonds are the basis assets that can replicate any default-free fixed income claim. This paper addresses the question of which risk factors explain the full cross-section of discount bond returns. This problem is challenging as the discount bonds and the underlying factors are unobserved and must be inferred from a relatively sparse set of noisy returns of Treasury bonds.[1] Existing approaches first use an ad-hoc method to estimate a curve of discount bond prices, and in a second step construct asset pricing models on a subset of the implied discount bond returns.

Our paper unifies term structure curve estimation with cross-sectional asset pricing. Term structure factor modeling is the joint problem of estimating the discount bond return curve and cross-sectional factor modeling. A low dimensional factor structure in a panel of Treasury bond returns is equivalent to a sparse set of basis functions that span the discount bond return curve. This dual perspective is novel and important for two reasons. First, a precise return curve is needed to discover all relevant asset pricing factors, and we demonstrate that popular return curve estimates are biased and distort asset pricing factors. Second, by formally showing the connection between curve fitting and factor modeling, we can explain the structure of the cross-sectional factors and why they arise. The optimal sparse selection of basis functions that span the discount bond return curve are the asset pricing factors that price the cross-section of Treasury bonds.

We develop a robust, flexible, and easy-to-implement method to estimate a low dimensional factor model for Treasury bonds. By the duality of the problem, our method can be interpreted as either non-parametric curve estimation or conditional latent factor modeling in an unbalanced panel. Importantly, our simple closed-form solution for the discount bond return curve is linear in a sparse set of optimal basis functions, which establishes the link to factor modeling, and implies tradable discount bonds and factors. The estimated discount bond return curve is given in closed form as a simple cross-sectional ridge regression of Treasury bond returns on their discounted cash flows. The ridge penalty rewards smoothness of the discount bond return curve. An additional lasso penalty selects the optimal sparse set of basis functions, which maps into a low dimensional factor representation. Our approach imposes minimal assumptions on the true underlying discount bond return curve, using only the core elements that define the estimation problem. We trade off return fitting against an economically motivated smoothness reward of the discount bond return curve, which is sufficient to uniquely determine the optimal basis functions that span the discount bond return curve. In contrast to existing methods, we do not pre-select any potentially misspecified functional form or ad-hoc non-parametric basis functions.

Our extensive empirical study on U.S. Treasury bonds provides four main findings. First, we identify smoothness as a fundamental principle of the term structure of returns and explain

[1] We use Treasury bond as synonym for Treasury security. These Treasury bonds are generally coupon bonds (except for short maturities) and are only available for few maturities.

the familiar slope and curvature shapes of PCA factors. Discount bond return curves that span Treasury bonds out-of-sample are smooth curves. The optimal unique basis functions that span the return curve are the smoothest basis functions in decreasing order. These basis functions capture exactly the slope and curvature type shapes, which can be consequently estimated with a PCA applied to a panel of discount bonds. In other words, slope and curvature factors arise naturally from a PCA applied to basis assets when the underlying problem is fitting a smooth functional relationship. Importantly, our cross-sectional factors do not only provide insight on the structure of PCAs but they are also applicable if we only observe a cross-section on a single day and, thus, cannot use a PCA estimator.

Second, the exposure of Treasury bonds to risk factors is fully explained by cash flow information. We provide a conditional cross-sectional factor model for Treasury bonds, and show how to construct bond characteristics. We show empirically that the cross-sectional loadings to risk factors based on cash flow information are identical to time-series regression loadings based on covariances with factors. In this sense, "cash flows are covariances" for Treasury bonds, which is similar to the observation that "characteristics are covariances" for equities. Hence, we do not require a time-series to obtain factor realizations and factor loadings, but it is sufficient to have cross-sectional cash flow information available on a given day. An important difference to equity modeling is that in our framework there are no ad-hoc assumptions on the functional transformation of the characteristics, but the law of one price and the optimal basis functions determine the Treasury loadings.

Third, four factors fully explain Treasury excess returns and the term structure premium. We "shrink" the term structure and show that the cross-sectional dimensionality is represented by four tradable factor portfolios. These factors are sufficient to replicate and price the excess return of any default-free fixed income claim. The term structure premium curve has a complex shape. Our fourth risk factor represents a basis function that approximates complex shapes, and is essential for spanning the term structure premium. Hence, we interpret the premium of the fourth factor as a complexity premium. It carries a large risk premium, which lifts the annualized out-of-sample Sharpe ratio of the implied stochastic discount factor projected on Treasuries to 0.85, while it lowers the pricing errors by a multiple of three. This important factor requires a precise term structure estimation in the first place, and it is ignored in discount bond panels based on an overly simplistic form like the parametric Nelson–Siegel–Svensson model.

Fourth, we document a time-varying complexity premium, which is informative about economic conditions. The complex fourth term structure factor is a hedge for bad economic times and pays off during recessions. It can be used in trading strategies to hedge against recession risk, and ignoring the fourth factor results in substantially worse investment performance during recessions. The exposure to term structure risk factors provides a measure for the state of the bond market. Whereas the factors in Treasury markets are stable over time, the cross-sectional variation that is explained by the factors is time-varying and can be informative about economic conditions. We introduce Treasury Market Complexity (T-COM) to measure the complexity of the bond market.

It captures how much variation is explained by higher order term structure factors. The bond market complexity is positively correlated with future unemployment rates and strongly related to the term structure risk premium. Remarkably, the first (slope) factor earns its risk premium during times when a linear curve explains bond returns well, while its risk premium turns negative during complex market conditions. In contrast, our fourth factor serves as a hedge against the risk of complex market conditions. It has a negative payoff during low T-COM periods, but earns a high positive risk premium during challenging times.

Our work also provides new conceptual perspectives on bond markets. First, we connect asset pricing among different asset classes. We show that the familiar concept of cross-sectional regressions commonly used in equities also arises naturally in the fixed income space, and that both, Treasury bonds and equities can be studied with similar tools and interpreted in similar ways after identifying the analogy between cash flows and firm characteristics. Our factor model can be viewed as a Barra model for Treasury bonds. In cross-sectional factor models such as Barra, the fundamental risk factors are obtained from a cross-sectional regression of stock returns on their firm characteristics. Similarly, our method implies fundamental factors through a cross-sectional regression on appropriately transformed cash flows. In contrast to Barra and most cross-sectional equity regression models, our method is not ad-hoc, but based on the economic first principles of the law of one price and the smoothness of the discount bond return curve. Second, this paper is a step further into unifying different methodological perspectives. So far, the literature has largely separated the two problems of 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 connects these two problems. Third, we show the importance of and how to use shrinkage in the term structure space. The smoothness of the discount curve, which follows from limits to excessive payoffs, naturally implies ridge and lasso shrinkage. Our novel methodology shows how a successful machine learning solution needs to be tailored to an economic problem.

We provide publicly available and regularly updated data sets of daily discount bond returns and factors based on our precise estimates.[2] This is the only fixed income data, where the discount bonds and factors are feasible portfolios of actually traded U.S. Treasury bonds. Our shared data also includes the complexity and pricing error measures over time. It is part of our larger discount bond database, which also includes the precise and robust yield estimates of discount bonds and the corresponding code to replicate our methods.

Related Literature

Our work connects fixed income modeling with recent advances in equity modeling. Our work shares insights with the Instrumented-PCA of Kelly, Pruitt, and Su (2019). We develop a cross-sectional conditional factor model for Treasury bonds, where the risk factor loadings are instrumented by the discounted cash flows of the coupon bonds. In other words, the cash flows take the role of firm characteristics. In contrast to equity modeling, instrumenting betas with cash flows is

[2]Our discount bond database is available at `https://www.discount-bond-data.org`.

an exact relationship grounded in the law of one price. Our work also draws on ideas similar to the Risk-Premium PCA of Lettau and Pelger (2020a,b). Our and their papers include an economically motivated penalty in the objective function to estimate latent factors. Our paper leverages the structure of fixed income modeling, and the penalty is not targeting first moments but the smoothness of the discount bond return curve. Our findings also relate to weak factors. Our fourth factor captures a high risk premium but explains less variation. Hence, similar to Lettau and Pelger (2020a,b), we confirm that weak factors, that explain less variation, can be crucial in explaining the first moment in the cross-section. As suggested by the title, our work also relates to "Shrinking the Cross-Section" by Kozak, Nagel, and Santosh (2020). Their paper constructs an SDF for equities with shrinkage methods. They use lasso shrinkage to select a sparse set of factors and ridge shrinkage to lower the weights of higher order principal components in the SDF. Their solution can be formulated as an elastic net regression of the second on the first moments of rotated equity returns. Our solution is also an elastic net regression. Both papers share a lasso penalty to select an optimal set of factors, and a ridge penalty to downweight factors by imposing limits on payoffs of near arbitrage opportunities. However, our solution is based on fundamentally different arguments, and the ridge regression arises to enforce the smoothness of the discount bond return curve.

This paper is complementary to Filipović, Pelger, and Ye (2022) and the literature on non-parametric yield curve estimation. Both papers introduce the novel reproducing kernel Hilbert space framework to estimate general discount bond curves. There are three important differences between the two papers. First, Filipović, Pelger, and Ye (2022) estimates discount bond price curves, while this paper estimates return curves. Second, this paper studies the question whether a sparse subset of all basis functions implied by the reproducing kernel is sufficient for the curve estimation. Third, and most importantly, this paper establishes the connection with cross-sectional asset pricing and factor modeling, and provides a comprehensive empirical study of the asset pricing implications of the resulting factors. We find that only four basis functions and term structure factors, respectively, are needed to approximate the true underlying discount bond excess return curve. Filipović, Pelger, and Ye (2022) provides one of the most extensive comparison studies between different yield curve estimation methods, which includes their own kernel ridge (KR) method, Fama and Bliss (1987), Liu and Wu (2021), Gürkaynak, Sack, and Wright (2007), and their own implementation of the Nelson–Siegel–Svensson model. They find that their KR estimator outperforms the benchmarks in all dimensions, and provides the most precise and robust yield and price estimates. Given the results in Filipović, Pelger, and Ye (2022), there is no need to repeat a similar horse race of yield curve estimators in this paper. Hence, the focus of this paper is on the comparison between a low dimensional factor KR model and the full KR model, which represents the best possible fit for discount curves.

Naturally, our work overlaps with Crump and Gospodinov (2022) and the debate on the number of term structure factors. Both papers highlight that the number of factors for a panel of discount bonds depends on the metric and the objective that is studied and that the first three PCA factors,

commonly referred to as level, slope and curvature, might not be sufficient to represent the cross-section of bond returns. Crump and Gospodinov (2022) assume a potential factor structure for a given panel of discount bond returns, and they show how the eigenvalues and eigenvectors are affected by the mechanical cross-sectional overlapping structure. They argue that the explained variation in excess returns might not be informative about the underlying number of factors. We confirm their finding, and argue that selecting the number of asset pricing factors should also be based on other asset pricing metrics. We show that Sharpe ratio and pricing error metrics suggest two factors in addition to level, slope and curvature. We also show that these two additional factors are crucial for explaining out-of-sample excess returns of the traded Treasury bonds, instead of a panel of discount bonds. The level, slope and curvature are the first three factors, because they are the optimal first three basis functions for approximating smooth functions. However, they are not sufficient, as the discount bond return curve is more complex than a second order polynomial type function. We need four factors in excess returns, which implies five factors in the returns, to adequately capture the functional form and moments of the cross-section. Crump and Gospodinov (2022) prove that the shape of the loadings of principal components is structural for problems when the cross-section has a mechanical overlapping dependency structure. We provide further insights into why the loadings of PCA factors take the form of polynomial functions with increasing degrees. Our paper provides an ordering of the optimal sparse non-parametric basis functions for discount bond excess return curves. The key element is that these curves are smooth. Crump and Gospodinov (2022) interpret smoothness as local cross-sectional correlations. We use a functional analytic perspective. The fact that good models for the term structure are sufficiently smooth, and that polynomial type basis functions provide the best approximations for this function space, explains why the observed factor structure arises in the first place.[3]

Our paper is complementary to the literature on bond price and yield forecasting. The forecasting literature often begins by assuming a low dimensional representation of the cross-section of Treasury bonds, typically in the form of a panel of factors or selected discount bonds. The future yields or returns of this low dimensional representation are predicted using prior information, including possibly macroeconomic variables, which allows to test important economic hypotheses like the spanning or expectation hypothesis. Notable contributions in this area include Cochrane and Piazzesi (2005), Cooper and Priestley (2008), Ludvigson and Ng (2009, 2010), Duffee (2013), Cieslak and Povala (2015), Greenwood and Vayanos (2014), Bauer and Hamilton (2018), and many others. Our cross-sectional factor model provides a low dimensional representation of the cross-

[3]The relationship between smooth curve estimation and slope and curvature type factors is a universal finding. Previous literature including Litterman and Scheinkman (1991), Cochrane and Piazzesi (2005), Crump and Gospodinov (2022) and Filipović, Pelger, and Ye (2022) have documented factor structures with these familiar shapes for PCA applied to panels of discount bond prices or yields. We study the returns of Treasuries as our paper focuses on the risk premium and investment implication of factors. Our method maps a non-parametric curve estimation into a linear factor model, where the factors are portfolios of traded Treasury bond returns. Importantly, the same cross-sectional mapping also applies to panels of discount bond prices or yields. As shown in Filipović, Pelger, and Ye (2022), the same optimal basis functions (up to a level shift) that explain discount bond returns also span discount bond prices or yields, because fundamentally all these problems represent smooth curves. By construction, if we work in the yield or price space the estimator maps into portfolios of discount bond prices or yields instead of returns.

section of Treasury bonds, which can be used as a forecasting target to study conditional time-series moments. In other words, as any default-free fixed income claim can be replicated as a portfolio of our factors, modeling the time-series properties of our factors is sufficient to describe the time-series of any fixed income assets.

Our paper is also complementary to the literature on dynamic no-arbitrage term structure models. In fact, the most common affine term structure models, studied in Duffie, Pan, and Singleton (2000), Dai and Singleton (2002) and Duffie, Filipović, and Schachermayer (2003), among others, and alternative stochastic factor models such as in Filipović, Larsson, and Trolle (2017), assume a specific low dimensional stochastic factor process that generates the bond data. These factors are not directly observable and need to be filtered from data, usually in conjunction with a statistical estimation method such as quasi-maximum likelihood. They are not linearly related to observable bond prices and thus do not represent tradable portfolios.[4] Estimating the parameters of such models can therefore be challenging and sensitive to the choice of data and estimation methods. In addition, these models often assume stationarity, which makes them inflexible, so they do not always fit well the yield or return curves, especially in times of market stress or structural changes in the economy. In contrast, our paper solves a non-parametric sequential cross-sectional curve fitting and dimension reduction problem, and as such is maximally flexible. As our cross-sectional estimator is linear in the return space, it maps into tradable portfolios, instead of abstract stochastic drivers, which facilitates the asset pricing and bond investment analysis.[5] As a result, our paper provides new insights into the cross-sectional relationship between Treasury bonds, which serves as a basis for studying the time-series properties of fixed income assets.

This paper is organized as follows. In Section 2, we formulate the term structure estimation problem and introduce the shrinkage solution and sparse factor model. In Section 3, we conduct a comprehensive empirical analysis of the low dimensional factor structure of the U.S. Treasury market. In Section 4 we study the cross-sectional asset pricing and investment implications of the factors. Section 5 concludes. The appendix collects all the formal theoretical statements, proofs, and additional empirical results. Supporting results are deferred to the Internet Appendix.

2 Term Structure Estimation and Cross-Sectional Asset Pricing

We provide a formal framework for bond asset pricing and show that it is intimately intertwined with term structure estimation. Our framework unifies non-parametric curve estimation with cross-sectional asset pricing. This allows us to estimate and explain the factors that explain the cross-section of bond returns.

[4] Zero-coupon yields are linear in the factors of an affine term structure model. However, the former are not directly observable and certainly not tradable.

[5] There is a similar complementarity in the literature on equity modeling. One stream of that literature studies the cross-sectional relationship between stocks in a linear reduced form using factors (either statistical or fundamental factors, but always in the context of cross-sectional regressions). This is analogous to what we are doing. Then another stream of literature models the dynamics of the stochastic discount factor in some parametric or reduced form (e.g., the ICAPM literature or many macro-finance models). This second type often performs worse in cross-sectional asset pricing, but has the advantage that it can give testable implications about certain economic structures.

2.1 Framework

Discount bonds, also known as zero-coupon bonds, of all maturities are the basis assets that span any default-free fixed income claim. Hence, asset pricing factors that price all discount bonds can price all default-free fixed income securities. As the prices and returns of discount bonds are not directly observed, they must be estimated from the noisy observations of a relatively sparse set of traded Treasury bonds. These Treasury bonds are generally coupon bonds (except for short maturities) and are only available for few maturities.

Formally, on a given business day t, we observe M_t Treasury bonds i with prices $P_{t,i}$. Each bond i is characterized by its future cash flows $C_{t,ij}$ at x_j time units after day t, for a common grid of cash flow dates $0 = x_0 < x_1 < \cdots < x_N$.[6] The non-zero cash flows for Treasury bonds are the semi-annual coupon payments and their principal repayment. The law of one price then connects the fundamental value of the observed bonds with the unobserved discount bonds through

$$P_{t,i} = \sum_{j=1}^{N} C_{t,ij}\, d_t(x_j), \tag{1}$$

for the *discount curve* $d_t : [0, \infty) \to \mathbb{R}$, where $d_t(x)$ denotes the price of the discount bond with time to maturity x and face value one. In other words, any Treasury bond can be represented as a portfolio of discount bonds, where the portfolio weights are given by the cash flows of the bond. Under the absence of arbitrage this portfolio has the same price as the bond.

A similar no-arbitrage identity holds between bond returns. To see this, we express the relationship (1) with the cash flow matrix from the prior business day $t-1$ by shifting the cash flow index j, that is, $P_{t,i} = \sum_{j=1}^{N} C_{t-1,ij+1}\, d_t(x_j)$. The gross return of bond i from $t-1$ to t, including its cash flow at t, is then related to discount bond gross returns through

$$\frac{P_{t,i} + C_{t-1,i1}}{P_{t-1,i}} = \frac{\sum_{j=0}^{N} C_{t-1,ij+1} d_t(x_j)}{P_{t-1,i}} = \sum_{j=0}^{N} Z_{t-1,ij} \frac{d_t(x_j)}{d_{t-1}(x_j + \Delta_t)}, \tag{2}$$

where Δ_t denotes the time units between business days $t-1$ and t, and $Z_{t-1,ij} \coloneqq \frac{C_{t-1,ij+1} d_{t-1}(x_j+\Delta_t)}{P_{t-1,i}}$ denote the (normalized) discounted cash flows of bond i.[7]

Given the asset pricing focus of this paper, the objects of interest are the bond excess returns $R^{\text{bond}}_{t,i} \coloneqq \frac{P_{t,i}+C_{t-1,i1}}{P_{t-1,i}} - \frac{1}{d_{t-1}(\Delta_t)}$ over the risk-free one-business day return, which is given by the gross return $\frac{1}{d_{t-1}(\Delta_t)}$ of the discount bond maturing at t. As $\sum_{j=0}^{N} Z_{t-1,ij} = 1$, we obtain from (2)

[6] We consider daily maturities for the actual/365 day count convention. That is, the cash flow dates are of the form $x_j = j/365$. This can easily be adapted to any other time unit convention.

[7] These formal expressions hold under the assumption that $t-1$ and t are consecutive calendar days. In general, we replace $C_{t-1,ij+1}$ by $C_{t-1,ij+k_t}$, where $k_t \geq 1$ denotes the number of calendar days between the business days $t-1$ and t. For regular consecutive days, we have $k_t = 1$. If a weekend or other bank holidays fall between $t-1$ and t, then $k_t \geq 2$. Accordingly, we set $\Delta_t = k_t/365$ for the time units between $t-1$ and t.

the no-arbitrage identity $R^{\text{bond}}_{t,i} = \sum_{j=1}^{N} Z_{t-1,ij} r_t(x_j)$, for the *return curve* $r_t : [0, \infty) \to \mathbb{R}$, where

$$r_t(x) = \frac{d_t(x)}{d_{t-1}(x + \Delta_t)} - \frac{1}{d_{t-1}(\Delta_t)} \tag{3}$$

denotes the excess return from $t-1$ to t of the discount bond with time to maturity x. In matrix notation, this no-arbitrage identity reads

$$R^{\text{bond}}_t = Z_{t-1} R_t \tag{4}$$

for the M_t-vector of bond excess returns $R^{\text{bond}}_t \coloneqq (R^{\text{bond}}_{t,1}, \ldots, R^{\text{bond}}_{t,M_t})^\top$, the $M_t \times N$-matrix of discounted cash flows $Z_{t-1} \coloneqq (Z_{t-1,ij})_{i=1,\ldots,M_t,\, j=1,\ldots,N}$, and the N-vector of discount bond excess returns

$$R_t \coloneqq (r_t(x_1), \ldots, r_t(x_N))^\top, \tag{5}$$

which is the return curve r_t evaluated at dates $x_1, \ldots, x_N$.[8] Hence, the excess returns of Treasury bonds are represented as portfolio of discount bond excess returns, where the portfolio weights are cash flow instruments.

Any asset pricing model for Treasury bonds requires first an estimate of the return curve, which in turn substantially affects the asset pricing implications. In other words, cross-sectional bond asset pricing and return curve estimation are always a joint problem. There are two reasons why it is important to connect the curve estimation and cross-sectional factor modeling for panels of discount bonds. First, a precise return curve, like our method, is needed to discover all relevant asset pricing factors, and we demonstrate that popular return curve estimates are biased and distort asset pricing factors. Second, by formally showing the connection between curve fitting and factor modeling, we can explain the structure of the cross-sectional factors.

2.2 Return Curve Estimation

In the literature so far, the discount curves d_{t-1} and d_t were first estimated using parametric or non-parametric methods in order to then derive the implied return curve from (3). In contrast, we directly estimate the curve r_t that can best explain the observed coupon bond returns. This has the advantage that the estimated discount bonds, and in turn the factors, become traded portfolios of Treasury bonds.

Due to market imperfections, the observed bond excess returns R^{bond}_t deviate from their fundamental values (4), resulting in return errors ϵ_t,

$$R^{\text{bond}}_t = Z_{t-1} R_t + \epsilon_t.$$

Hence, a naive starting point to estimate the unobserved return curve would be to minimize the

[8] The excess return of the discount bond maturing on the next business day is zero, $r_t(0) = 0$, by definition. Therefore, we omit the components $r_t(x_j)$ and $Z_{t-1,ij}$ for $j = 0$ in the specification of the vector R_t and matrix Z_{t-1}.

mean squared return errors of the observed coupon bonds for each day t,[9]

$$\min_{r_t} \left\| R_t^{\text{bond}} - Z_{t-1} R_t \right\|_2^2.$$

However, the number of observed bond returns M_t is substantially smaller than the number of cash flow dates N. On a typical trading day we observe around $M_t = 300$ different Treasury bond returns, while the return curve for 10 years needs estimates for around $N \approx 3{,}650$ discount bond returns. Thus, any estimation approach needs to impose regularizing assumptions to reduce the dimensionality of the parameter space. All feasible estimators restrict the discount curves, and hence also return curves, either in terms of their functional form or their smoothness properties. The most common methods impose some ad-hoc parametric form on the discount curve, which Filipović, Pelger, and Ye (2022) show to be empirically misspecified and hence lead to spurious errors. More flexible methods impose either explicitly or implicitly some form of regularization for the discount curve.

This raises two questions. First, what is the class of plausible functions for the return 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. Smoothness of the return curve is founded in economic principles. In fact, smooth return curves imply that bonds with similar cash flows and maturities have similar returns. Large sudden changes along the return curve imply trading strategies in bonds with nearby maturities with extreme payoffs, such as butterfly returns $r_t(x - \Delta_t) - 2r_t(x) + r_t(x + \Delta_t) \approx r_t''(x)\Delta_t^2$. Hence, the degree of smoothness sets limits on excessive returns that are infeasible out-of-sample.

To answer the first question, we therefore consider the extremely large class of functions that are twice weakly differentiable. In fact, Filipović, Pelger, and Ye (2022) show that any arbitrage-free discount curve $d_t(x)$ is twice weakly differentiable in x, subject to mild technical first and second moments conditions on the short rates and its drift. Hence, we can assume that the return curve $r_t(x)$ is twice weakly differentiable.

To answer the second question, we regularize the problem by selecting the optimal level of smoothness. A natural measure of the smoothness of a twice weakly differentiable function h on $[0, \infty)$ is given by the weighted average of its second derivative

$$\|h\|_{\mathcal{H}_\alpha}^2 := \int_0^\infty h''(x)^2 \mathrm{e}^{\alpha x}\, dx. \tag{6}$$

Penalizing the second derivative avoids kinks and minimizes the curvature, while the function $\mathrm{e}^{\alpha x}$ with maturity weight $\alpha > 0$ allows the smoothness to be maturity dependent.[10] Accordingly, we consider the infinite-dimensional space $\mathcal{H}_\alpha$ consisting of all twice weakly differentiable functions

[9] For any vector $v = (v_j)$, we denote by $\|v\|_2 = (\sum_j |v_j|^2)^{1/2}$ its Euclidean ℓ^2-norm, and by $\|v\|_1 = \sum_j |v_j|$ its ℓ^1-norm.

[10] A more general smoothness measure could also include the first derivative, but Filipović, Pelger, and Ye (2022) show that the second derivative measured with this metric is the most relevant for term structure fitting.

$h : [0, \infty) \to \mathbb{R}$ with boundary conditions $h(0) = 0$ and $\lim_{x\to\infty} h'(x) = 0$ and with finite smoothness measure (6). Technically speaking, $\mathcal{H}_\alpha$ is a reproducing kernel Hilbert space (RKHS), which we formally discuss in the appendix.

The following proposition states that it is essentially without loss of generality to study the extremely large space of return curves in $\mathcal{H}_\alpha$. This means that we include all economically relevant curves, while being fully flexible without restrictive parametric or other ad-hoc assumptions. The appendix provides the formal statement of the following proposition with its proof.

Proposition 1
All arbitrage-free return curves are elements in $\mathcal{H}_\alpha$ under mild technical assumptions.

In summary, the fundamental problem of estimating the return curve is solved by trading off the fitting error against a smoothness reward. We formally express this by the objective function

$$\min_{r_t \in \mathcal{H}_\alpha} \left\{ \underbrace{\frac{1}{M_t} \left\| R_t^{\text{bond}} - Z_{t-1} R_t \right\|_2^2}_{\text{return error}} + \underbrace{\lambda \left\| r_t \right\|_{\mathcal{H}_\alpha}^2}_{\text{smoothness}} \right\} \tag{7}$$

for the smoothness parameter $\lambda > 0$, and where the vector of discount bond excess returns R_t is the evaluation of the return curve r_t at the cash flow dates as in (5). This problem is completely determined up to the two parameters α and λ, which we select empirically via cross-validation to minimize return errors out-of-sample. Hence, our framework is transparent in the choices we make, and at the same time our estimated curves are flexible and data-driven to obtain the most precise out-of-sample fit.

2.3 A Linear Factor Model Solution

We provide a remarkably simple closed-form solution to the non-parametric curve estimation problem (7). Importantly, the solution is linear, which establishes the link to factor modeling, and implies tradable discount bonds and factors. In more detail, our solution boils down to a kernel ridge regression, which is straightforward to implement. The solution builds on insights in functional analysis and machine learning by leveraging the structure of $\mathcal{H}_\alpha$ being a RKHS. Such spaces are particularly important in machine learning because of the celebrated representer theorem, which states that every function in a RKHS that minimizes an empirical 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 (return error and smoothness measure) and space of functions (twice differentiable, $\mathcal{H}_\alpha$), the representer theorem uniquely pins down the basis functions to solve the non-parametric problem. As the solution is linear in these basis functions, it provides the formal link between factor modeling and non-parametric curve fitting.

We do not impose any ad-hoc basis functions. Instead, the basis functions, which provide a linear solution with the desired properties, are implied by the reproducing kernel. This reproducing

kernel $k : [0,\infty) \times [0,\infty) \to \mathbb{R}$ of $\mathcal{H}_\alpha$ is given in closed form in the appendix. The kernel ridge (KR) solution curve $\hat{r}_t(\cdot)$ to (7) is unique in $\mathcal{H}_\alpha$ and given as linear combination of the N kernel basis functions $k(\cdot, x_1), \ldots, k(\cdot, x_N)$. The coefficients of this linear combination are linear portfolios of traded coupon bond excess returns, which are given in closed form in terms of the $N \times N$ kernel matrix $\boldsymbol{K} \coloneqq (k(x_i, x_j))_{1 \le i,j \le N}$. We define our full set of factors as the rotated coefficients after an orthonormalization of the kernel basis functions. This rotation is desirable for interpreting the factors, but without loss of generality. Thereto, we consider the singular value decomposition of the symmetric kernel matrix

$$\boldsymbol{K} = VSV^\top \tag{8}$$

with the diagonal matrix $S = \operatorname{diag}(s_1, \ldots, s_N)$ of eigenvalues in descending order $s_1 \ge \cdots \ge s_N > 0$ and the corresponding orthonormal eigenvectors V such that $V^\top V = I_N$, which denotes the $N \times N$ identity matrix.[11] These vectors V are the normalized basis functions implied by the reproducing kernel.

The basis functions V have special properties. In fact, their order corresponds to their smoothness. Hence, the first vector in V is the smoothest vector in $\mathbb{R}^N$, whereas the second vector is the second smoothest vector orthogonal to the first one, and so on. We provide the formal function analytical statement in the appendix and summarize it with the following proposition.

Proposition 2
The basis functions V implied by the kernel are orthonormal vectors ordered by their smoothness.

The following theorem is our main result. It states that the fundamental curve estimation problem (7) is equivalent to a finite-dimensional ridge regression problem, which can be viewed as a conditional full factor model. The proof is given in the appendix.

Theorem 1 (Conditional Full Factor Model Representation)
The vector of discount bond excess returns $\hat{R}_t = (\hat{r}_t(x_1), \ldots, \hat{r}_t(x_N))^\top$, where $\hat{r}_t$ denotes the solution to (7), can be represented as the full factor model

$$\hat{R}_t = \beta \hat{F}_t, \quad \text{with } \beta = VS^{1/2}, \tag{9}$$

where $\hat{F}_t$ denotes the unique solution to the cross-sectional ridge regression problem

$$\min_{F_t \in \mathbb{R}^N} \left\{ \frac{1}{M_t} \left\| R_t^{\text{bond}} - \beta_{t-1}^{\text{bond}} F_t \right\|_2^2 + \lambda \left\| F_t \right\|_2^2 \right\}, \tag{10}$$

where the conditional loadings $\beta_{t-1}^{\text{bond}}$ are given in terms of the discounted cash flows (bond characteristics) Z_{t-1} by

$$\beta_{t-1}^{\text{bond}} = Z_{t-1}\beta. \tag{11}$$

[11] We show in Lemma A.1 in the appendix that the kernel matrix $\boldsymbol{K}$ is strictly positive definite, so that all eigenvalues are strictly positive, $s_N > 0$.

The full factors $\hat{F}_t$ are given in closed form by

$$\hat{F}_t = \omega_{t-1} R_t^{\text{bond}}, \tag{12}$$

which are the excess returns of traded bond portfolios with portfolio weights[12]

$$\omega_{t-1} = \left(\beta_{t-1}^{\text{bond}\top} \beta_{t-1}^{\text{bond}} + \lambda M_t I_N\right)^{-1} \beta_{t-1}^{\text{bond}\top}. \tag{13}$$

Our solution expresses a general curve fitting problem as a simple ridge regression. The ridge regression penalizes higher basis functions of V, which as we show are the less smooth functions. Most importantly, the linearity of our solution implies tradability. The factors are investable portfolios of traded Treasury bonds. The discount bonds are portfolios of factors and thus also investable portfolios. Hence, our solution allows to replicate and hedge any default-free fixed income security.

The full factor model (9) is an unconditional factor model for discount bonds as it implies constant factor loadings. It implies a conditional full factor model for the coupon bond excess returns $\hat{R}_t^{\text{bond}} = Z_{t-1}\beta\hat{F}_t = \beta_{t-1}^{\text{bond}}\hat{F}_t$, where the time-varying loadings are conditional on the bond characteristics. The full factors can also be represented as portfolios of discount bonds with portfolio weights given in terms of the discount bond loadings:

$$\hat{F}_t = (\beta^\top\beta)^{-1}\beta^\top\hat{R}_t = S^{-1/2}V^\top\hat{R}_t.$$

Note that $\beta^\top\beta = S$ is a diagonal matrix, and hence up to a scaling normalization the columns of β (and V) are the factor portfolio weights in terms of discount bonds.

One of our empirical main contributions is to explain the structure of PCA factors estimated from discount bond panels. As we will show, the principal components extracted from a precisely estimated panel of discount bond excess returns coincide with the basis functions V. We show formally in the appendix that this relationship is not mechanical, but evidence that smoothness is a fundamental principle of Treasury bond returns. We thus provide, for the first time in the literature, empirical evidence that smoothness, as captured by our measure, is a fundamental principle of Treasury bond returns.

So far, our representation is not a low dimensional factor model. The factors of the full model correspond to the number of basis functions, which equals N. In the next section, we discuss how to optimally select a parsimonious low-rank factor model.

[12]Expression (13) contains the inverse of a $N \times N$ matrix. As N may be large, for computations one may thus prefer the equivalent expression $\omega_{t-1} = \beta_{t-1}^{\text{bond}\top}(\beta_{t-1}^{\text{bond}}\beta_{t-1}^{\text{bond}\top} + \lambda M_t I_{M_t})^{-1}$, which requires only the inversion of a $M_t \times M_t$ matrix.

2.4 Shrinking the Term Structure

A low-rank factor model corresponds to a sparse selection of the basis functions in Theorem 1. The challenge is to find an optimal selection. As we will see, the optimal selection contains the factors associated with the n largest eigenvalues of the kernel matrix (8). A priori, a researcher can select an arbitrary subset of n vectors from the N orthogonal eigenvectors V. The researcher then restricts the optimization problem in Theorem 1 to search over the function space that is spanned by the subset of basis functions corresponding to the selected eigenvectors V. The resulting sparse loadings $\beta_{t-1}^{(n)} \in \mathbb{R}^{M_t \times n}$ are given by the n columns of $\beta_{t-1}^{\text{bond}}$ that correspond to the selected indices. The regularized factors, which solve the restricted problem (7), are given by

$$\underbrace{\hat{F}_t^{(n)}}_{n\times 1} = \left(\beta_{t-1}^{(n)\top}\beta_{t-1}^{(n)} + \lambda M_t I_n\right)^{-1} \beta_{t-1}^{(n)\top} R_t^{\text{bond}}.$$

We now show how to select an optimal set of sparse factors from the data. Importantly, our factor selection uses the right objective. We aim to find the optimal sparse selection of factors that minimize the excess return errors of traded coupon bonds, while including a penalty on excessive payoffs of trading bonds with nearby maturities in terms of a smoothness reward. We achieve this by adding a LASSO selection penalty to the optimization problem in Theorem 1, which leads to

$$\min_{F_t \in \mathbb{R}^N} \left\{ \underbrace{\frac{1}{M_t}\left\| R_t^{\text{bond}} - \beta_{t-1}^{\text{bond}} F_t \right\|_2^2}_{\text{return error}} + \underbrace{\lambda \left\| F_t \right\|_2^2}_{\text{smoothness}} + \underbrace{\lambda_1 \left\| F_t \right\|_1}_{\text{selection}} \right\}. \tag{14}$$

The LASSO penalty $\lambda_1 \|F_t\|_1$ in problem (14) selects a sparse solution. There is a one-to-one mapping between the values of the LASSO parameter λ_1 and a target number of factors n. Hence, we can select the best n factors in terms of trading off return errors and smoothness. The resulting estimator combines ridge and LASSO shrinkage, which corresponds to an elastic net penalty. Introducing the LASSO penalty into the kernel regression provides a link between basis function and factor selection.

The effect of LASSO and ridge shrinkage are closely related as, in our case, they both reward the dominating eigenvalues of the kernel space. The ridge shrinkage implies a "soft" shrinkage of small eigenvalues, while the "hard" shrinkage of LASSO sets non-selected eigenvalues to zero. In order to illustrate this intuition, we provide a closed-form solution for problem (14) in the appendix. We show that both forms of regularization select or upweight the first basis functions in V, which are associated with the dominating kernel eigenvalues.

Equipped with our novel estimator, we will perform an extensive empirical analysis. We refer to the estimator given in Theorem 1 as the full KR model, and to the model with the n selected factors corresponding to the largest eigenvalues of the kernel matrix as the KR-n factor model.

2.5 Cross-Sectional Asset Pricing

Cross-sectional asset pricing is fundamentally a spanning problem, that explains the relative return relationship of bonds with different characteristics through exposure to risk. So far, the linearity of our estimator allows us to approximate Treasury bonds and discount bonds as portfolios of tradable factors. However, this curve fitting solution does not necessarily imply risk factors, which demand two requirements. First, the loadings to factors have to capture risk, that is, we have to show that the time-series covariances of our factors with fixed income claims imply the same exposure as our cross-sectional loadings. Second, the factors have to explain the risk premia of fixed income claims. Hence, in addition to explaining the variation with the test assets, the factors also need to explain the average expected excess returns.

We evaluate the cross-sectional asset pricing performance with two standard metrics. First, we test how well different factors explain the cross-section of expected excess returns of discount bond. Second, we construct the stochastic discount factor (SDF) implied by risk factors and evaluate its feasible Sharpe ratio.

In more detail, we test empirically whether the KR-n factor models price the full cross-section of discount bond returns. We use the conventional asset pricing setup and run time-series regressions of discount bond excess returns R_t on the first n cross-sectionally estimated KR factors, say $f_t \equiv \hat{F}_t^{(n)}$,

$$R_t = a + bf_t + \epsilon_t.$$

The KR factors f_t price the cross-section of discount bond excess returns if and only if the pricing errors are zero, i.e. $a = 0$, for loadings measuring risk, i.e. $b = \mathrm{Cov}(R_t, f_t)\,\mathrm{Cov}(f_t)^{-1}$. Then, the risk premium of the discount bonds satisfies $\mathbb{E}[R_t] = b\mathbb{E}[f_t]$ and is fully explained by the risk exposure. We show empirically that b equals the corresponding columns of the loading matrix β from Theorem 1, which justifies its interpretation as loadings. Overall, this is just the canonical cross-sectional asset pricing perspective that is pursued among many others for example in Fama and French (1993).

The SDF is a projection on the asset space that prices all assets. As all default-free fixed income claims are spanned by the full set of discount bonds, a factor model, that can span the cross-section of discount bonds, can also span all fixed income claims in a conditional factor model specification. In this case, it is sufficient to project the SDF on the factors to price themselves, which is equivalent to constructing the tangency portfolio based on the factors. The Sharpe ratio of the tangency portfolio is a direct measure of the pricing information that is captured by the factors, as well as a measure for the profitability of trading the factors. Therefore, a higher feasible out-of-sample Sharpe ratio of a factor tangency portfolio indicates that factors better approximate the true underlying SDF projected on Treasury bonds, which theoretically should achieve the highest possible Sharpe ratio.

Given the traded factors f_t, we compute the maximal Sharpe ratio $\mathrm{SR}_f = \sqrt{\mathbb{E}[f_t]^\top \mathrm{Cov}(f_t)^{-1}\mathbb{E}[f_t]}$ of the tangency portfolio spanned by the f_t with the weights $b_f = \mathrm{Cov}(f_t)^{-1}\mathbb{E}[f_t]$ of the mean-

variance frontier that is spanned by f_t. As discussed, for example, in Kozak, Nagel, and Santosh (2020) and Lettau and Pelger (2020b), the implied SDF spanned by the factors f_t is an affine transformation of the tangency portfolio given by

$$\text{SDF}_t^f = 1 - b_f^\top \left(f_t - \mathbb{E}[f_t] \right).$$

Explaining the cross-section of expected returns with a low rank factor model has direct investment implications. The four KR factors, which we show to price all discount bonds, represent tradable portfolios that replicate and hedge the full term structure. In this sense, the effective dimensionality of the Treasury market is $n = 4$, and investors only need access to these four portfolios. Hence, a cross-sectional factor model can also be understood as a dimension reduction, where the factors capture the first and second moments of the panel.

The conditional term structure premium follows the same arguments but models the conditional moments of the factors. Importantly, for both, the unconditional and conditional risk premium analysis, the cross-sectional relationship between discount bonds is completely captured by the cross-sectional loadings β.

3 Factor Structure

We conduct a comprehensive empirical analysis of the KR method applied to the U.S. Treasury market and study its low dimensional factor structure. We analyze the number and structure of the KR factors and relate them to PCA factors.

3.1 Data

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 of securities maturing within 10 years. The sample period is 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 day prices. Throughout our analysis we measure time in years. Our data set is very close to Filipović, Pelger, and Ye (2022), but limited to bonds maturing within 10 years, since our study is based on a balanced panel of Treasury bonds and longer maturities are not available in the first part of the sample.

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.[13] This step ensures 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

[13]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".

(2010) 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 (2010) and Liu and Wu (2021) and exclude the two most recently issued securities with maturities of 2, 3, 4, 5, 7, 10 years for securities issued in 1980 or later. Since our analysis requires daily returns of securities, on each date, we retain only securities whose prices on the subsequent business day are available. After applying the filters, this gives us a total of 5,335 issues of Treasury bonds and 2,168,382 end-of-day price quotes for 14,865 days.

We study the excess returns of the traded bonds. We use the risk-free one-business day return estimated in Filipović, Pelger, and Ye (2022). This is the return of the discount bond maturing on the next business day. During regular weekdays this corresponds to a one-calendar day return, but over weekends and bank holidays it includes additional calendar days, which we accordingly adjust for as explained in Footnote 7. We specify the discounted cash flows Z_{t-1} based on the discount curve prevailing at $t-1$ estimated in Filipović, Pelger, and Ye (2022).

3.2 Term Structure Estimation

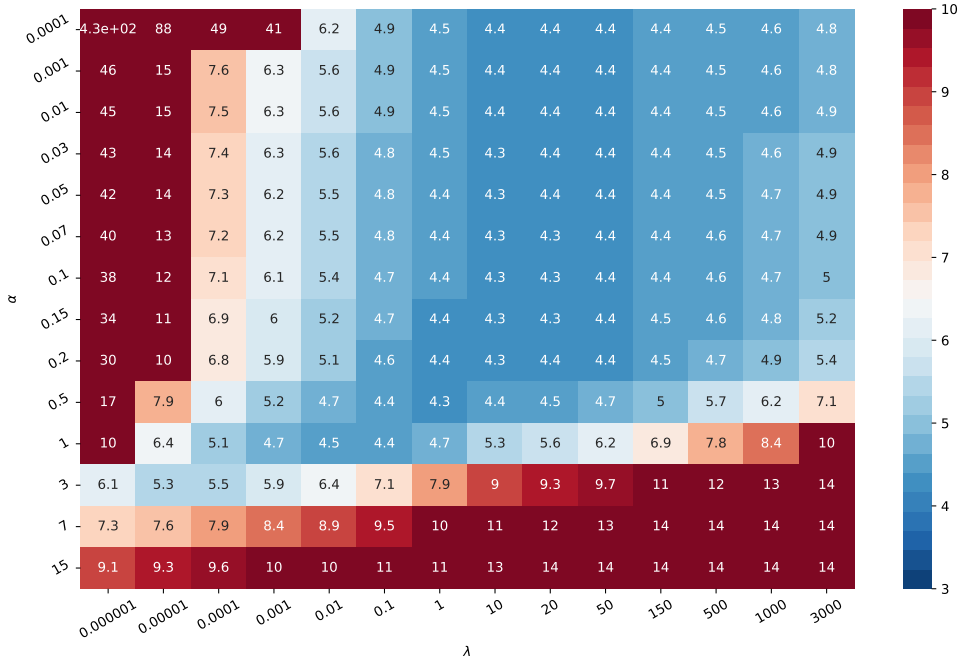
We start our analysis by selecting the optimal parameter values for our KR method. The KR method is completely specified up to the smoothness parameter λ and the maturity weight α. We denote as full KR the estimator with all factors, while we refer to KR-n as the estimator with n factors. Our main metric is the error in excess returns of observed Treasury bonds. We select the optimal set of parameter values in a data-driven way by applying cross-sectional leave-one-out cross-validation. This means that in a cross section, each security is left out only once, and the remaining securities form a training set, on which we fit the model. Out-of-sample performance is then evaluated on the single security being left out, and results are averaged across all securities in the cross-section to get out-of-sample performance.

Note that the estimation error can be interpreted as a hedging or replication error. Our KR method constructs a portfolio of all traded coupon bonds that mimics a target coupon bond. The error of this replicating portfolio is the estimation error. This interpretation is uniqued to our method as no other term structure fitting approach maps directly into tradable portfolios.

Figure 1 shows the cross-validated root-mean-squared error (RMSE) in excess returns for the full KR estimator as a function of λ and α. First, we observe that for a wide range of values, the choice of α has a negligible effect on excess returns. Selecting nearly equal maturity weights (α close to zero) is not optimal, but as long as alpha is within a reasonable range below 0.2, the results are robust. We select $\alpha = 0.05$ as the baseline, which based on Filipović, Pelger, and Ye (2022) allows us to maintain flexibility when extending the maturities beyond 10 years. In contrast, the choice of the smoothness parameter λ has a large effect on the estimation. The optimal value is attained for $\lambda = 10$.[14] Estimating the term structure of returns without a smoothness penalty results in

[14]For better interpretability we normalize the smoothness parameter λ by the number of calendar days, that is, by 1/3650. In other words, we trade off the average daily smoothness with the average daily return error, which are on a more comparable scale.

Figure 1: Cross-Validation Excess Return RMSE for λ and α



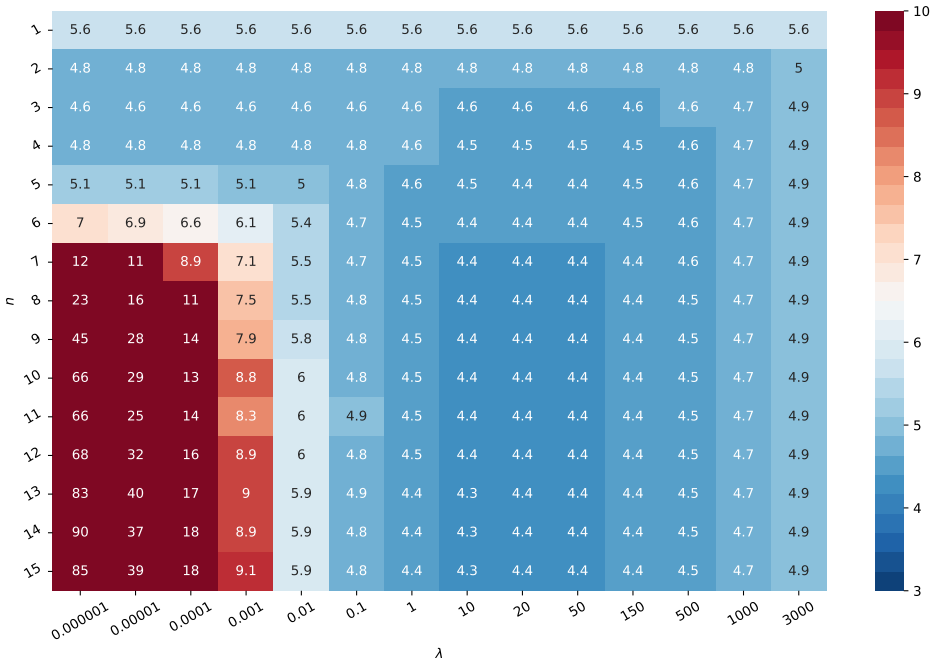
This figure shows the cross-validation excess return RMSE of the full KR model in basis points (BPS) as a function of the smoothness parameter λ and the maturity weight α. We apply cross-sectional leave-one-out cross-validation, this means on each day, each each security is left out only once for out-of-sample evaluation and the model is estimated on the remaining securities. The out-of-sample results are averaged across all securities. We use the last day of each quarter from June 1961 to December 2020 to speed up the calculation.

excessive overfitting. In fact, even an extreme smoothness penalty provides better out-of-sample results than no penalization. The RMSE is very robust to the choice of λ and values between 1 to 100 provide close results. We select $\alpha = 0.05$ and $\lambda = 10$ as the baseline parameters for the full KR model.

Figure 2 compares the cross-validated RMSE as function of the number of factors n and the smoothness penalty, keeping the maturity weight fixed at $\alpha = 0.05$. Factors are selected by the order implied by the kernel matrix decomposition, that is, in decreasing order of smoothness. First, we observe that a sparse set of basis functions can achieve the same RMSE as the full model. In fact, a very small number of factors between 3 and 6 explains already most of the term structure returns. The ridge shrinkage λ can be interpreted as a soft thresholding of lower order factors, while the choice of a small n corresponds to a hard thresholding. In this sense, sparsity in the factors and a smoothness penalty are partial substitutes. Hence, a lower order factor model with $n = 4$ achieves a relatively small RMSE even without the smoothness penalty. However, using a smoothness penalty of $\lambda = 10$ benefits also sparse factor models.[15] This means we can almost perfectly replicate any traded coupon bonds by trading only four to six investable factor portfolios.

[15]The Internet Appendix shows the corresponding in-sample results. They confirm the overall robustness of the parameter choices. As the in-sample results are prone to overfitting, a more regularized estimation with either larger smoothness penalty λ or fewer factors n has obviously higher in-sample RMSE.

Figure 2: Cross-Validation Excess Return RMSE for λ and n



This figure shows the cross-validation excess return RMSE of the KR-n factor model in basis points (BPS) as a function of the smoothness parameter λ and the sparse number of factors n. Factors are added by the order implied by the kernel decomposition, that is, in decreasing order of the eigenvalues of the kernel matrix $\boldsymbol{K}$. We apply cross-sectional leave-one-out cross-validation, this means on each day, each each security is left out only once for out-of-sample evaluation and the model is estimated on the remaining securities. The out-of-sample results are averaged across all securities. We use the last day of each quarter from June 1961 to December 2020 to speed up the calculation.

Figure 2 can be interpreted as showing the amount of variation in traded coupon bonds that is explained for different number of KR factors. The results show that with coupon bonds as test assets at most the first six factors are systematic. In other words, the higher order KR factors do not capture systematic variation in coupon bonds.

Based on the cross-validation, we define our baseline models as $\lambda = 10$ and $\alpha = 0.05$ for the full KR and $n = 1, \ldots, 6$, $\lambda = 10$ and $\alpha = 0.05$ for the low rank factor model KR-n. We study these models in more detail in the next sections.[16]

It is sufficient to compare the KR factor model with the full KR specification. Filipović, Pelger, and Ye (2022) have already shown in a comprehensive empirical study that the KR method dominates existing benchmark methods in terms of pricing and yield errors. Hence, the KR method provides the most precise and robust estimates of discount bond prices. Alternative estimators for discount bond returns, for example Fama and Bliss (1987), Gürkaynak, Sack, and Wright (2010) or Liu and Wu (2021), would first estimate the discount bond prices on two consecutive days. Those

[16] In the Internet Appendix we illustrate the estimated return curve on three representative example days. Including more factors allows the estimated curves to capture more complex patterns and move closer to the full KR curves. A KR factor model with four to six factors approximates the full KR curve well. We also apply the sparse basis functions to yield estimation instead of return estimation. Our results illustrate that the yield estimates of Filipović, Pelger, and Ye (2022) can also be approximated well by a small number of yield factors.

price estimates are then used to calculate the implied returns. Note that in contrast to our KR methodology the returns of alternative parametric or nonparametric estimators do not represent returns of traded assets. In this paper we are the first to estimate the discount bond returns directly from the returns of observed returns instead of first estimating discount bond prices.

Figure A.2 in the appendix illustrates the small magnitude of the errors of the KR method. We report the RMSE of observed securities for the full KR model, KR-n with $n = 1, ..., 4$ factors and estimates of Gürkaynak, Sack, and Wright (2010) labeled as GSW. The RMSE are calculated in-sample for all maturities or separately for seven maturity buckets. In the case of GSW we use their estimated discount bond prices to obtain the discount bond returns. The discount bond returns and discounted cash flows are sufficient to model the coupon bond returns. It is striking that the in-sample RMSE of GSW is larger than the out-of-sample RMSE with KR. In fact, the out-of-sample errors of only two factors in KR-n, as shown in Figure 2, are smaller than the in-sample errors of GSW.

3.3 Selection of Term Structure Factors

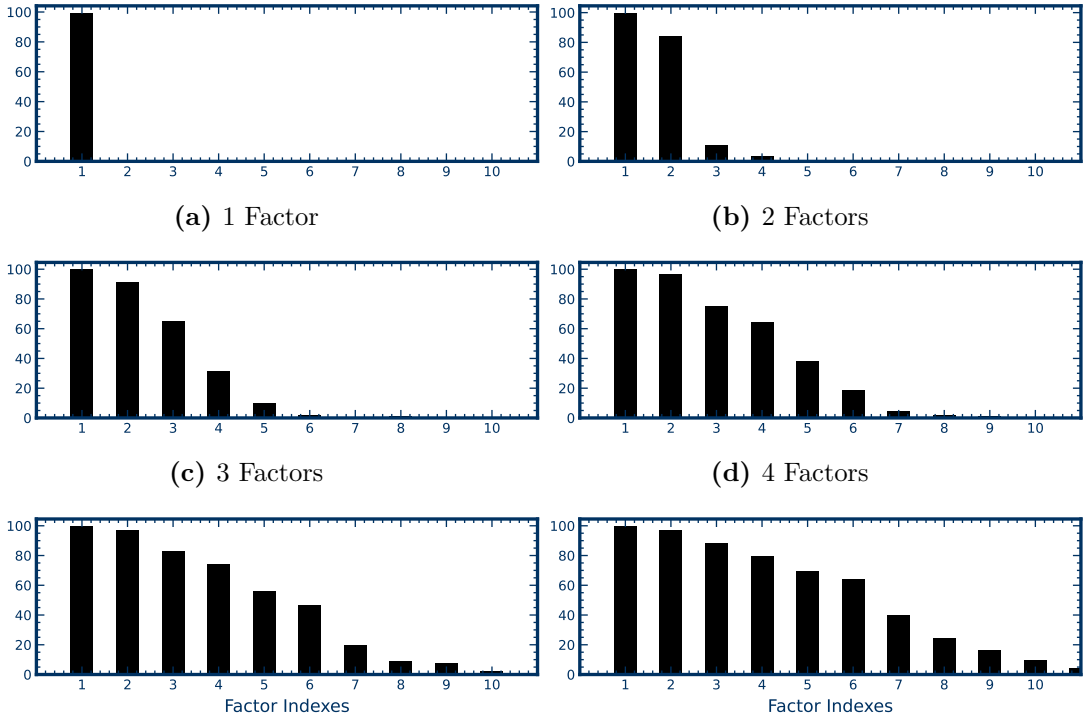
We show that the optimal factors are the dominating kernel factors. In our n-factor KR model, we retain the top n factors associated with the largest eigenvalues of the kernel matrix. We show that it is not beneficial to replace them by factors based on kernel eigenvectors of smaller eigenvalues. For this purpose, we use a data-driven approach to select an optimal sparse factor model. We use the LASSO estimation formulated in Equation (14) to select n factors, instead of selecting them based on the kernel order.

Figure 3 displays the fraction of times when the ith factor is selected by LASSO in an n-factor model. The selection is run separately on the last day in every quarter and the results are aggregated over the full sample. We observe that with a high probability an n-factor model will exactly select the first n factors based on the kernel order. For the first two factors this happens almost surely. For a 4-factor model there is a low probability to also select the fifth or sixth factor, but for most days the first four kernel factors are chosen. We conclude that the selected factors closely follow the kernel order and hence in the following we use the kernel order to describe an n-factor model.

3.4 Explaining Slope and Curvature by Smoothness

We provide an explanation for slope and curvature shapes frequently encountered in PCA based term structure factors. The crucial insight is that smoothness is a fundamental principle of the term structure of returns. The optimal basis functions that span the return curve are the smoothest basis functions in decreasing order. These basis functions capture exactly the slope and curvature type shapes, which are then consequently estimated with a PCA applied to a panel. In other words, slope and curvature factors arise when the underlying problem is fitting a smooth functional relationship.

Figure 3: Frequency (%) of factors selected by LASSO



(a) 1 Factor **(b)** 2 Factors

(c) 3 Factors **(d)** 4 Factors

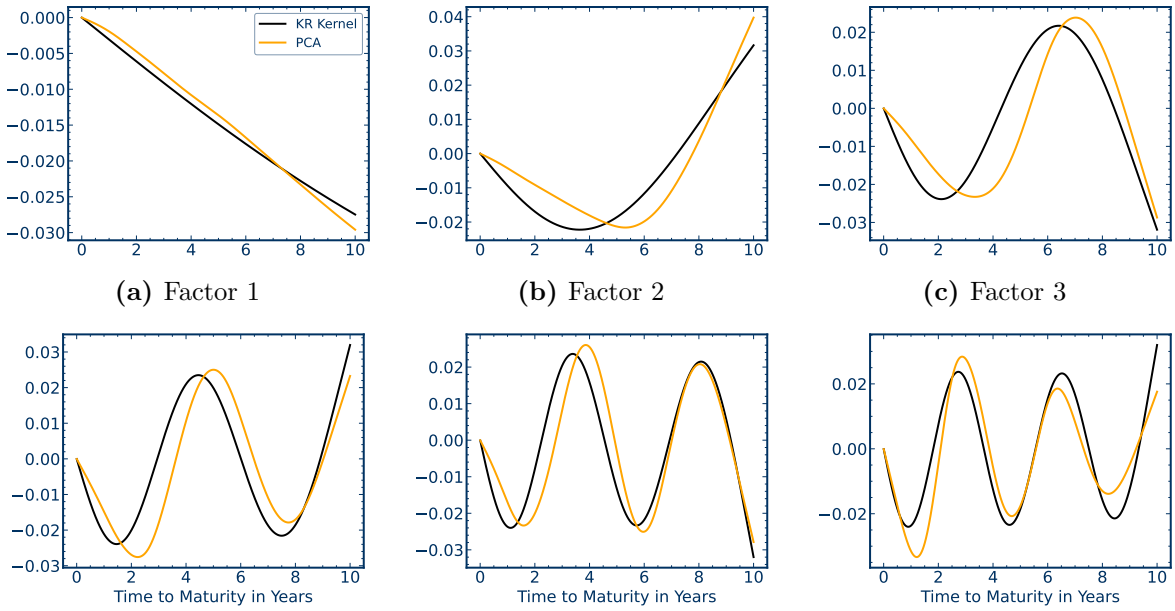
(e) 5 Factors **(f)** 6 Factors

This figure shows fraction of times the i-th factor (horizontal axis) is selected by LASSO when at most n factors can have non-zero coefficients for n ranging from 1 to 6. We run LASSO separately on the last day of every quarter in the sample period, and aggregate results to calculate the fraction of times each factor has non-zero coefficient. We use the baseline model with parameters $\alpha = 0.05$ and $\lambda = 10$. The sample is daily data from June 1961 to December 2020.

First, we study the portfolio weights of KR factors and show that PC factors simply extract the KR factors. We can interpret the structure of the KR factors from their portfolio weights on discount bonds. The KR loadings β correspond to the relative portfolio weights in discount bonds. Hence, the shape of the KR loadings tell us which maturities the factors load on. Figure 4 shows the loadings β of the KR factors and the eigenvectors of PCA factors obtained from the covariance matrix of a panel of discount bond excess returns which are estimated with the full KR model. The norms of the loadings are normalized to one for better comparability. The eigenvectors of the PCA also correspond to the loadings and portfolio weights of the PCA factors.

The striking finding is that the portfolio weights of KR and PCA factors are almost identical and capture exactly the same patterns. Both set of factors are based on familiar slope and curvature type patterns. More specifically, the loadings have a polynomial or sinusoidal structure, where the number of roots corresponds to the order of the factors. As we work with excess returns, we have mechanically removed a "level" factor, which would appear if we modeled factors in returns. Hence, a two factor model would correspond to the common level, slope and curvature three factor

Figure 4: Cross-sectional factor loadings of KR and PCA factors on discount bond excess returns



(a) Factor 1 **(b)** Factor 2 **(c)** Factor 3

(d) Factor 4 **(e)** Factor 5 **(f)** Factor 6

This figure shows the normalized loadings β of the first 6 KR factors and the loadings of the first 6 PCA factors on a panel of discount bond excess returns. The norm of the loadings is normalized to one. The loadings correspond to portfolio weights on discount bond excess returns to construct the KR and PCA factors. The KR loadings β equal the eigenvectors of the KR kernel matrix $\boldsymbol{K}$. The PCA loadings are the eigenvectors of the sample covariance matrix from the panel of discount bond excess returns estimated with the full KR model. The sample is daily data from June 1961 to December 2020.

Table 1: Similarity between KR and PCA Factors

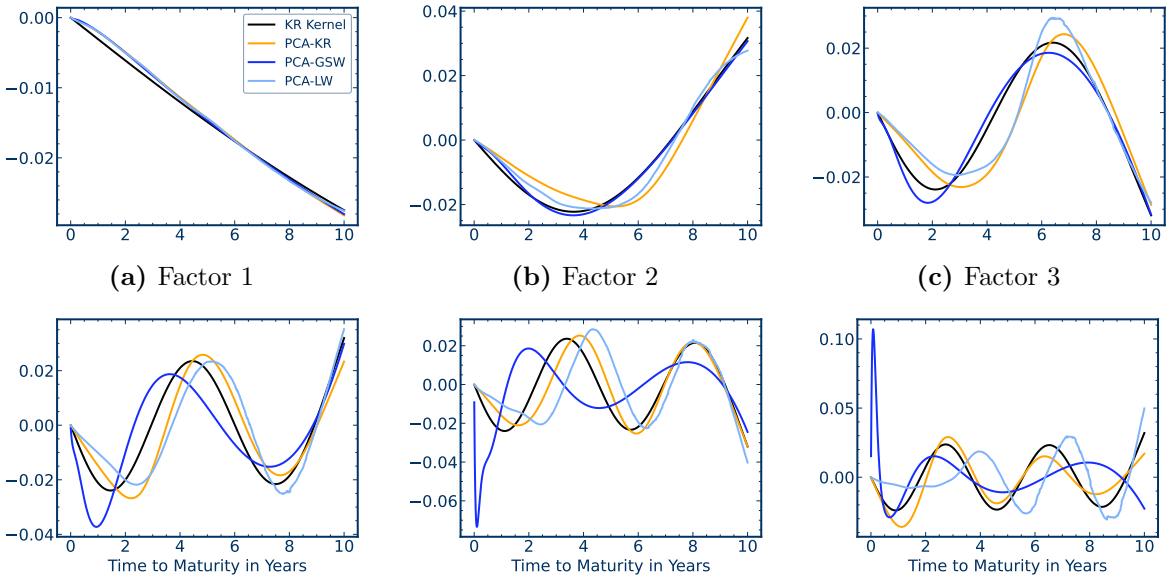
Number of factors n	1	2	3	4	5	6
Explained variation of KR factors	0.916	0.976	0.992	0.997	0.999	1.000
Explained variation of PCA factors	0.920	0.982	0.994	0.997	0.999	1.000
Generalized correlation between KR and PCA	0.998	0.936	0.901	0.943	0.922	0.986

This table shows the cumulative fraction of variation in daily discount bond excess returns of all maturities explained by the first n KR and PCA factors, for n from 1 to 6. The PCA factors are extracted from a panel of discount bond returns estimated with the full KR model. The KR loadings β equal the eigenvectors of the KR kernel matrix $\boldsymbol{K}$, that is, they are estimated without time-series information. The PCA loadings are the eigenvectors of the sample covariance matrix from the panel of discount bond excess returns, that is, they are equivalent to the coefficients from a time-series regression. The last row reports the average generalized correlations between KR and PCA factors.

representation.[17] Higher order factors have a higher curvature and capture more complex patterns.

[17]Our slope factor in Figure 4 is not centered at zero as we work with returns in excess of the one-business day return. If we would consider returns in excess of an equally weighted "market" portfolio of discount bonds, then we would mechanically center the slope factor at zero with the usual long-short portfolio interpretation. In other words, if we apply PCA to returns then the first PCA is essentially an equally weighted market portfolio, and the second PCA a slope factor centered at zero. We work with excess return as it provides the linear mapping between coupon bond and discount bond returns, and it is the basis for the usual risk premium estimation. If we would instead study returns in excess of a bond market return, we would simply have a mechanical level shift for Figure 4, which is inconsequential for the results.

Figure 5: PCA factor loadings of KR, GSW and LW discount bond excess returns



(a) Factor 1 **(b)** Factor 2 **(c)** Factor 3

(d) Factor 4 **(e)** Factor 5 **(f)** Factor 6

This figure shows the normalized loadings of the first 6 PCA factors from a panel of discount bond excess returns estimated with the full KR model, GSW, and LW, respectively. The sample is daily data from June 1961 to December 2020.

Intuitively, a two factor model can be interpreted as a polynomial function of degree two, while a four factor model corresponds to a polynomial function of degree four which can approximate a more complicated curve.

Importantly, we know why KR factors arise, and that allows us to explain why we observe PCA factors. The smoothest basis functions have slope and curvature type patterns and these are exactly the optimal low dimensional basis functions that span the return curve on each day. As the cross-section on each day is optimally spanned by a low dimensional set of these basis functions, this also implies that a panel is spanned by the same basis functions, but with coefficients that can vary for each day. However, it is not mechanically the case that the order and rotation of PCA factors has to coincide with the smoothness order of the KR basis functions. It is an empirical finding that the factor structure estimated with PCA represents the same model as the KR factor structure.

Second, we show that KR factors have the same time-series properties as PCA factors. We start by measuring the correlations between KR and PCA factors. Table 1 reports the generalized correlation and explained variation for the two sets of factors. The high generalized correlations show that the two set of factors span the same space.[18] Indeed, the average generalized correlation

[18]The generalized correlation is also called canonical correlation. Let F be the n KR factors and G are n_G candidate factors. To what degree can a linear combination of the candidate factors G replicate some or all of the factors F? The first generalized correlation is the highest correlation that can be achieved through a linear combination of the factors F and the candidate factors G. For the second generalized correlation we first project out the subspace that spans the linear combination for the first generalized correlation and then determine the highest

between the first four KR and PCA factors is 94%, which means that rotated PCA factors are almost perfectly correlated with KR factors. Next, the explained variation of KR and PCA factors is essentially the same. Note that this is not mechanically given, as the cross-sectional KR factor loadings of discount bonds are given by the smoothness kernel, while the PCA factor loadings are obtained from a time-series regression. The fact, that the factors weights and the explained variation of KR and PCA factors are the same, implies that the cross-sectional loadings and time-series loadings have to coincide as well. We will analyze this point in more detail in the next section. The important finding is that the smoothest basis functions also explain most of the variation in a discount bond panel.

The KR factors are available for each single day without the need for panel data. The loadings and factors weights are simply given by the kernel, which is derived from the smoothness norm. Hence, the KR factors do not only provide insight on the structure in PCAs but they are also applicable if we only observe a cross-section on a single day and, thus, cannot use a PCA estimator.

Third, we show that smoothness is an underlying principle of Treasury markets. Whereas our estimates of discount bond returns have many advantages in terms of precision, explainability and tradability, the dominant PCA factors of discount bond panels estimated with other precise methods reflect the same structure. In other words, precise estimates of discount bonds either explicitly or implicitly try to extract a low dimensional structure based on smoothness, even if the curve estimates themselves are not smooth to begin with. Figure 5 includes the PCA loadings from a panel of discount bond excess returns estimated with GSW and the method of Liu and Wu (2021), labeled as LW. Filipović, Pelger, and Ye (2022) show that the LW estimates are almost as precise as KR estimates, but not smooth. Interestingly, the PCA loadings of LW capture very similar shapes for the first four factors as the KR PCA and kernel loadings. The shapes for factor five and six are still close, and the variation can be due to the imprecision of the LW estimates. In summary, the patterns found in the PCA are not due to the fact that we use the discount bond panel estimated with the full KR model, but are a universal feature of Treasury bond markets.

We want to emphasize that an imprecise estimate of discount bond returns can distort the factors structure. Figure 5 also includes the PCA loadings from a panel of discount bond excess returns estimated with GSW. Filipović, Pelger, and Ye (2022) have demonstrated that GSW results in imprecise and biased estimates of discount bond yields. The first three PCA loadings have similar shapes. However, the loadings of the higher order GSW PCA factors are distorted. A return curve with a large error can omit certain basis patterns, which then do not appear in the cross-sectional modeling. In other words, it is not surprising that higher order term structure effects might not play a role in discount bond panels based on GSW, as the Nelson–Siegel–Svensson model omits those patterns. In the subsequent sections, we demonstrate the importance of the fourth KR factor

possible correlation that can be achieved through linear combinations of the remaining $n-1$ respectively n_G-1 dimensional subspaces. This procedure continues until we have calculated the $\min\{n, n_G\}$ generalized correlation. Mathematically the generalized correlations are the square root of the $\min\{n, n_G\}$ largest eigenvalues of the matrix $\mathrm{Cov}(G,F)\,\mathrm{Cov}(F)^{-1}\,\mathrm{Cov}(F,G)\,\mathrm{Cov}(G)^{-1}$. If $n = n_G = 1$, it is simply the conventional correlation. We report the average generalized correlation between the first n KR and PCA factors. Pelger (2020) provides a further discussion on generalized correlations.

which could not be extracted from panels of GSW discount bonds.

The relationship between smooth curve estimation and slope and curvature type factors is a universal finding. Previous literature including Litterman and Scheinkman (1991), Cochrane and Piazzesi (2005), Crump and Gospodinov (2022) and Filipović, Pelger, and Ye (2022) have documented factor structures with these familiar shapes for PCA applied to panels of discount bond prices or yields. We study the returns of Treasuries as our paper focuses on the risk premium and investment implication of factors. Importantly, the same cross-sectional mapping also applies to panels of discount bond prices or yields. As shown in Filipović, Pelger, and Ye (2022), the same optimal basis functions (up to a level shift) that explain discount bond returns also span discount bond prices or yields, because fundamentally all these problems represent smooth curves. By construction, if we work in the yield or price space the KR estimator maps into portfolios of discount bond prices or yields instead of returns.[19]

The relevance of our findings extends beyond term structure modeling. Slope and curvature patterns in PCAs of panels of asset returns have been demonstrated for a variety of asset classes, including equities and foreign exchange markets. In equities it is common to use portfolio sorts to reflect the conditional impact of firm characteristics. For example, for book-to-market ratios or momentum, one can obtain a panel of deciles sorts for the different quantiles of the characteristics. These basis assets are the analogue to our discount bonds as they have constant exposure to the sorting variable in contrast to individual stocks. As shown in the Internet Appendix, a PCA on equity decile sorts leads to the same slope and curvature patterns as for "maturity sorted" discount bonds. Our findings suggest that these type of factors simply arise because the functional dependency of stock returns on momentum or book-to-market is inherently smooth. This non-parametric perspective is novel, and we are able to make it because our approach unifies non-parametric curve estimation with cross-sectional factor modeling.

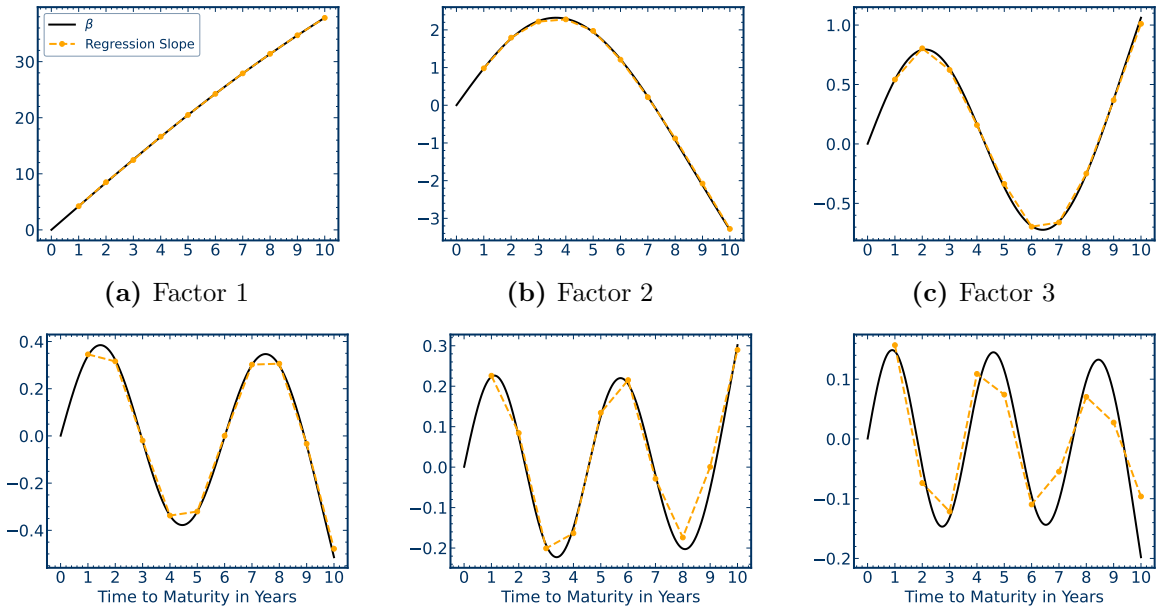
3.5 Cash Flows are Covariances: Equivalence of Cross-sectional and Time-Series Loadings

We show that the conditional KR factor loadings, based on cash flows, equal the time-series regression loadings. The cross-sectional estimation of the KR model implies loadings for KR factors that are completely determined by the cash flows of the bonds without using any time-series information. The loadings of risk factors have to capture the time-series co-movement of assets with factors, and hence are usually estimated in time-series regressions. We show empirically that cross-sectional cash flow information and time-series covariances capture the same term structure risk.

We start with a panel of 10 discount bond excess returns estimated with the full KR method. The maturities of the discount bonds range from 1 to 10 years. The KR loadings are given by β. The discount bond returns have constant cash flows in this panel and hence constant loadings for term structure factors. Figure 6 shows the loadings of these 10 discount bond returns on the first

[19] Appendix A.7 discusses how fitting the cross-sectional relationship between discount bond returns is related to fitting the cross-section of discount bond yields.

Figure 6: Conditional loadings β and time-series regression slopes on discount bond excess returns



(a) Factor 1 **(b)** Factor 2 **(c)** Factor 3

(d) Factor 4 **(e)** Factor 5 **(f)** Factor 6

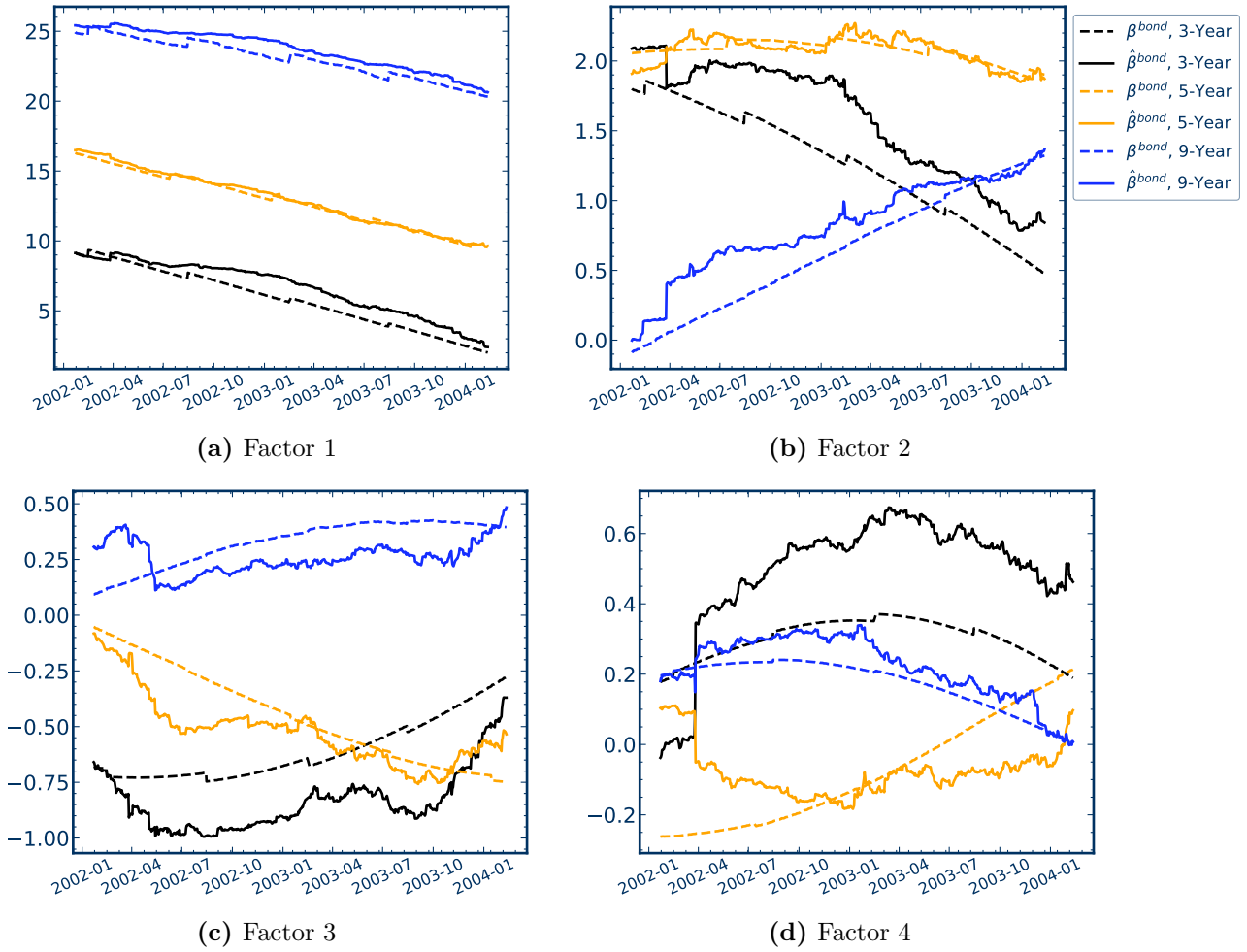
This figure shows the conditional loadings β and time-series regression slopes of KR factors on discount bond excess returns. The panel of ten discount bond excess returns with maturities ranging from 1 to 10 years is estimated with the full KR model. The time-series regression slopes are the coefficients on the first 6 KR factors. The sample is daily data from June 1961 to December 2020.

six KR factors. We compare them with the time-series regression coefficients on these factors. The loadings based on time-series regressions result in essentially the same loadings. Hence, we can either use the cash flow information or the time-series covariances of bond returns to obtain the loadings on discount bonds.

This is an important empirical finding. A priori, it does not need to hold that loadings that are obtained from purely cross-sectional non-parametric estimations are related to the time-series. The result that cash flows and time-series covariances lead to the same betas is a refinement of the observation from the previous subsection that KR factors equal PCA factors of discount bond panels.

Coupon bonds have time-varying discounted cash flows and hence their conditional betas $\beta_{t-1}^{\text{bond}}$ are time-varying. The conditional KR loadings are given by $\beta_{t-1}^{\text{bond}} = Z_{t-1}\beta$ and hence do not require time-series information. As the exposure to term structure factors cannot be constant for coupon bonds, we use a rolling window regression to obtain time-series loadings. Figure 7 compares the conditional loadings and time-series estimates for three example coupon bonds with 3, 5 and 9 years of maturity for the first four KR factors. The rolling window estimate, based on daily excess returns in the past year, is only a crude local estimator. However, we can see clearly that both loading estimators capture the same underlying pattern. We conclude that we obtain approximately the same loadings from the cash flows as from the time-series.

Figure 7: Conditional loadings $\beta_{t-1}^{\text{bond}}$ and local time-series regression slopes on coupon bonds



(a) Factor 1 **(b)** Factor 2

(c) Factor 3 **(d)** Factor 4

This figure shows the conditional loadings $\beta_{t-1}^{\text{bond}}$ and time-series regression slopes of the first four KR factors on three representative coupon excess returns. The dashed lines display the conditional loadings $\beta_{t-1}^{\text{bond}} = Z_{t-1}\beta$, which are determined by the cash flows of the coupon bonds. The solid lines show the slope estimates of a rolling window regression on the four KR factors. The rolling window consists of the daily excess returns in the past one year.

In summary, cash flows are the "characteristics" of coupon bonds. Discount bonds are a special case, where these characteristics are constant. The risk exposure of bonds is completely determined by their characteristics. Similar to "characteristics are covariances" in equity, we find that "cash flows are covariances" in fixed income. We obtain the same loadings for term structure factors from the cash flows of bonds as from the time-series correlation with factors. However, different from equity modeling, this relationship is exact as coupon bonds are exact portfolios of discount bonds. In contrast, the conditional factor loadings for equity can be a complicated function of firm characteristics, which is often modeled ad-hoc as a linear function. In this sense, our factor model is related to the instrumented PCA (IPCA) of Kelly, Pruitt, and Su (2019). Our factors are obtained in only one step by running a cross-sectional regression without the need of iterating it with time-series regressions, and the functional form of our conditional loadings is exact.

Our finding has important consequences for asset pricing. First, exposure to risk in a factor model and its implied stochastic discount factor is fundamentally based on conditional covariances, that is, time-series moments. We show that the cash flow based loadings are equivalent and hence fully explain risk. Second, cash flow information allows us to price bonds even if they do not have a past time-series available.

3.6 How Many Term Structure Factors Explain the Variation?

The number of factors depends on the evaluation objective and the choice of test assets that we try to explain. The estimation of the number of latent term structure factors is often guided by explaining variation in a chosen panel. We show that for informative test assets we require at least four factors to explain the systematic variation. In the next section we demonstrate that we also require four factors for the objective of explaining the risk premium in Treasury returns.

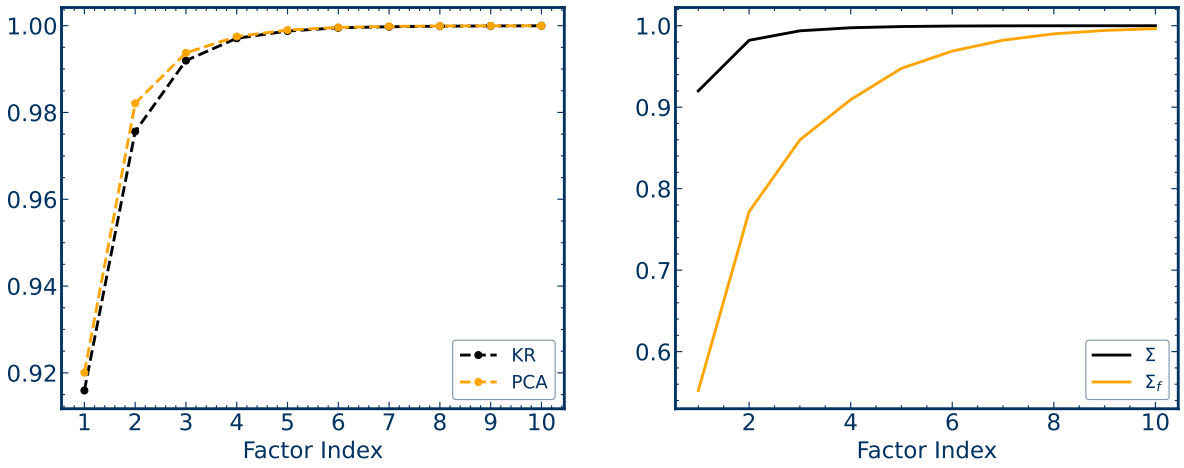
A natural choice of test asset are the traded coupon bonds. As the loadings of factors vary with the time-varying cash flows, we have to use the conditional factor model formulation. Figure 2 shows in an out-of-sample cross-validation analysis that we need four factors to explain the variation in traded coupon bond excess returns. The improvement from a slope and curvature two-factor model to a four factor model is 0.2 basis points in terms of RMSE. All the factors beyond the fourth factor explain cumulatively another 0.2 basis points RMSE.

The most widely used test assets are panels of discount bond (excess) returns or yields. The common perception is that two to three factors are sufficient to explain essentially all the variation in discount bond returns. This perspective is misleading. Figure 8(a) plots the explained variation in the full panel of discount bond excess returns for all daily maturities based on the full KR model for different numbers of factors. As expected given our previous results, the explained variation for KR and PCA factors is essentially identical. The explained variation of PCA factors corresponds to the normalized cumulative sum of eigenvalues of the second moment matrix of the panel.[20] It seems that the first two KR or PCA factors explain over 97% of the variation. However, using a cutoff, for example, explaining 97% of the variation, would be misleading. The returns of discount bonds have a mechanical overlapping dependency structure which inflates the eigenvalues.

Forward returns are a more meaningful set of test assets for measuring dependencies in the cross-section. This builds on the insights of Crump and Gospodinov (2022), who have shown that the eigenvalues in a panel of discount bond returns (and excess returns) are mechanically inflated, whereas the eigenvalues in a panel of forward returns are more informative. Forward returns are based on the cross-sectional differences of discount bond returns, that is $R^{\text{forward}}_{t,i} = R_{t,i} - R_{t,i-1}$. They correspond to a forward agreement for a future one-period risk-free investment. The excess

[20] The portfolio weights of latent factors estimated with a PCA applied to a covariance matrix of a panel of discount bond excess returns are identical to those estimated with a PCA applied to the second moment matrix of the panel. The variation is based on the uncentered second moment, while covariances uses demeaned second moments. All the results of this section are the same when we study demeaned data, that is, measure the unexplained covariance in the panel. We refer to Lettau and Pelger (2020b) for a detailed discussion about applying PCAs to second moment matrices or covariances. As the mean is orders of magnitude smaller than the variance, it has a negligible effect on the variation.

Figure 8: Explained variation and eigenvalue structure for different number of factors



(a) Explained variation in discount bond excess returns for KR and PCA factors

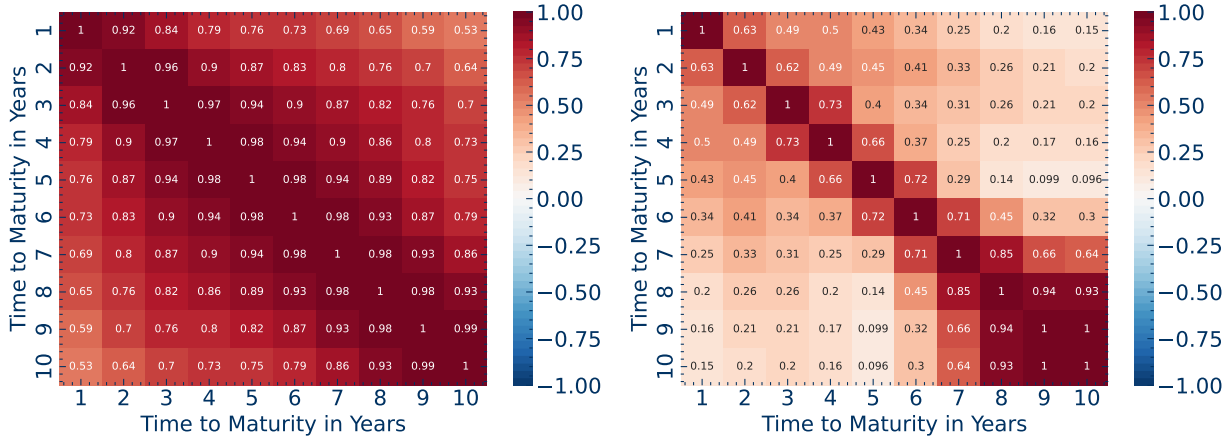
(b) Normalized cumulative sum of eigenvalues of Σ and Σ_f

This figure shows the explained variation and eigenvalue structure for different number of factors. The left subplot (a) displays the cumulative fraction of variation in the panel of discount bond excess returns of all daily maturities estimated by the full KR model, which is explained by the first 10 KR factors and the first 10 PCA factors. The right plot shows the normalized cumulative sum of eigenvalues of Σ and Σ_f. Σ is the covariance matrix of the full panel of discount bond excess returns with all daily maturities estimated with the full KR model. Σ_f is the covariance matrix of forward returns, which is the cross-sectional first difference of discount bond excess returns. Each eigenvalue is normalized by the sum of all eigenvalues. The sample is daily observations from June 1961 to December 2020 for all daily maturities.

return of a discount bond with maturity of i periods is a portfolio of i forward returns, that is, $R_{t,i} = \sum_{s=1}^{i} R_{t,s}^{\text{forward}}$. Hence, even if forward returns were independent, discount bond excess returns have a strong dependency structure with correlations close to one for nearby bonds. This structure is closely related to the unit-root problem in time-series analysis. This mechanical dependency is removed for forward returns, similarly to first differences of unit-root processes. Figure 8(b) plots the explained covariance in a panel of forward returns, that is the normalized eigenvalues of the covariance matrix of forward returns. The explained covariance of a PCA in forward returns shows that four factors are needed to explain at least 90% of the covariance dependency. A two factor model (corresponding to slope and curvature) would only explain around 77% of the covariance structure.

Figure 9 shows the correlation structure of a panel of discount bond excess returns and forward returns. We consider 10 assets with yearly maturities. The heatmap in Figure 9(a) illustrates the mechanical dependency in discount bond excess returns, which inflates the eigenvalues. In contrast, the correlation in forward returns is substantially lower, as expected given the eigenvalue results. This means that there is a dependency structure in forward returns, but it can be "hidden" by the mechanical dependency in discount bond excess returns. Figure A.7 in the appendix shows that the PCA loadings on forward returns have essentially the same structure (besides a mechanical shift in maturities). Hence, the choice between excess returns and forward returns affects mainly the number of factors but not their structure. We conclude that we also need four term structure

Figure 9: Correlation matrices of discount bond excess returns and forward returns



(a) Correlation of excess returns

(b) Correlation of forward returns

This figure shows the correlation matrices of discount bond excess returns and forward returns. The discount bond excess returns are estimated with the full KR method. The figures show the correlations for the subset of 10 discount bonds, which form a panel for yearly maturities ranging from 1 to 10 years. The sample is daily data from June 1961 to December 2020.

factors to explain the correlation for the more informative forward returns.

In summary, the first four KR factors explain most of the variation in Treasury returns. The discussion about the strength of the factors depends on the choice of test assets and the objective.

4 Asset Pricing

Next, we study the cross-sectional asset pricing implications of the factors. We start by establishing that the term structure premium depends on the estimation method for discount bond returns. Then, we show that four KR factors explain the term structure premium and how the factor premium depends on economic conditions.

4.1 Term Structure Premium

We study the expected return curve for the full maturity spectrum of discount bonds. The term structure premium equals the expected return of discount bonds with different maturities minus the average risk-free one-business day return. It measures the compensation for risk associated with different maturities.[21]

Statements about the term structure premium depend on the estimation method. We confirm the findings of Filipović, Pelger, and Ye (2022) for different methods, and also study how the number of factors affects the spanning of the term structure premium. We compare our method (KR) with the estimates from Fama and Bliss (1987) (FB), Gürkaynak, Sack, and Wright (2007)

[21] Modeling expected returns is without loss of generality as we can directly revert them back to expected excess returns. Hence, the term structure premium curve is obtained by simple shifting the intercepts of the curves to zero.

(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 used in the literature. The methods are discussed in detail in Filipović, Pelger, and Ye (2022), and we follow exactly the same implementation as in that paper. The benchmark methods do not estimate models for returns but for discount bond prices, which we then use to calculate returns. The FB curve is constructed from piece-wise constant forward rate estimates. The FB dataset of estimated forward rates is available from June 1961 to December 2013, which we use to construct the yield curves and to price Treasury bonds.[22] As the 10-year rate is unavailable in the FB data set until November 1971, we use the shorter sample from November 1971 to December 2013 to compare the different estimation approaches. The appendix collects all the results for the KR model for the full sample from June 1961 to December 2020 with the same findings. We implement the NSS model on the same underlying dataset 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. 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.[23] Importantly, only our KR estimates 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.

Figure 10 shows the term structure premium implied by different estimators. The sample is restricted to the time-series for which the Fama and Bliss (1987) estimates are available. The figure shows the mean of the estimated discount bond returns. The term structure premium follows from subtracting the average risk-free one-business day return. Overall, the different estimates share a similar shape with a premium increasing in maturity. However, the benchmark estimates show systematic biases and instabilities. The returns of FB estimates result in extremely unstable time-series, which is reflected by the kinks and spikes in the average returns. This makes them also problematic as an input for portfolio optimization. The less precise methods GSW and NSS, and LW systematically overestimate the average returns for 6 to 10 years of maturity. The instability in the price estimates of FB and LW also results in more volatile estimated return time-series. These observations confirm the findings in Filipović, Pelger, and Ye (2022).[24]

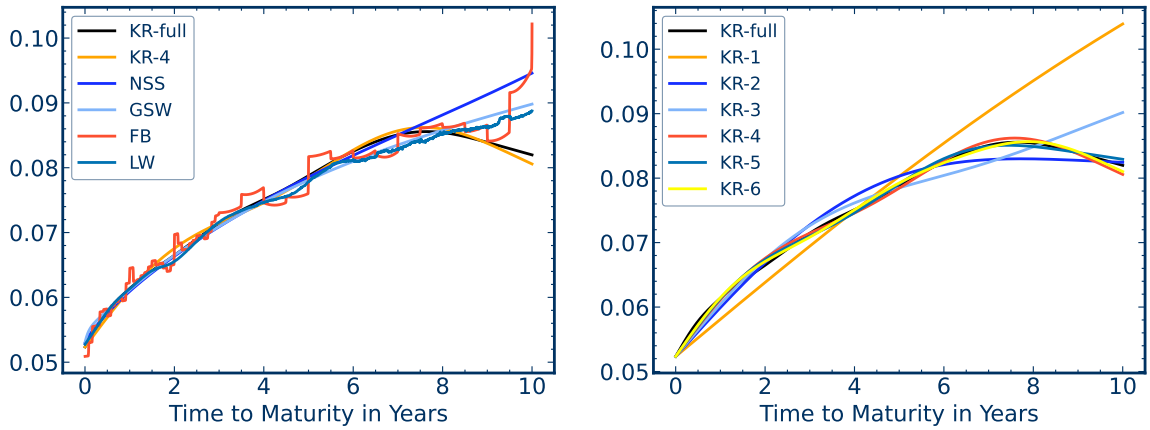
We need four KR factors to explain the term structure premium. Figure 10(b) shows the term

[22]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.

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

[24]We can use the KR approach to either directly estimate the return curve, or to estimate discount bond prices curves on two consecutive days to calculate the discount bond returns. We obtain very similar return estimates from the KR method applied to prices or returns. As our focus is on the implication for the return factor structure, we only report the KR estimates for returns. The results for returns based on the KR estimates from bond prices, are available upon request.

Figure 10: Expected returns of discount bonds for different estimators



(a) Different estimators **(b)** Number of KR factors

This figure shows expected one-day discount bond returns for different estimators. The left subplot shows the expected discount bond returns implied by the full KR, four-factor KR, GSW, NSS, FB, and LW estimates. The KR estimator directly estimates the term structure of returns, while GSW, NSS, FB and LW estimate yields for discount bonds. The discount bond returns are averaged over time from November 1971 to December 2013, as the 10-year rate is unavailable in the FB data set until November 1971. The right subplot shows the average discount bond returns given by KR models for different number of factors. We consider the full KR model and one to six KR factors. For better comparison we use the time period as in the left plot. Figure A.4 in the appendix shows average returns for daily data from June 1961 to December 2020 with the same results. The average one-business day returns are annualized.

structure premium for the full KR model and for one to six KR factors. The difference between the curve for the full KR model and the respective curves based on KR factors represents the cross-sectional pricing errors in discount bonds. The first three factors are not sufficient to capture the term structure premium, and induce pricing errors, in particular for the long end. The KR-4-factor model provides a very close approximation of the term structure premium of the full KR model. We conclude that the fourth KR factor is essential for explaining the term structure premium. Figure A.4 in the appendix shows that the same findings hold if we use the full sample until the end of 2020.

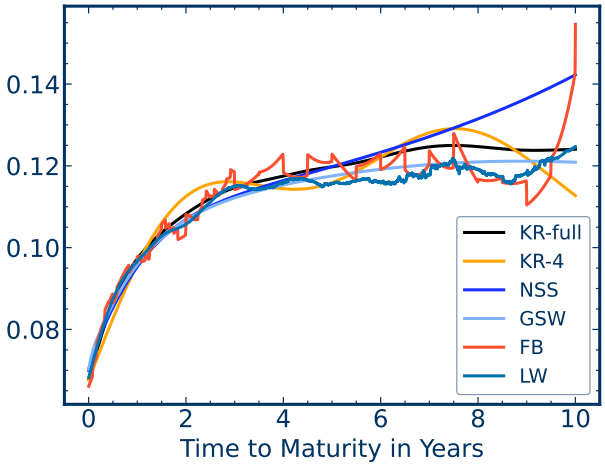
The term structure premium has a complex shape. The first factor captures a linear slope and hence can only represent a linearly increasing term structure premium. The second curvature factor is too restrictive to approximate the non-trivial shape of the term structure premium. As the fourth factor represents a basis function that approximates complex shapes, and is necessary for spanning the term structure, we interpret the premium of the fourth factor as a complexity premium.

4.2 Complexity Premium

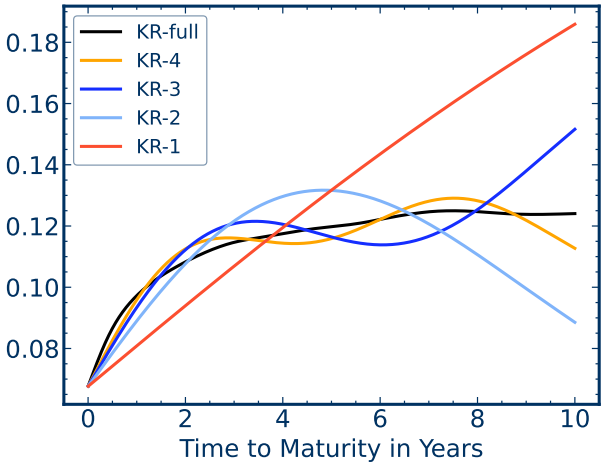
The interpretation of a "complexity premium" can be better understood from studying the conditional term structure premium for different number of factors. We consider transparent and relevant conditioning variables, namely NBER boom and recession indicators and the yield spread.

Figure 11: Average discount bond returns conditioned on boom/recession

Panel A: Recession

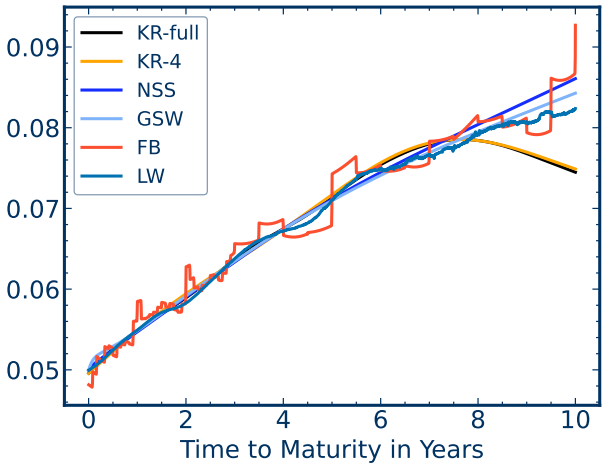


(a) Recession: Different estimators

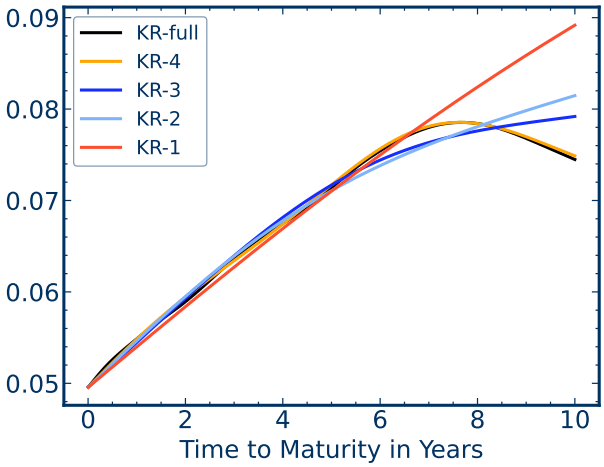


(b) Recession: Number of factors

Panel B: Boom



(c) Boom: Different estimators



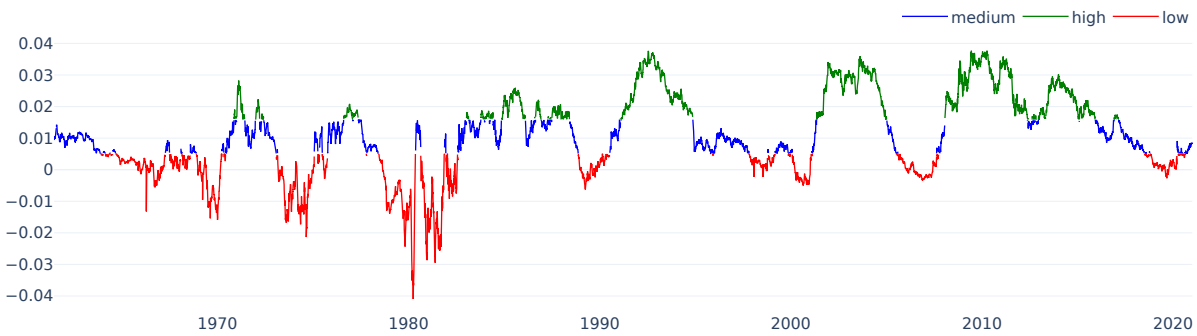
(d) Boom: Number of factors

The figure show the average returns of estimated discount bonds conditional on boom and recession dates. The left subfigures display the conditional returns implied by the NSS, GSW, FB and LW estimates, and KR models with different number of factors. The right subfigures report the conditional returns for different number of KR factors. Business cycle dates are from NBER. The returns are averaged over time from November 1971 to December 2013. Figure A.5 in the appendix shows the results for different numbers of KR factors for the full sample of daily data from June 1961 to December 2020. The average one-business day returns are annualized.

The yield spread is directly related to macroeconomic conditions and allows us to isolate the effect for time periods with inverted yield curves. Figure 12 shows the times-series of the yield spread between 10 years and 1 year of maturity. The yield estimates are from Filipović, Pelger, and Ye (2022) based on 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 11 shows the average returns of estimated discount bonds conditional on boom and

Figure 12: Yield spread terciles



The plot shows the time-series of the yield spread between yields of 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.

recession dates.[25] The right subfigures report the conditional returns for different number of KR factors. During times of booms, the shape of the term structure takes a simpler form. In fact, a simple slope factor captures most of the information about the term structure premium for up to 6 years of maturity. However, the situation is entirely different for average returns during recession periods. The term structure is substantially more complex during recessions. Using models with only the first three factors result in strongly biased estimates of the term structure premium. Including the fourth factor is key for capturing the same shape as with the full KR model. This provides an argument for why including the fourth factor is the most relevant for investments during recession periods. The term structure premium can be described by a simple and close to linear function during good times, where the "complex" higher order factors carry a low conditional risk premium. The complexity premium matters during bad times, which require higher order factors to adequately describe the term structure premium.

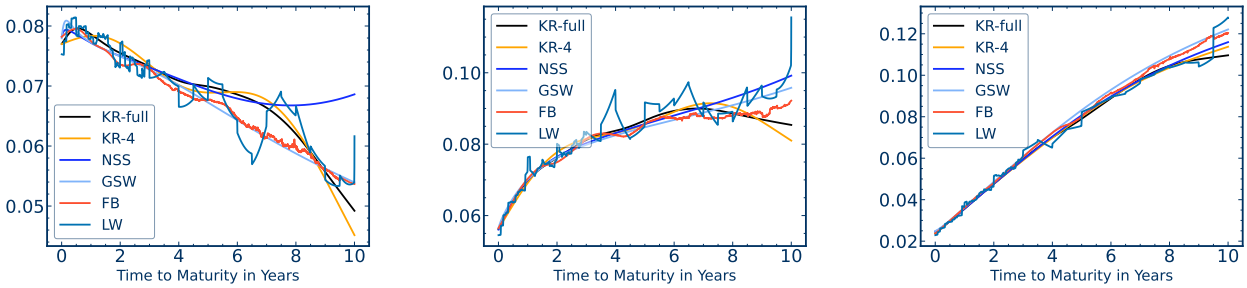
Conditioning on yield spreads leads to similar findings as conditioning on NBER recessions. Figure 13 shows the average returns conditional on yield spread terciles. The bottom panel compares the conditional discount bond returns for the KR models with different number of factors. The times of high yield spreads coincide with simple shapes of the term structure premium. The complexity premium matters the most during low and medium yield spread terciles. Omitting the fourth factors leads to biased estimates relative to the full KR model.

The precision of the KR method matters for a conditional investment analysis of discount bond returns. Figures 11 and 13 also show the conditional average returns of estimated discount bonds for alternative estimation approaches. The left subfigures of Figure 11 and the top panel of Figure 13 display the conditional returns implied by the NSS, GSW, FB and LW estimates, and KR models with different number of factors conditional on boom and recession dates respectively yield spread

[25] To assess the economic significance in the differences between curves, we report the annualized mean of the business day returns. We obtain it by scaling the average business day return by 252, the number of business days. The results are very similar to annualizing average calendar day returns.

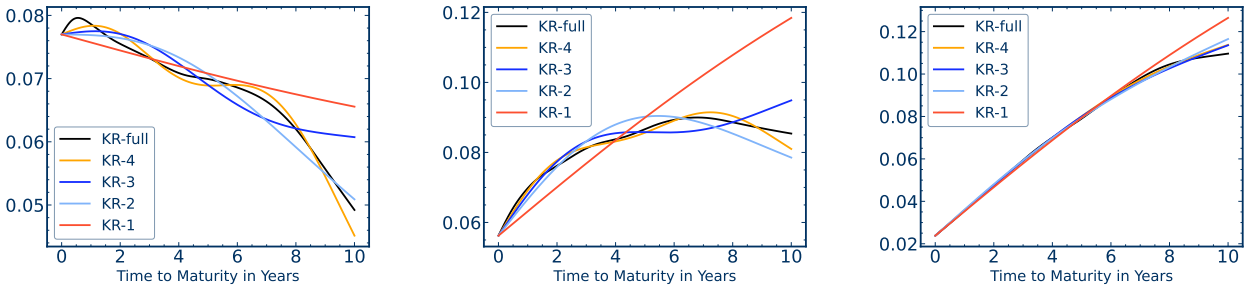
Figure 13: Average discount bond returns conditioned on yield spreads

PANEL A: Different estimators



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

PANEL B: Number of factors

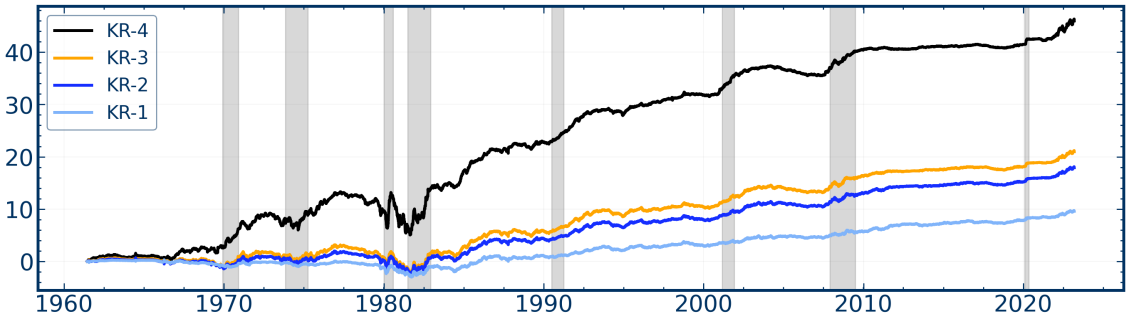


(d) low yield spread **(e)** medium yield spread **(f)** high yield spread

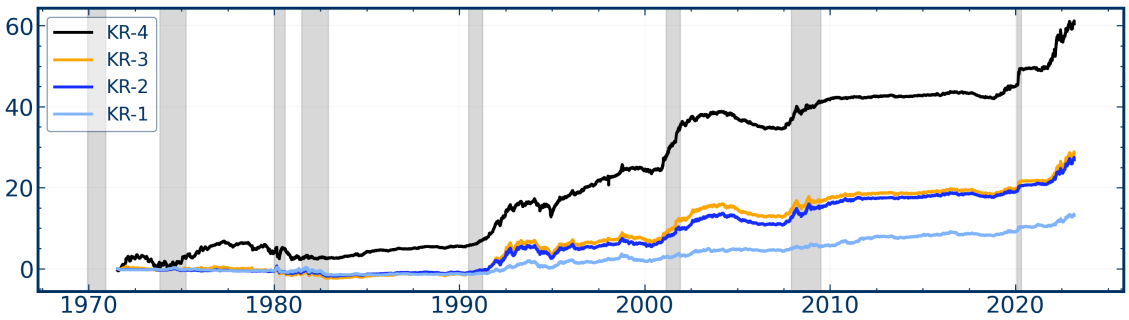
The figure shows the average discount bond returns conditioned on yield spreads. The top panel compares the conditional discount bond returns implied by the KR, NSS, GSW, FB and LW estimates for low/medium/high yield spread dates. The bottom panel compares the conditional discount bond returns for KR models with different number of factors for the corresponding time periods. The time periods are sorted into terciles based on the 10 year minus 1 year KR yield spreads. The returns are averaged over time from November 1971 to December 2013. Figure A.6 in the appendix shows the results for different numbers of KR factors for the full sample of daily data from June 1961 to December 2020. The average one-business day returns are annualized.

terciles. Bad times require more complex curvature patterns and alternative methods are unstable and/or omit flexible patterns. The estimates of imprecise approaches like GSW and NSS are very similar to the KR-1 or KR-3 model, that is, the too simplistic form of parametric models can be interpreted as omitting important term structure factors. The returns implied by FB and LW are unstable during recessions or low and medium yield spread terciles. The benchmark methods have systematic biases in particular for maturities larger than 6 years, which are the most pronounced during recessions and low and medium yield spread periods. As bad economic states require more complex curvature patterns, these are also the times when alternative benchmarks methods are either the most unstable or lack flexibility. This suggests, that the important fourth factor of our method requires a precise and robust method to detect it in the first place.

Figure 14: Cumulative excess returns of implied SDF



(a) In-sample cumulative excess returns



(b) Out-of-sample cumulative excess returns

The figure shows the in-sample and out-of-sample cumulative excess returns of mean-variance efficient portfolios formed by 1 to 4 KR factors. The in-sample mean-variance efficient portfolio weights are estimated using the full panel of daily data from June 1961 to December 2020. The out-of-sample mean-variance efficient portfolio weights are estimated on a 10-year rolling window and are re-balanced monthly. Shaded areas indicate recession periods.

4.3 Stochastic Discount Factor with Complexity Premium

Asset pricing is fundamentally a problem about spanning asset returns and replicating the components of the payoffs that are compensated for risk with a pricing kernel. We have first shown that four tradable KR factors explain almost all variation in coupon and discount bond returns. Second, we have also demonstrated that these factors also very precisely span the mean returns of discount bonds. Importantly, our KR factors and discount bonds are investable portfolios. This is distinctively different from all other term structure estimation approaches, where the discount bond returns are based on estimated "artificial" prices. In conclusion, the four KR factors provide tradable assets that are sufficient to replicate any fixed income claim and thus serve as basis assets for the stochastic discount factor (SDF).

We construct the SDF projected on different candidate factor models. This is equivalent to constructing the mean-variance efficient tangency portfolio based on tradable factors. The Sharpe ratio of the tangency portfolio is a direct measure of the pricing information that is captured by the factors, as well as a measure for the profitability of trading the KR factors.

Table 2: Sharpe ratio of implied SDF

	Out-of-sample				In-sample			
	KR-1	KR-2	KR-3	KR-4	KR-1	KR-2	KR-3	KR-4
	Unconditional							
	0.51	0.69	0.75	0.85	0.47	0.67	0.74	0.84
	Conditional on boom and recessions							
Boom	0.46	0.53	0.48	0.63	0.31	0.38	0.37	0.54
Recession	0.15	0.68	0.77	1.65	0.59	0.99	1.26	1.99
	Conditional on yield spread terciles							
Low yield spread	0.21	0.39	0.33	0.50	-0.21	0.04	0.04	0.39
Medium yield spread	0.21	0.43	0.48	1.18	0.53	0.74	0.88	1.38
High yield spread	0.71	0.75	0.74	0.93	0.81	0.90	0.99	1.22

This table shows the out-of-sample and in-sample annualized Sharpe ratios of the mean-variance efficient portfolio formed by 1 to 4 KR factors conditioned on NBER boom or recession periods and yield spread terciles. The time periods are sorted into terciles based on 10 year - 1 year KR yield spreads. For different number of KR factors, we form the mean-variance efficient portfolio either on the full-sample (in-sample) or on a rolling window using past observations (out-of-sample) and evaluate the performance of the portfolio during different time periods. The out-of-sample mean-variance efficient portfolio weights are estimated on a 10-year rolling window and are re-balanced monthly using the full panel of daily data from June 1961 to December 2020.

The KR factors imply profitable investment opportunities. We construct the implied SDFs based on the first n KR factors and compare the Sharpe ratios of the candidate SDFs. Table 2 show that the fourth factor boosts the Sharpe ratio out-of-sample to 0.85, emphasizing again its important asset pricing implications. The in- and out-of sample Sharpe ratios do not increase further when using more than four factors as shown in Table A.3.

The complex fourth term structure factor is a hedge for bad economic times and pays off during recessions. It can be used in trading strategies to hedge against recession risk, and omitting the fourth factor results in worse investment performance during recessions. In order to illustrate this point, Figure 14 shows the in-sample and out-of-sample cumulative returns of mean-variance efficient portfolios formed by 1 to 4 KR factors. As before, the in-sample mean-variance efficient portfolio weights are estimated using the full panel, while the out-of-sample mean-variance efficient portfolio weights are estimated on a 10-year rolling window and are re-balanced monthly. The shaded areas indicate recession periods. We have three main findings. First, the investment patterns for the in-sample and out-of-sample analysis are qualitatively very similar, confirming that our results are stable and do not suffer from overfitting. Second, the largest increase in cumulative returns is due to including the fourth factor. Third, a large part of this increase in performance from including the fourth factor occurs during recession periods. This motivates the economic label for the fourth KR factor as a hedging factor for bad times. As the fourth factor captures complex term structure shapes, its premium can be understood as a complexity premium, which pays off during recessions.

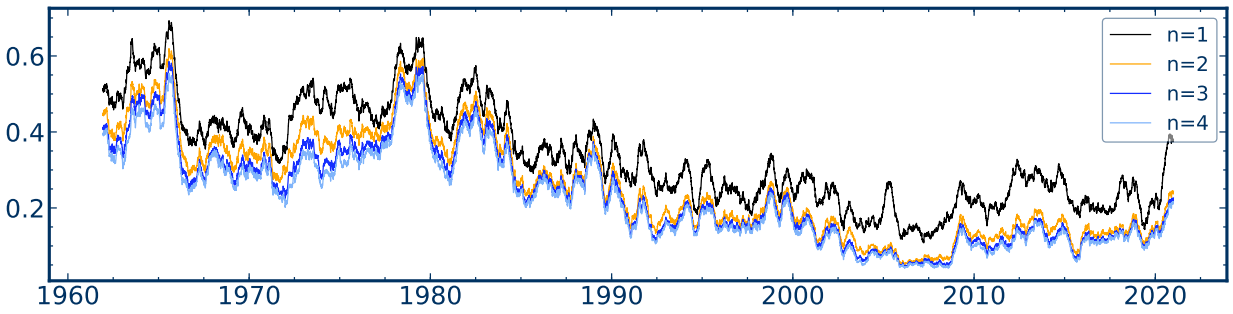
Table 2 shows the out-of-sample and in-sample annualized Sharpe ratios of the implied SDF based on one to four KR factors conditioned on either NBER boom or recession periods or yield spread terciles. For different number of KR factors, we form the mean-variance efficient portfolio either on the full-sample (in-sample) or on a rolling window using past observations (out-of-sample) and evaluate the performance of the portfolio during different time periods. The out-of-sample mean-variance efficient portfolio weights are estimated on a 10-year rolling window and are re-balanced monthly. We observe a substantial increase in Sharpe ratios from including the fourth factor during bad times. During recessions, the fourth factors doubles the out-of-sample Sharpe ratio relative to a three-factor model and increases the Sharpe ratio by a factor of ten relative to a slope factor. The Sharpe ratios during booms are very similar for different number of factors. The same findings hold for the in-sample analysis. When evaluating the implied SDF during different yield spread periods, we obtain an analogous result. The largest boost in Sharpe ratio occurs from including the fourth factor during low and medium yield spread time periods. During high yield spread time periods, the number of factors has only a minor effect on the Sharpe ratio, and a simple slope factor captures most of the information.

4.4 Bond Market Complexity over Time

The exposure to term structure risk factors provides a measure for the state of the bond market. The cross-sectional variation that is explained by our term structure risk factors is time-varying and can be informative about economic conditions. In a panel we can measure the relative importance of factors based on the eigenvalues of covariances matrices. In contrast, given our conditional factor structure, we do not require a time-series to obtain factor loadings. In our case, we can calculate the cross-sectional importance of different factors for each single day. In Figure 15 we show the 120-day moving average (MA) of the unexplained variation for the KR-1 to KR-4 factor models over time. First, we observe that the explained variation in coupon bond excess returns is time-varying. The amount of unexplained variation is generally higher in the first half of the sample. Second, we also observe that the relative importance of the factors is time-varying as well. For example at the end of 2000 around half of the explained variation is captured by KR factors 2 to 4, while in 1979, almost all variation is explained by the first KR factor. Having established in the previous sections the need for four KR factors, we study here the time variation in the cross-sectional fit and relative importance of higher order factors.

We introduce two novel measures for the state of the bond market: IT-VOL and T-COM. The Idiosyncratic Treasury Volatility (IT-VOL) measures the idiosyncratic volatility normalized by the overall volatility. It captures how hard it is to explain the observed bond returns even with a flexible model. The Treasury Market Complexity (T-COM) measures the complexity of the bond market. It captures how much variation is explained by more complex, non-linear return curves. Formally, we define IT-VOL as the percentage of unexplained variation by the KR-4 factor model, which corresponds to the light blue line in Figure 15. This means that IT-VOL is one minus the

Figure 15: Unexplained variation by the first 4 KR factors



This figure shows the 120-day moving average of the unexplained cross-sectional variation for 1 to 4 KR factors. The unexplained cross-sectional variation of a factor model is normalized by the overall cross-sectional variation on that day.

cross-sectional R^2 of a four-factor model. T-COM is essentially[26] equal to the difference between the explained variation of the KR-1 and KR-4 models, which corresponds to the difference between the black and light blue line in Figure 15 :

$$\text{IT-VOL}_t = \frac{\left\| R_t^{\text{bond}} - \sum_{k=1}^{4} \beta_{t-1,k}^{\text{bond}} F_{t,k} \right\|_2^2}{\left\| R_t^{\text{bond}} \right\|_2^2}, \qquad \text{T-COM}_t = \frac{\left\| \sum_{k=2}^{4} \beta_{t-1,k}^{\text{bond}} F_{t,k} \right\|_2^2}{\left\| R_t^{\text{bond}} \right\|_2^2}.$$

Figure 16 plots the 120-day moving average time-series of IT-VOL and T-COM. It is obvious that they capture different time-series information. The plot suggests that T-COM seems to be a better indicator for times of macroeconomic distress. These are the time periods when the term structure has more curvature and an estimate based on a simple slope factor would be strongly misspecified.

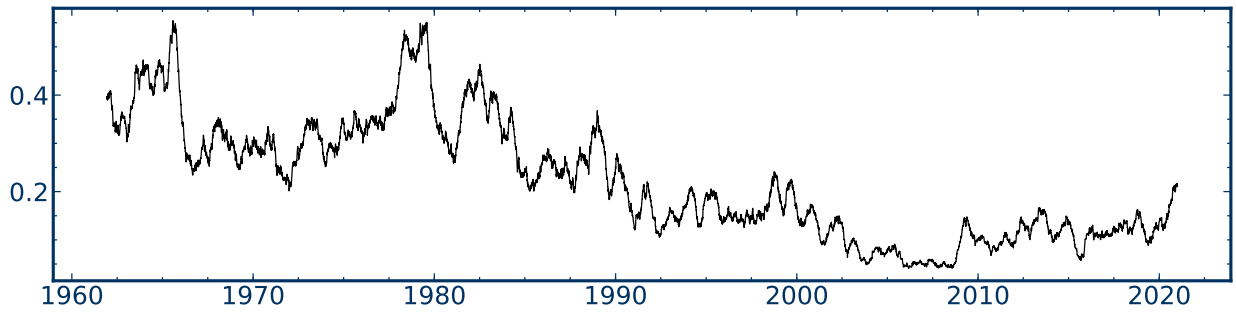
The examples in Figure 17 illustrate the meaning of the two measures. In Figure 17 we plot the observed and estimated Treasury excess returns for two distinctively different days. On 1979-01-03 returns are very volatile and are not well explained even with four factors. This is captured by the large idiosyncratic volatility measure IT-VOL of 0.08. The complexity of the term structure is small, as indicated by T-COM=0.01. In contrast, 2005-03-03 requires a term structure with more curvature, and hence has a higher complexity of T-COM=0.11. A linear slope factor alone does not explain the term structure well. The fluctuation around the fitted returns is small. Thus, IT-VOL takes a low value of 0.003.[27]

The measures of bond market conditions directly relate to real economic conditions. The bond market complexity predicts future unemployment rates. More specifically, higher complexity in the term structure is correlated with higher future unemployment rates. Figure A.9 compares the times-series of T-COM with unemployment rates. Given the lower frequency of unemployment rates, we take a local average of T-COM using the prior 120 business day observations of T-COM.

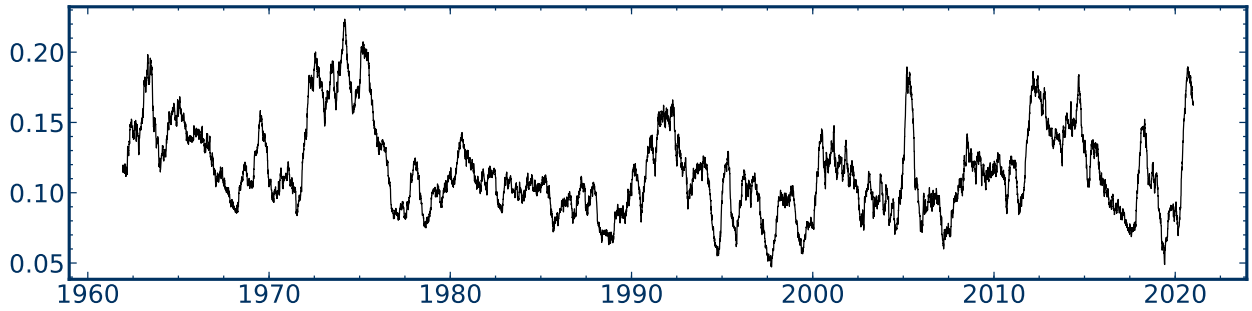
[26] It would be exactly equal if the columns of $\beta_{t-1}^{\text{bond}}$ were orthogonal.

[27] Our novel measures also detect high-frequency changes in bond markets. While Figure 16 shows the average time-series behavior of the market condition measures, the Internet Appendix illustrates the two measures and the fitted coupon bond excess returns for a time window around example days, and shows that we can capture sudden changes.

Figure 16: Time-Series of Market Condition Measures



(a) Idiosyncratic Treasury Volatility measure (IT-VOL)



(b) Treasury Market Complexity measure (T-COM)

This figure shows the 120-day moving average of the two market condition measures. Subfigure (a) displays the Idiosyncratic Treasury Volatility measure (IT-VOL), which is the unexplained variation by the 4-factor KR model. Subfigure (b) presents the Treasury Market Complexity measure (T-COM), which is essentially the difference between unexplained variation by the 1- and 4-factor KR model. The unexplained cross-sectional variation of a factor model is normalized by the overall cross-sectional variation on that day.

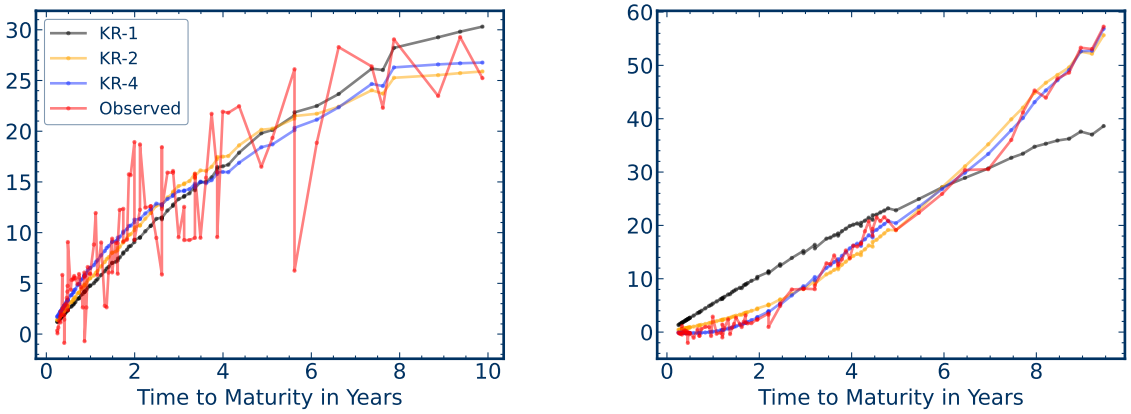
The results are robust to this choice. The top figure shows the two time-series of the 120-day moving average of T-COM at daily frequency and the unemployment rate at monthly frequency. The red line corresponds to the unemployment rate, and the blue line is the T-COM measure. We observe a clear co-movement in the level of both time-series. The bottom figure displays the changes in the moving average of T-COM and of the monthly unemployment rate relative to their values in the previous year. The changes are at monthly frequency. It shows that changes in T-COM precede changes in unemployment rates, which is particularly visible for the spikes in the time-series.

Changes in T-COM predict changes in unemployment rate one year ahead. We calculate the correlation of changes in T-COM and the unemployment rate relative to last year for different lags. The largest correlation of 0.21 occurs at the 14-month lag. This means that changes in T-COM have the highest correlations with changes in unemployment rates roughly one year in the future. This prediction is statistically significant at a 1% level.[28]

The idiosyncratic volatility in bond markets is related to the volatility in equity markets. Figure A.10 compares the times-series of IT-VOL with the VIX index. The VIX is the Chicago Board Options Exchange (CBOE) volatility index, and measures market expectation of near term volatility

[28]The Internet Appendix collects the detailed results.

Figure 17: Market Conditions on Example Days



(a) 1979-01-03
IT-VOL: 0.08, T-COM: 0.01

(b) 2005-03-03
IT-VOL: 0.003, T-COM: 0.11

This figure shows fits of coupon bond excess returns by KR factor models with 1, 2, and 4 factors on two example dates: 1979-01-03 (left) and 2005-03-03 (right). Numbers are in basis points. On 1979-01-03, IT-VOL takes a high value, while the value of T-COM is relatively small. The reason is that return observations are noisy, but the term structure shape is not very complex. On the other hand, on 2005-03-03, IT-VOL is small while T-COM is large. This is because return observations are not very noisy, but the term structure shape has a large deviation from the straight line suggested by the first KR factor.

Table 3: KR factors conditioned on IT-VOL and T-COM

	SR				Mean				Standard Deviation			
	F_1	F_2	F_3	F_4	F_1	F_2	F_3	F_4	F_1	F_2	F_3	F_4
IT-VOL												
Tercile 1	0.70	0.19	-0.14	-0.32	24.49	12.55	-11.29	-37.19	34.76	64.51	82.16	115.16
Tercile 2	0.38	0.19	0.01	0.53	9.50	22.44	1.20	94.23	25.01	115.48	189.98	177.92
Tercile 3	-0.53	0.34	0.11	0.59	-6.55	29.16	18.6	112.72	12.25	84.94	166.36	189.58
T-COM												
Tercile 1	1.03	-0.01	-0.19	-0.35	34.27	-0.37	-16.15	-40.43	33.22	44.39	87.20	117.09
Tercile 2	0.04	0.23	0.09	0.57	0.85	18.88	12.88	90.65	23.91	83.82	146.66	158.06
Tercile 3	-0.44	0.36	0.06	0.58	-7.68	45.64	11.77	119.54	17.50	125.35	203.49	205.33

This table shows the annualized Sharpe ratio, mean, and standard deviation of the four KR factors conditioned on terciles of the Idiosyncratic Treasury Volatility (IT-VOL) measure and the Treasury Market Complexity (T-COM) measure. Mean and standard deviation are annualized and in basis points. The sample is daily data from June 1961 to December 2020.

captured by stock index option prices. The left axis and blue line corresponds to the log VIX, while the right axis and red line captures IT-VOL. Both volatility measures co-move together. Periods of high stock market volatility, for example during the second half of 2008 or the beginning of March 2020, also imply large cross-sectional idiosyncratic volatility in bond markets. While a connection between equity and bond markets is not surprising, the dependencies between the measures represents still an interesting finding. IT-VOL is not a time-series volatility, but based on cross-sectional variation in bond returns. The implied volatility measured by the VIX is a proxy

for the spot volatility of the market return time-series. While the volatility measures for both asset markets are related, there seem to be also differences, which suggests that they capture distinct information.

The changes in bond market conditions detected by our measures impact the term structure risk premium. Table 3 shows the annualized Sharpe ratio, mean, and standard deviation of the four KR factors conditioned on the terciles of T-COM and IT-VOL. The risk premium of the first (slope) factor is strongly affected by T-COM and IT-VOL. Remarkably, the sign of the risk premium of the first KR factor changes from positive to negative. This means that the first factor earns its risk premium during times when a linear curve explains bond returns well. Conversely, it "loses" this premium during complex or noisy market conditions. In contrast, the fourth KR factor serves as a hedge against the risk of changing market conditions. It has a negative payoff during low T-COM and IT-VOL periods, but earns a high positive risk premium during challenging times.

The first and fourth KR factors exhibit the largest conditional and unconditional Sharpe ratios. The second and fourth factors carry only a small risk premium as reflected by the lower Sharpe ratios. We emphasize that the dependency of the mean and Sharpe ratio of the KR factors on T-COM and IT-VOL is not mechanical and is evidence that these measures capture a source of risk in the bond market. The time-variation in explained variation of the risk factors is mechanically linked to the standard deviation of the factors. The standard deviation of higher order factors increases when the complexity goes up. However, the effect on the magnitude and the sign of the mean of the factors as well as the the effect on the Sharpe ratio is not mechanical. In fact, it is an important empirical finding that times of more complex term structures require a higher conditional risk premium for higher order factors.

Figures A.11 and A.12 provide a refinement of the results in Table 3. These figures show the non-parametric estimation with local kernel smoothing of the conditional Sharpe ratio, mean and standard deviation.[29] The mean returns and standard deviation are normalized by their unconditional values, that is a value of 1 corresponds to the unconditional values. We condition on the logarithm of IT-VOL and T-VOL. The vertical dotted lines represent the tercile values of IT-VOL and T-COM, which are used in Table 3. We observe that our novel T-COM measure has a stronger effect on the risk premium and Sharpe ratio of the first KR factor. Hence, the complexity of bond markets is crucial to assess the term structure premium of long maturity over short maturity bonds. We also observe, while T-COM and IT-VOL have similar effects on the first and fourth factor, they are capturing different information as illustrated by the second and third factor.[30]

Our results complement prior findings on the dependency of the term structure premium on

[29] Conditional variance is defined as conditional second moment minus the squared conditional mean.

[30] The results are robust to alternative formulations of T-COM and IT-VOL based on the full KR model. In the Internet Appendix we show the results when the IT-VOL and T-COM measures are constructed with the full KR model, that is, the errors for the unexplained variation are based on the full KR model. Similarly, the alternative complexity measure is based on the normalized difference in cross-sectionally explained variation between the first KR factor and the full KR model instead of only the first four KR factors. The results are virtually identical, showing that our findings do not depend on the choice of the number of factors.

market conditions. We show that the sign of the term structure premium captured by the first (slope) factor switches based on challenging bond market conditions. This shares some similarity with inverted return curves during recessions. The inversion of the slope can be at least partly hedged by the convexity effects of higher order risk factors. This suggests that it is crucial to consider all relevant term structure factors when studying the impact of market conditions on the return curve.

4.5 Cross-Sectional Asset Pricing

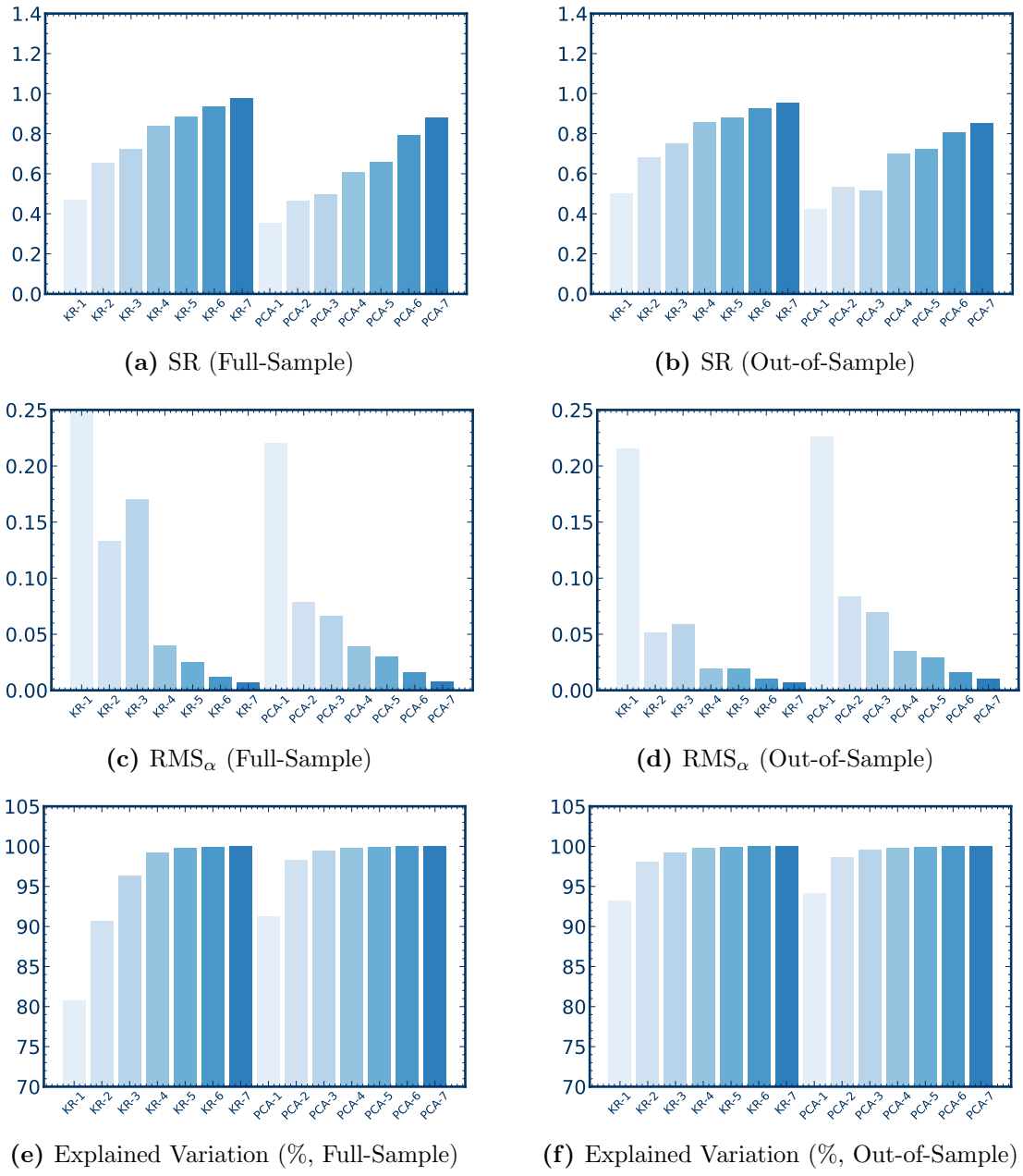
Lastly, we study the cross-sectional asset pricing implications of different SDFs for the term structure in a conventional panel setup. We use this standard asset pricing setup with common metrics for three reasons. First, we want to demonstrate the pricing results in a familiar setup that allows us to compare different factor models. Second, we want to emphasize the dependence on the test assets, and clarify that imprecisely estimated discount bonds affect the asset pricing results. Third, we consider directly the SDF as the tangency portfolio for pricing the test assets. This is achieved by construction when we run time-series regressions on factors to obtain the pricing errors, instead of using the cross-sectional conditional factor loadings.

We use the conventional asset pricing setup for panels. The exposure to risk factors is estimated from time-series regressions. Our test asset panel consists of daily returns of 10 discount bonds estimated with the full KR method for yearly maturities ranging from one to ten years. We compare how well different sets of factors explain the risk premium and variation of the test assets. The factors include our KR factors, the PCA factors estimated on the panel, and the Fama–French 3 and 5 factors. We report the Sharpe ratios of the implied SDF (SR), the root-mean-squared pricing errors of the term structure (RMS_α), and the explained variation. The in-sample pricing errors are obtained as the intercept of the time-series regression on the factors. The out-of-sample pricing errors follow from a cross-sectional regression, that is, they are the mean of residuals of a regression on cross-sectional loadings estimated on a rolling window of past observations. The explained variation equals the model implied variation normalized by the total variation, that is, the average in- or out-of-sample squared residuals normalized by the average squared excess return. Overall, this is a standard cross-sectional asset pricing analysis similar to Lettau and Pelger (2020b). The sampling period of the comparison study is from July 1963 to December 2020, since the data for daily Fama–French 5 factors only starts in July 1963.

Figure 18 summarizes the asset pricing results. Panels (c) and (d) show that we require four KR factors to obtain small pricing errors in- and out-of-sample. These results are consistent with the Sharpe ratios for the factors. Similarly, four PCA factors strongly lower the pricing errors. A large part of the variation is already explained by only two factors. Therefore, higher order factors like the fourth KR or PCA factor can be interpreted as "weak" factors in this panel. This means they explain only a small part of the variation, but carry important pricing information.

In Figure 19, we study in more detail the pricing errors for the ten discount bonds. We report the in-sample pricing errors from time-series regressions for each of the ten discount bonds. We also

Figure 18: Asset Pricing Results



(a) SR (Full-Sample)

(b) SR (Out-of-Sample)

(c) RMS_α (Full-Sample)

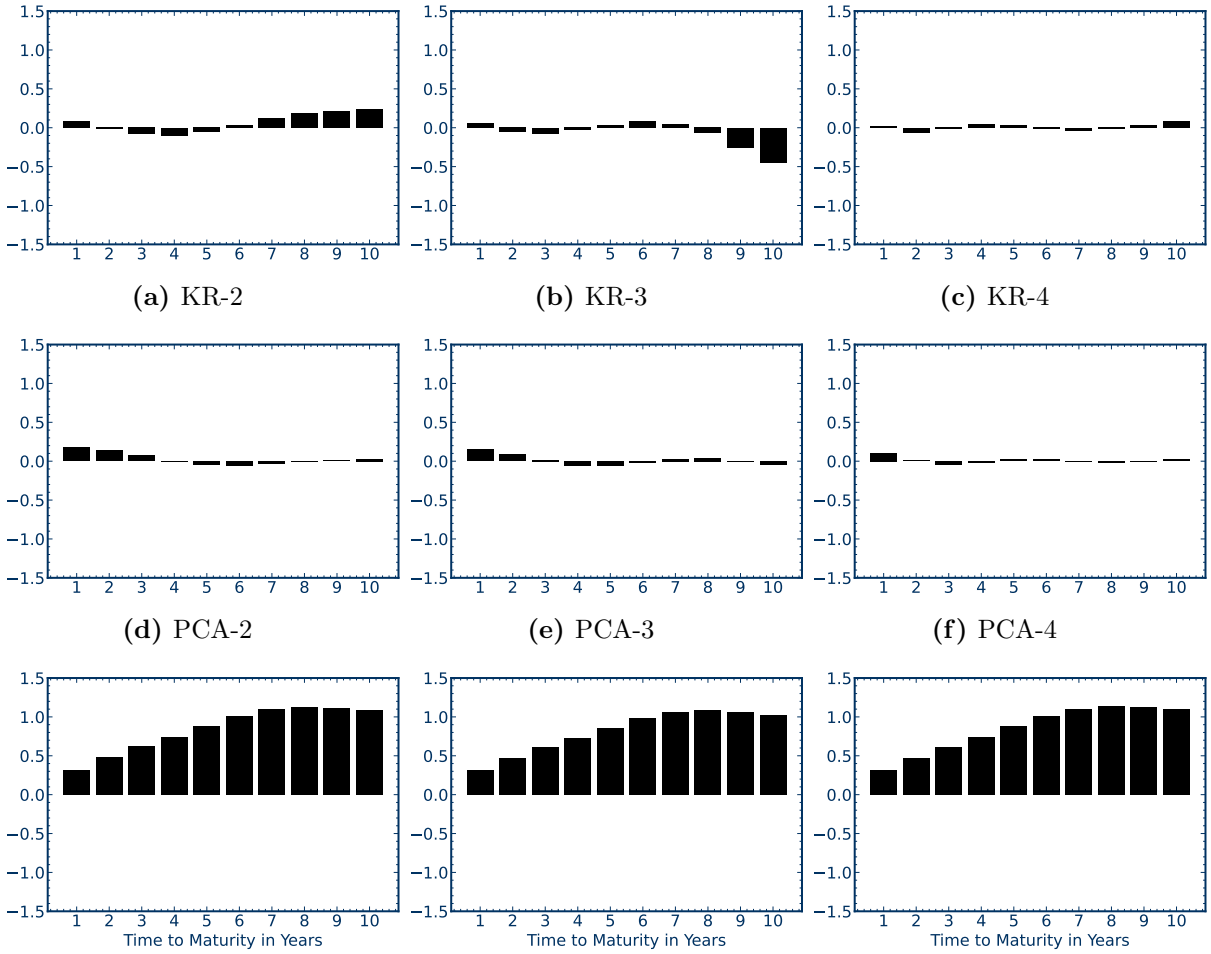
(d) RMS_α (Out-of-Sample)

(e) Explained Variation (%, Full-Sample)

(f) Explained Variation (%, Out-of-Sample)

This figure summarizes the asset pricing results for different term structure factors on a panel of discount bond excess returns. We compare the the first seven KR and PCA factors in- and out-of-sample in terms of their Sharpe ratios, root-mean-squared pricing errors and explained variation. The test assets are the panel of 10 discount bond excess returns estimated with full KR with yearly maturities ranging from 1 to 10 years. The PCA factors are extracted from this panel of 10 test assets. The annualized Sharpe ratios of the implied SDF are based on the mean-variance efficient combination of successively adding factors. The annualized pricing errors are the intercepts in time-series regressions of the discount bond excess returns on factors. The explained variation is the percentage of variation spanned by the factors in the time-series regressions. Out-of-sample results are obtained from a regression and moments on a monthly updated 10-year rolling window. The out-of-sample pricing errors are estimated with a cross-sectional Fama Macbeth type regression with loadings from past data. The risk-free one-business day return is the estimate from the KR discount curve. The sample is daily data from July 1963 to December 2020.

Figure 19: Pricing errors of discount bonds for different factors



(a) KR-2 **(b)** KR-3 **(c)** KR-4

(d) PCA-2 **(e)** PCA-3 **(f)** PCA-4

(g) Mean Excess Returns **(h)** FF-3 **(i)** FF-5

This figure shows the time-series pricing errors (in basis point) from time-series regressions of 10 discount bond excess returns on different factors. The test assets are the panel of 10 discount bond excess returns estimated with KR with yearly maturities ranging from 1 to 10 years. We compare KR, PCA and Fama–French factors. The PCA factors are extracted from the panel of 10 discount bond excess returns. The Fama–French 3 and 5 factors are taken from Kenneth French's website. The risk-free one-business day return is the estimate from the KR discount curve. The sampling period is from July 1963 to December 2020.

report the term structure premium as the mean of the excess returns. The two-factor models KR-2 and PCA-2, which correspond to the common slope and curvature (as the level is taken out when we work with excess returns), result in non-negligible pricing errors for very short or long maturity bonds. However, the four factor models KR-4 and PCA-4 price the complete panel extremely well. The Fama–French 3 and 5 factors do not explain the term structure premium. In fact, the pricing errors are close to the mean returns of the bonds. Table A.2 in the appendix shows that pricing errors with 4 KR factors are not only economically small, but for most test assets statistically insignificant.

The asset pricing results on an alternative cross-section highlight that cross-sectional asset

pricing and term structure estimation are a joint problem. In Table A.2 and Figure A.3 we use a panel of 10 discount bonds estimated with GSW as test assets and to construct the PCA factors. We observe that fewer factors already result in smaller pricing errors. This result is baked into the over-simplistic discount curve estimated with GSW. We only need two factors (which correspond to the three factors level, slope and curvature in the return space) to price the complete cross-section. Table A.2 shows that pricing errors with 2 KR factors are already insignificant. This is not surprising, as the GSW test assets neglect important term structure patterns. In conclusion, our KR factors price the KR and GSW test assets. As the KR test assets better reflect the underlying term structure information, they require four factors. Our results are robust to the choice of the risk-free one-business day return.

5 Conclusion

We introduce a conditional factor model for the term structure of Treasury bonds, which unifies non-parametric curve estimation with cross-sectional asset pricing. Our robust, flexible and easy-to-implement method learns the return curve directly from observed returns of Treasury bonds. This curve lies in a reproducing kernel Hilbert space, which is based on economic first principles, and optimally trades off smoothness against return fitting. We show that a low dimensional factor model arises because a sparse set of basis functions spans the estimated return curves. The estimated factors are investable portfolios of traded assets, which replicate the full term structure and are sufficient to hedge against interest rate changes.

In an extensive empirical study on U.S. Treasury bonds, we show that the return curve is well explained by four factors, which capture polynomial shapes of increasing order and are necessary to explain the term structure premium. We show that our factors are essentially the same as PCA factors obtained from a panel of discount bond excess returns. However, importantly, our factors are not only available for any single day without the need for panel data, and we can also explain why they occur. We identify smoothness as a fundamental principle of the term structure of returns. Our latent factors correspond to the optimal sparse set of non-parametric basis functions in decreasing order of smoothness to approximate smooth curves. The structure of PCA factors is a consequence of the nature of our estimated curves. The cash flows of coupon bonds fully explain the factor exposure, and play the same role as firm characteristics in equity modeling. In this sense, "cash flows are covariances". We introduce a new measure for the time-varying complexity of bond markets based on the exposure to higher-order factors, and show that changes in market complexity affect the term structure premium. The "complex" fourth term structure factor is a hedge for bad economic times and pays off during recessions. It can be used in trading strategies to hedge against recession risk, and omitting the fourth factor results in worse investment performance during recessions. This also provides an economic label for the fourth factor as a hedging factor for bad times, and its premium can be interpreted as a *complexity premium*.

We provide publicly available and regularly updated data sets of daily discount bond returns

and factors based on our precise estimates. Our zero-coupon bonds and factors are investable portfolios of traded U.S. Treasury bonds. Our shared data also includes the complexity and pricing error measures over time. It is part of our larger discount bond database, which also includes the precise and robust yield estimates of discount bonds for all maturity ranges.

References

Bauer, M., and J. D. Hamilton (2018): "Robust Bond Risk Premia," *Review of Financial Studies*, 31(2), 399–448.

Cieslak, A., and P. Povala (2015): "Expected Returns in Treasury Bonds," *Review of Financial Studies*, 28, 28592901.

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

Cooper, I., and R. Priestley (2008): "Time-Varying Risk Premiums and the Output Gap," *Review of Financial Studies*, 22, 2801–2833.

Crump, K. R., and N. Gospodinov (2022): "On the factor structure of bond returns," *Econometrica*, 90(1), 295–314.

Dai, Q., and K. Singleton (2002): "Expectation puzzles, time-varying risk premia, and affine models of the term structure," *Journal of Financial Economics*, 63, 415–441.

Duffee, G. (2013): "Forecasting Interest Rates," *Handbook of Economic Forecasting*, 2A.

Duffie, D., D. Filipović, and W. Schachermayer (2003): "Affine processes and applications in finance," *Annals of Applied Probability*, 13, 984–1053.

Duffie, D., J. Pan, and K. Singleton (2000): "Transform analysis and asset pricing for affine jump-diffusions," *Econometrica*, 68, 1343–1376.

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.

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

Fama, E. F., and K. R. French (1993): "Common risk factors in the returns on stocks and bonds," *Journal of Financial Economics*, 33(1), 3–56.

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

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): "Stripping the Discount Curve – a Robust Machine Learning Approach," *Working paper*.

Greenwood, R., and D. Vayanos (2014): "Bond Supply and Excess Bond Returns," *Review of Financial Studies*, 27, 2291–2304.

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.

Kelly, B., S. Pruitt, and Y. Su (2019): "Characteristics Are Covariances: A Unified Model of Risk and Return,," *Journal of Financial Economics*, 134, 501–524.

Kozak, S., S. Nagel, and S. Santosh (2020): "Shrinking the Cross Section," *Journal of Financial Economics*, 135(2), 271–292.

Lettau, M., and M. Pelger (2020a): "Estimating Latent Asset-Pricing Factors," *Journal of Econometrics*, 318(1), 1–31.

——— (2020b): "Factors that Fit the Time-Series and Cross-Section of Stock Returns," *Review of Financial Studies*, 33(5), 2274–2325.

Litterman, R., and J. Scheinkman (1991): "Common Factors Affecting Bond Returns," *Journal of Fixed Income*, 1(54-61).

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

Ludvigson, S., and S. Ng (2009): "Macro Factors in Bond Risk Premia," *Review of Financial Studies*, 22(5027-5067).

——— (2010): "A Factor Analysis of Bond Risk Premia," *Handbook of Empirical Economics and Finance*, 313.

Pelger, M. (2020): "Understanding Systematic Risk: A High-Frequency Approach," *Journal of Finance*, 75(4), 2179–2220.

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

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

A Theory and Proofs

This appendix contains further theoretical details and the proofs for all the theorems and statements in the main text.

A.1 Recap of Reproducing Kernel Hilbert Space $\mathcal{H}_\alpha$

We first recap a the general definition of a reproducing kernel Hilbert space, which is a fundamental notion in statistical machine learning. For more background and applications we refer the reader to, e.g., Rasmussen and Williams (2006); Hastie, Tibshirani, and Friedman (2009). In functional analysis, a reproducing kernel Hilbert space (RKHS) is a Hilbert space of functions in which point evaluation is a continuous linear functional. Roughly speaking, this means that if two functions in the RKHS are close in norm, then they are also pointwise close. An RKHS is associated with

a kernel that reproduces every function in the space in the sense that for every "x" in the set on which the functions are defined, "evaluation at x" can be performed by taking an inner product with a function determined by the kernel.

The formal definition reads as follows. Let E be an arbitrary set and $\mathcal{H}$ a Hilbert space of functions $h : E \to \mathbb{R}$. Then $\mathcal{H}$ is called a *reproducing kernel Hilbert space* 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.

Reproducing kernel Hilbert spaces are particularly important in the field of statistical learning theory because of the celebrated representer theorem which states that every function in an RKHS that minimizes an empirical loss functional can be written as a linear combination of the kernel function evaluated at the training points. This is a practically useful result as it effectively simplifies the empirical loss minimization problem from an infinite dimensional to a finite dimensional optimization problem. Our results are a variant thereof.

Our hypothesis space $\mathcal{H}_\alpha$ is the reproducing kernel Hilbert space containing all twice weakly differentiable functions $h : [0, \infty) \to \mathbb{R}$ with $h(0) = 0$ and $\lim_{x\to\infty} h'(x) = 0$ with finite norm $\|h\|^2_{\mathcal{H}_\alpha} < \infty$ given by the smoothness measure (6). Filipović, Pelger, and Ye (2022) show that $\mathcal{H}_\alpha$ is a RKHS and its reproducing kernel $k : [0, \infty) \times [0, \infty) \to \mathbb{R}$ of $\mathcal{H}_\alpha$ is given in closed form as

$$k(x, y) = -\frac{\min\{x, y\}}{\alpha^2} \mathrm{e}^{-\alpha \min\{x,y\}} + \frac{2}{\alpha^3} \left(1 - \mathrm{e}^{-\alpha \min\{x,y\}}\right) - \frac{\min\{x, y\}}{\alpha^2} \mathrm{e}^{-\alpha \max\{x,y\}}. \tag{15}$$

Filipović, Pelger, and Ye (2022) also derive a Bayesian distribution theory, which directly carries over to this setup and provides valid confidence intervals for our curve estimators.

A.2 Return Curve is Element of $\mathcal{H}_\alpha$ (Proposition 1)

The following proposition is the formal statement of Proposition 1 and shows that, under the absence of arbitrage and a technical condition, $\mathcal{H}_\alpha$ is the appropriate hypothesis space for the return curve. The proof uses the theorem of Dybvig, Ingersoll, and Ross (1996), which states that long forward rates can never fall. Including all curves in $\mathcal{H}_\alpha$ makes our approach fully flexible and general.

Proposition A.1
Assume absence of arbitrage and that forward curves $f_t = -d'_t/d_t$ are regular in the sense that they flatten out and converge to a finite limit fast enough such that

$$\int_0^\infty f'_t(x)^2 \mathrm{e}^{\alpha x} \, dx < \infty. \tag{16}$$

Then the return curve r_t given in (3) *is an element in $\mathcal{H}_\alpha$, that is, $\|r_t\|^2_{\mathcal{H}_\alpha} < \infty$.*

Proof. We can assume that forward curves f_t are weakly differentiable. Indeed, Filipović, Pelger, and Ye (2022) show that arbitrage-free discount curves d_t are twice weakly differentiable, subject to mild technical first and second moments conditions on the short rates and its drift.

In the proof of (Filipović, 2001, Theorem 5.1.1) it is shown that (16) implies that the limit

$$f_t(\infty) = \lim_{x\to\infty} f_t(x) \text{ exists in } \mathbb{R}, \tag{17}$$

and the term spread function $\phi_t(x) = f_t(x) - f_t(\infty)$ is integrable such that

$$\int_0^\infty |\phi_t(x)|\, dx < \infty, \tag{18}$$

$$\int_0^\infty \phi_t(x)^4 \mathrm{e}^{\alpha x} dx < \infty. \tag{19}$$

The absence of arbitrage implies that long forward rates can never fall, according to Dybvig, Ingersoll, and Ross (1996), in the sense that

$$\Delta f_t(\infty) = f_t(\infty) - f_{t-1}(\infty) \geq 0. \tag{20}$$

Consequently, the return curve r_t is twice weakly differentiable and can be written as

$$\begin{aligned} r_t(x) &= \frac{1}{d_{t-1}(\Delta_t)} \left(\frac{d_{t-1}(\Delta_t) d_t(x)}{d_{t-1}(x+\Delta_t)} - 1 \right) = \frac{1}{d_{t-1}(\Delta_t)} \left(\mathrm{e}^{\int_{\Delta_t}^{x+\Delta_t} f_{t-1}(s)\, ds - \int_0^x f_t(s)\, ds} - 1 \right) \\ &= \frac{1}{d_{t-1}(\Delta_t)} \left(\mathrm{e}^{\int_0^x (f_{t-1}(s+\Delta_t) - f_t(s))\, ds} - 1 \right) \\ &= \frac{1}{d_{t-1}(\Delta_t)} \left(\mathrm{e}^{\int_0^x (\phi_{t-1}(s+\Delta_t) - \phi_t(s))\, ds} \mathrm{e}^{-\Delta f_t(\infty) x} - 1 \right). \end{aligned}$$

Differentiating gives its first and second order derivatives,

$$\begin{aligned} r_t'(x) &= \frac{1}{d_{t-1}(\Delta_t)} \left(\phi_{t-1}(x+\Delta_t) - \phi_t(x) - \Delta f_t(\infty) \right) \mathrm{e}^{\int_0^x (\phi_{t-1}(s+\Delta_t) - \phi_t(s))\, ds} \mathrm{e}^{-\Delta f_t(\infty) x}, \\ r_t''(x) &= \frac{1}{d_{t-1}(\Delta_t)} \left(\phi_{t-1}'(x+\Delta_t) - \phi_t'(x) + \left(\phi_{t-1}(x+\Delta_t) - \phi_t(x) - \Delta f_t(\infty) \right)^2 \right) \\ &\quad \times \mathrm{e}^{\int_0^x (\phi_{t-1}(s+\Delta_t) - \phi_t(s))\, ds} \mathrm{e}^{-\Delta f_t(\infty) x}. \end{aligned}$$

Combining this with properties (16), (17), (18), (19), and (20), and the elementary fact that

$$\begin{aligned} &\left(\phi_{t-1}'(x+\Delta_t) - \phi_t'(x) + \left(\phi_{t-1}(x+\Delta_t) - \phi_t(x) - \Delta f_t(\infty) \right)^2 \right)^2 \\ &\qquad \leq C \left(\phi_{t-1}'(x+\Delta_t)^2 + \phi_t'(x)^2 + \phi_{t-1}(x+\Delta_t)^4 + \phi_t(x)^4 + \Delta f_t(\infty)^4 \right), \end{aligned}$$

implies that $\lim_{x\to\infty} r_t'(x) = 0$ and $\int_0^\infty r_t''(x)^2 \mathrm{e}^{\alpha x} dx < \infty$. As $r_t(0) = 0$ by definition, this shows that $r_t \in \mathcal{H}_\alpha$, which completes the proof of Proposition A.1. □

A.3 Kernel Ridge Solution

We derive a closed-form solution to the curve estimation problem (7), which boils down to a kernel ridge regression with kernel (15). We first show that the kernel matrix $\boldsymbol{K} = (k(x_i, x_j))_{1 \le i,j \le N}$ is strictly positive definite. This is the content of the following lemma.

Lemma A.1
For the reproducing kernel Hilbert space $\mathcal{H}_\alpha$, for any points $0 < x_1 < \cdots < x_N$, the kernel functions $k(\cdot, x_1), \ldots, k(\cdot, x_N)$ are linearly independent. Consequently, the kernel matrix $\boldsymbol{K}$ is non-singular.

Proof of Lemma A.1. We argue by contradiction and assume that there exists a linear combination $h(\cdot) = \sum_{j=1}^N c_j k(\cdot, x_j)$ such that $h = 0$ in $\mathcal{H}_\alpha$ for some coefficients $c_1, \ldots, c_N$ such that $c_i \neq 0$ for some i. As $\mathcal{H}_\alpha$ contains all C^2-functions with compact support, there exists $\phi_i \in \mathcal{H}_\alpha$ such that $\phi_i(x_i) = 1$ and $\phi_i(x_j) = 0$ for all $j \neq i$. Then $c_i = \langle \phi_i, h \rangle_{\mathcal{H}_\alpha} = 0$, which is trivially wrong. Hence the first claim follows. The second part follows, because $\boldsymbol{K}$ is the Gram matrix of the linearly independent kernel functions. □

As a consequence of Lemma A.1, all eigenvalues of $\boldsymbol{K}$ shown in the singular value decomposition (8) are strictly positive, $s_N > 0$. We can thus define the N rotated basis functions

$$u_i(\cdot) = s_i^{-1/2} \sum_{j=1}^N k(\cdot, x_j) V_{ji}, \tag{21}$$

which are orthonormal in $\mathcal{H}_\alpha$. Indeed, by the reproducing kernel property of k, we obtain $\langle u_i, u_j \rangle_{\mathcal{H}_\alpha} = s_i^{-1/2} s_j^{-1/2} \sum_{s,t=1}^N V_{si} k(x_s, x_t) V_{tj} = 1$ if $i = j$ and zero otherwise, as desired. Using the expression in (15), the functions u_i in (21) are given in closed form.

To streamline the notation, we stack all non-zero cash flow dates into the column vector $\boldsymbol{x} = (x_1, \ldots, x_N)^\top$ and evaluate functions elementwise. That is, we denote $h(\boldsymbol{x}) = (h(x_1), \ldots, h(x_N))^\top$ for any function $h(\cdot)$. We can now state the functional-analytic counterpart to Theorem 1.

Theorem A.2 (Kernel–Ridge (KR) Solution)
The fundamental term structure estimation problem (7) *has a unique solution, which is given in closed form by*

$$\hat{r}_t(x) = \sum_{j=1}^N u_j(x) \hat{F}_{t,j} \tag{22}$$

where the factors $\hat{F}_t = \omega_{t-1} R_t^{\text{bond}}$ are the excess returns of traded bond portfolios given in (12)–(13). *The portfolio weights can also be expressed as*

$$\omega_{t-1} = \beta^\top Z_{t-1}^\top \left(Z_{t-t} \boldsymbol{K} Z_{t-1}^\top + \lambda M_t I_{M_t} \right)^{-1} \tag{23}$$

for the loadings $\beta = V S^{1/2}$ given in (9).

Proof of Theorem A.2 (Functional Representation). It is shown in (Filipović, Pelger, and Ye, 2022, Theorem A.1) that the unique solution of (7) can be represented as $\hat{r}_t(\cdot) = k(\cdot, \boldsymbol{x})^\top \tilde{F}_t$, for the coefficients given by $\tilde{F}_t = Z_{t-t}^\top \left(Z_{t-t} \boldsymbol{K} Z_{t-1}^\top + \lambda M_t I_{M_t}\right)^{-1} R_t^{\text{bond}}$. Note that we have $\hat{F}_t = S^{1/2} V^\top \tilde{F}_t$, for $\hat{F}_t = \omega_{t-1} R_t^{\text{bond}}$ with ω_{t-1} given in (23), which again is equivalent to (13) in view of (11) and Footnote 12. By definition (21), we have $u(\cdot) = (u_1(\cdot), \dots, u_N(\cdot))^\top = S^{-1/2} V^\top k(\cdot, \boldsymbol{x})$, so that we can write

$$\hat{r}_t(\cdot) = u(\cdot)^\top S^{1/2} V^\top \tilde{F}_t = u(\cdot)^\top \hat{F}_t,$$

which shows (22) and thus completes the proof. □

Given Theorem A.2, we can easily derive the conditional full factor model representation in Theorem 1.

Proof of Theorem 1 (Conditional Full Factor Representation). It follows by standard arguments as, e.g., in (Filipović, Pelger, and Ye, 2022, proof of Theorem A.1), that $\hat{F}_t$ given in (12)–(13) is the unique solution of the ridge regression problem (10). Alternatively, using (22) and that the functions u_j are orthonormal in $\mathcal{H}_\alpha$, we also note that $\|\hat{r}_t\|_{\mathcal{H}_\alpha} = \|\hat{F}_t\|_2$, which shows the equivalence of (7) and (10).

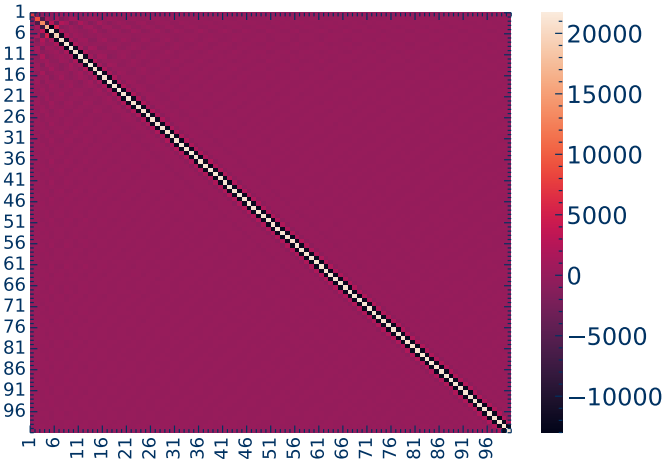
Equation (9) now follows from (22) and because $(u_1(\boldsymbol{x}), \dots, u_N(\boldsymbol{x})) = \boldsymbol{K} V S^{-1/2} = V S^{1/2} = \beta$. This completes the proof of Theorem 1. □

A.4 KR Basis Vectors (Proposition 2)

The basis functions V in (8) of the full KR estimator have the special property that their order corresponds to their smoothness. Hence, the first vector in V is the smoothest vector in $\mathbb{R}^N$, whereas the second vector is the second smoothest vector orthogonal to the first one, and so on. We now provide the formal function analytical statement for Proposition 2.

Our objective function in (7) rewards smoothness of the estimated return vector, where the smoothness measure on $\mathbb{R}^N$ is induced by the $\mathcal{H}_\alpha$-norm. Concretely, any vector $w \in \mathbb{R}^N$ can be represented by $w_j = h(x_j)$, $j = 1, \dots, N$, for a unique function h in the span of $k(\cdot, x_1), \dots, k(\cdot, x_N)$ in $\mathcal{H}_\alpha$. This function is given by $h = [k(\cdot, x_1), \dots, k(\cdot, x_N)] \boldsymbol{K}^{-1} w$. Its $\mathcal{H}_\alpha$-norm equals $\|h\|_{\mathcal{H}_\alpha} = \|\boldsymbol{K}^{-1/2} w\|_2$, where $\boldsymbol{K}^{-1/2} = V S^{-1/2} V^\top$ denotes the positive definite inverse square root of the kernel matrix. Hence $\|\boldsymbol{K}^{-1/2} w\|_2$ measures the smoothness of w; the smaller, the smoother. Indeed, Figure A.1 confirms that $\mathbf{K}^{-1/2}$ behaves approximately like a second-order difference operator on w with left boundary condition zero, that is, for extending w to the left by $(0, w)$. This also shows that the penalty term in the objective function (10), which can be written as $\|F_t\|_2^2 = \|\mathbf{K}^{-1/2} R_t\|_2^2$, just measures the smoothness of $R_t = \beta^\top F_t$.

Figure A.1: Inverse square root of kernel matrix $\boldsymbol{K}^{-1/2}$



This figure shows the upper left 100-by-100 submatrix of $\boldsymbol{K}^{-1/2}$, where $\boldsymbol{K}$ is the kernel matrix. The pattern in $\boldsymbol{K}^{-1/2}$ closely resembles a second-order difference operator with left boundary condition zero.

The smoothness measure scales linearly in w. Thus, to compare the smoothness of two vectors w_1 and w_2, they must be normalized. The unit vector w_1 in $\mathbb{R}^N$ with maximal smoothness solves

$$\min_{\|w\|_2=1} \|\boldsymbol{K}^{-1/2}w\|_2.$$

The solution is the normalized eigenvector of $\boldsymbol{K}$ with largest eigenvalue, $w_1 = v_1$.[31] The smoothest unit vector w_2 in $\mathbb{R}^N$ that is orthogonal to v_1 solves

$$\min_{\|w\|_2=1,\, w^\top v_1=0} \|\boldsymbol{K}^{-1/2}w\|_2.$$

The solution is the normalized eigenvector of $\boldsymbol{K}$ with second largest eigenvalue, $w_2 = v_2$. By induction, this shows that the eigenvector v_n is the smoothest unit vector in $\mathbb{R}^N$ that is orthogonal to $v_1, \ldots, v_{n-1}$.

A.5 KR Basis Vectors and PCA

The KR basis vectors V and the PCA of a panel of discount bond excess returns are empirically the same. Here we clarify that this is an economic and not mechanical finding.

In view of (9), the discount bond excess return vector R_t and the KR factors F_t are generically related such that $\mathrm{Cov}(R_t) = VS^{1/2}\,\mathrm{Cov}(F_t)S^{1/2}V^\top$. Therefore, the following properties are equivalent: (i) $S^{1/2}(F_t - \mathbb{E}[F_t])$ are the PCA factors of R_t; (ii) V are the normalized PCA loadings of R_t; (iii) $\mathrm{Cov}(F_t)$ is diagonal and the scaled variances $s_j\,\mathrm{Var}(F_{t,j})$ are descending in j. We show empirically that (i) and (ii) are approximately fulfilled. This also shows that these properties are very unlikely to hold for rotated factors if we represented the estimated vector R_t with respect to

[31] This equality holds modulo the sign, $w_1 = \pm v_1$. We assume that the eigenvalues are mutually distinct.

another orthonormal basis in $\mathbb{R}^N$.

Figure 4 reveals that the slope, curvature and higher order polynomial shapes frequently encountered in PCA are just the orthogonal smoothest vectors in decreasing order.[32] Importantly, this also applies to PCA of return curves obtained using other independent methods, including Liu and Wu (2021). We thus provide, for the first time in the literature, empirical evidence that smoothness, as captured by our measure, is a fundamental principle of Treasury bond returns. Any rotated factors and bases, other than KR, are unlikely to behave like PCA factors and loadings of the return curve.

A.6 Shrinking the Term Structure

A low-rank factor model corresponds to a sparse selection of the basis functions in Theorem 1. We show how to select an optimal set of sparse factors from the data by adding a LASSO selection penalty to the optimization problem in Theorem 1, which leads to (14).

The effect of LASSO and ridge shrinkage are closely related as, in our case, they both reward the dominating eigenvalues of the kernel space. The ridge shrinkage implies a "soft" shrinkage of small eigenvalues, while the "hard" shrinkage of LASSO sets non-selected eigenvalues to zero. In order to gain some intuition, we assume for illustration that all observed bonds are discount bonds, such that $R_t^{\text{bond}} = R_t$, $M_t = N$, $Z_{t-1} = I_N$, and $\beta_{t-1}^{\text{bond}} = \beta = VS^{1/2}$. In this case, the optimization problem (14) has a closed-form solution given by

$$\hat{F}_{t,j} = \left(1 + s_j^{-1}\lambda N\right)^{-1} s_j^{-1/2} \left(|v_j^\top R_t^{\text{bond}}| - s_j^{-1/2}\lambda_1 \frac{N}{2}\right)^+ \cdot \text{sign}(v_j^\top R_t^{\text{bond}}), \tag{24}$$

where $a^+ = \max\{a, 0\}$, and v_j denotes the jth eigenvector of $\boldsymbol{K}$, that is, the jth column of V. Indeed, as $\|R_t^{\text{bond}} - VS^{1/2}F_t\|_2^2 = \|V^\top R_t^{\text{bond}} - S^{1/2}F_t\|_2^2$, the objective function $\Phi(F_t)$ in (14) splits into $\Phi(F_t) = \sum_{j=1}^N \Phi_j(F_{t,j})$, with components $\Phi_j(F_{t,j}) \coloneqq \frac{1}{N}(v_j^\top R_t^{\text{bond}} - s_j^{1/2}F_{t,j})^2 + \lambda F_{t,j}^2 + \lambda_1|F_{t,j}|$. The function $\Phi_j(F_{t,j})$ is differentiable in all $F_{t,j} \neq 0$, with derivative $\Phi_j'(F_{t,j}) = -\frac{2}{N}s_j^{1/2}v_j^\top R_t^{\text{bond}} + \frac{2}{N}s_jF_{t,j} + 2\lambda F_{t,j} + \lambda_1\text{sign}(F_{t,j})$. It follows that $F_{t,j} = 0$ is optimal, all else equal, if and only if $-\lambda_1\frac{N}{2} \leq s_j^{1/2}v_j^\top R_t^{\text{bond}} \leq \lambda_1\frac{N}{2}$. Hence the optimal $F_{t,j}$ is given by $\hat{F}_{t,j} = (s_j+\lambda N)^{-1}(s_j^{1/2}|v_j^\top R_t^{\text{bond}}| - \lambda_1\frac{N}{2})^+ \cdot \text{sign}(v_j^\top R_t^{\text{bond}})$, which is equivalent to (24), as claimed.

From (24) we infer that without regularization, $\lambda = \lambda_1 = 0$, the solution would be $\hat{F}_{t,j} = s_j^{-1/2}v_j^\top R_t^{\text{bond}}$. Without LASSO penalty, $\lambda_1 = 0$, the ridge penalty $\lambda > 0$ shrinks factors towards zero, where λ upweights factors corresponding to larger eigenvalues s_j. Hence shrinkage is concentrated on small eigenvalues. This is confirmed by adding a LASSO penalty. For $\lambda_1 > 0$, basis functions are cut off if they do not have high cross-sectional correlations with the excess returns, $|v_j^\top R_t^{\text{bond}}|$, or sufficiently large eigenvalues s_j.

[32] There is no level component in the excess return curve. Level corresponds to the risk-free one-business day return. A similar KR estimator can be derived for total discount bond returns, on the direct sum RKHS $\mathbb{R} \oplus \mathcal{H}_\alpha$ with reproducing kernel given by $1+k$ and by imposing the short end to be equal to the observed risk-free one-business day return. The PCA and KR loadings would very likely exhibit the level, slope, curvature and higher order polynomial shapes. However, this is beyond the scope of this paper.

A.7 Cross-Sectional Modeling of Discount Bond Returns and Yields

A large part of the bond pricing literature considers discount bond (zero-coupon) yields $y_t(x) = (-1/x)\log d_t(x)$. Therefore, it is important to understand that fitting the cross-sectional functional relationship between discount bond returns of different maturities is conceptually and empirically similar to fitting the cross-section of discount bond yields. However, there are also important differences. In fact, there is a mechanical relationship between the return curve r_t and (increments of) the yield curve y_t. Assuming that simple discount bond excess returns are approximately the same as logarithmic excess returns (first order approximation), we obtain[33]

$$\begin{aligned} r_t(x) &= \left(\frac{d_t(x)}{d_{t-1}(x+\Delta_t)} - 1\right) - \left(\frac{1}{d_{t-1}(\Delta_t)} - 1\right) \\ &\approx \log \frac{d_t(x)}{d_{t-1}(x+\Delta_t)} - \log \frac{1}{d_{t-1}(\Delta_t)} \\ &= -(xy_t(x) - (x+\Delta_t)y_{t-1}(x+\Delta_t)) - \Delta_t y_{t-1}(\Delta_t). \end{aligned}$$

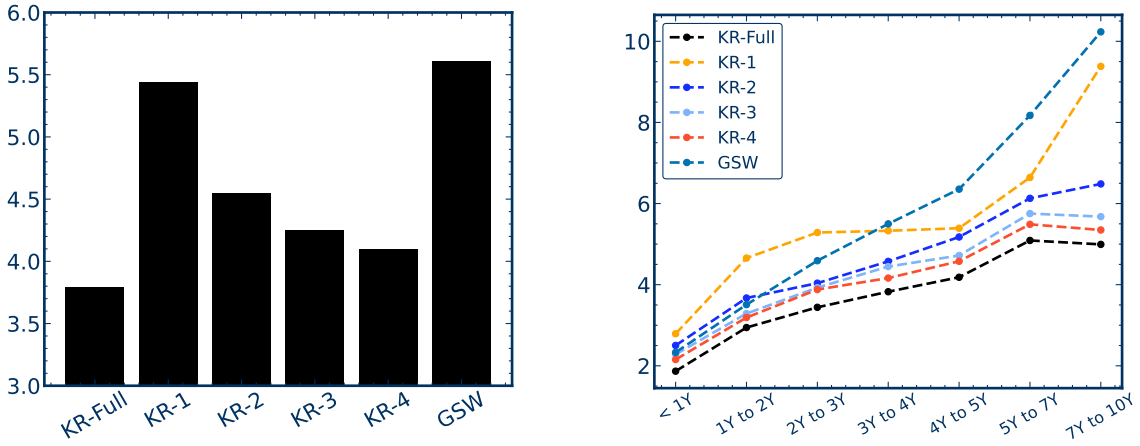
Hence, the shapes that we find in return curves are reflected in yield curves, and vice versa. In particular, the yield and thus the price curves of discount bonds are smooth. As the optimal basis functions to span smooth curves are given by KR factors, the same type of basis functions span the discount curve. In fact, Filipović, Pelger, and Ye (2022) show that leading eigenvectors of discount bond price curves are the same as the ones for discount bond excess returns (besides the mechanical level component). The fact that the time-series of returns are more volatile than those of yields does not mean that the cross-sectional fitting of return curves is more noisy or less stable than the one for yield curves. The advantage of directly estimating the return curve, instead of consecutive yield curves, is that it implies linear factors that are tradable portfolios of Treasuries. A linear estimation of the yield curve would imply non-tradable abstract yield factor portfolios.

[33] If we further assume that the yield curve does not change much over consecutive business days, $y_{t-1} \approx y_t$, we thus find that $r_t(x) \approx (f_{t-1}(x) - f_{t-1}(0))\Delta_t$, for the instantaneous forward rates $f_{t-1} = -d'_{t-1}/d_{t-1}$. This suggests the hypothesis that the forward spread curve $f_{t-1}(x) - f_{t-1}(0)$ may be a predictor of the annualized return curve.

B Additional Empirical Results

B.1 Term Structure Estimation

Figure A.2: Excess Return RMSE Comparison



(a) Average excess return RMSE

(b) Excess return RMSE by maturity bucket

This figure shows the average RMSE of observed coupon bonds in basis points (BPS). We compare the full KR model, KR with 1 to 4 factors and the Gürkaynak, Sack, and Wright (2010). The RMSE are calculated in-sample separately on each date and averaged over time. Subfigure (a) shows the aggregated RMSE, while subfigure (b) reports the RMSE for the six maturity buckets. GSW discount curves at $t-1$ and t are used to calculate excess returns of discount bonds and implied excess returns of observed securities. The discounted cash flow matrices use the KR respectively GSW discount curves. We use the KR risk-free one-business day return to obtain excess returns. The results are calculated using daily data from June 1961 to December 2020.

Table A.1: Excess return RMSE by maturity bucket

	Full	< 1Y	1Y to 2Y	2Y to 3Y	3Y to 4Y	4Y to 5Y	5Y to 7Y	7Y to 10Y
KR-Full	3.79	1.87	2.94	3.44	3.83	4.18	5.09	4.99
KR-1	5.44	2.79	4.66	5.29	5.33	5.39	6.64	9.38
KR-2	4.55	2.50	3.67	4.04	4.57	5.18	6.13	6.48
KR-3	4.25	2.28	3.29	3.93	4.45	4.72	5.75	5.68
KR-4	4.09	2.15	3.19	3.88	4.16	4.58	5.49	5.35
KR-5	4.00	2.07	3.16	3.74	4.09	4.46	5.36	5.16
KR-6	3.94	2.02	3.15	3.61	4.05	4.35	5.25	5.07
KR-7	3.90	1.97	3.11	3.57	3.95	4.32	5.19	5.04
KR-8	3.87	1.94	3.07	3.55	3.91	4.25	5.15	5.02
KR-9	3.84	1.92	3.03	3.52	3.89	4.23	5.13	5.01
KR-10	3.83	1.90	3.00	3.49	3.86	4.21	5.12	5.01
GSW	5.61	2.33	3.51	4.59	5.50	6.35	8.17	10.23

This table shows the average RMSE in fitting excess returns of observed securities. RMSE is calculated separately on each date and then averaged over time. The first column displays RMSE aggregated over all maturities. The rest of the columns show RMSE broken up by the six maturity buckets. GSW discount curves at $t-1$ and t are used to calculate excess returns of discount bonds and implied excess returns of observed securities. KR and GSW discount curves are used to form the transformed cashflow matrix when evaluating KR-based and GSW methods. We use the KR risk-free one-business day return to obtain excess returns. Results are calculated using daily data from June 1961 to December 2020. Numbers are in basis points (BPS).

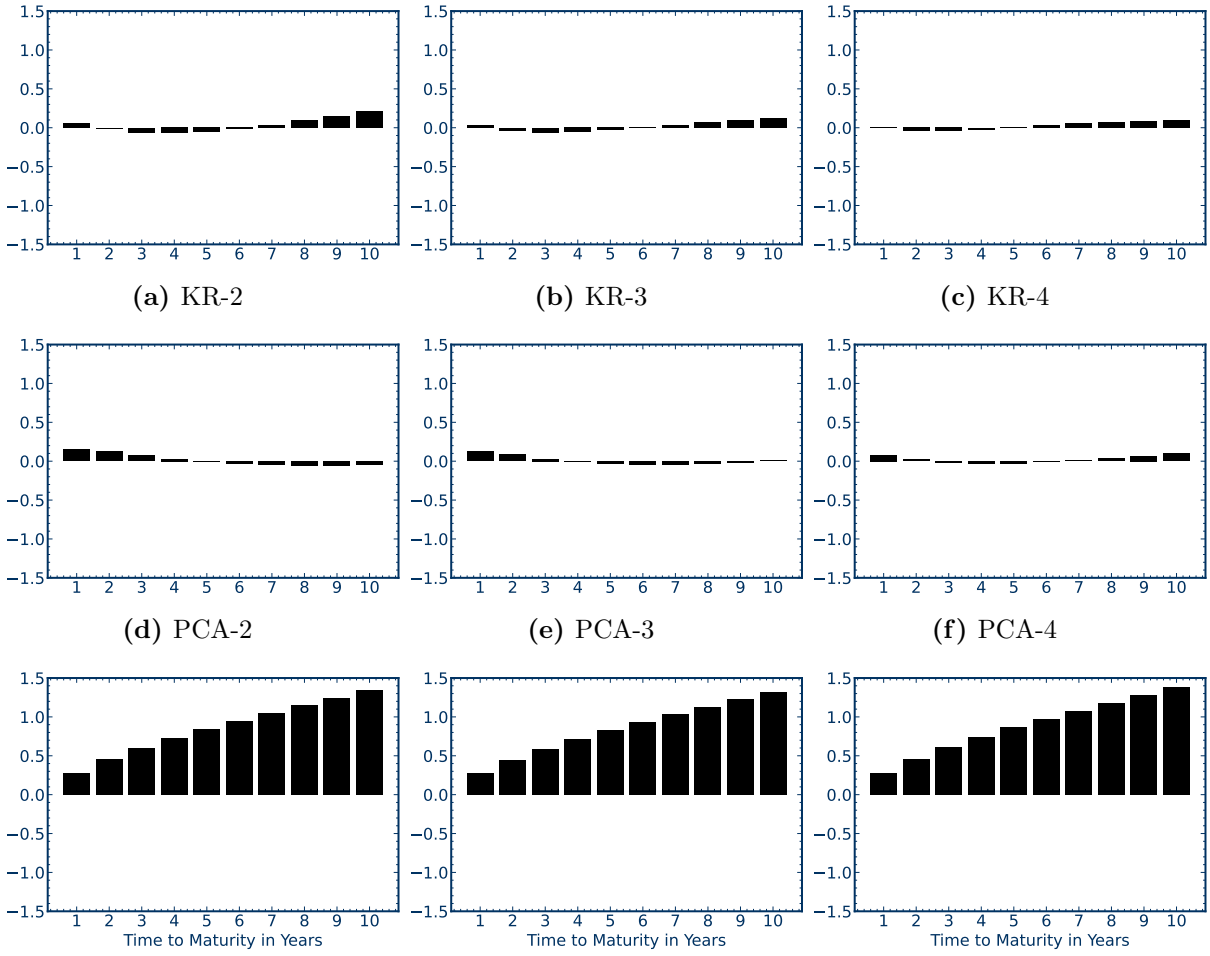
B.2 Cross-Sectional Asset Pricing

Table A.2: Cross-Sectional Asset Pricing Results

	1Y	2Y	3Y	4Y	5Y	6Y	7Y	8Y	9Y	10Y	RMS_α
KR discount bond returns											
KR-1	0.16***	0.11***	0.04	-0.04	-0.09 *	-0.14 *	-0.21 *	-0.32 *	-0.47 *	-0.62 *	0.29
KR-2	0.08***	-0.02	-0.08***	-0.10**	-0.06	0.03	0.12*	0.18**	0.22	0.24	0.13
KR-3	0.05***	-0.06***	-0.07***	-0.03	0.04	0.08**	0.05	-0.07	-0.25***	-0.45***	0.17
KR-4	0.02**	-0.06***	-0.02	0.04**	0.03	-0.02	-0.04	-0.01	0.03	0.08	0.04
KR-5	0.00	-0.04***	0.02	0.03*	-0.01	-0.03	0.01	0.00*	0.01	-0.03	0.03
KR-6	-0.01*	-0.02**	0.02***	-0.00	-0.01	0.01	0.01	-0.01	-0.01	0.01	0.01
KR-7	-0.01***	-0.01	0.01*	-0.01*	0.01*	0.00	0.00	0.00	-0.00	-0.01	0.01
PCA-1	0.22***	0.24***	0.23***	0.20**	0.19*	0.16	0.09	-0.04	-0.22**	-0.41**	0.22
PCA-2	0.17***	0.14***	0.08	0.00	-0.04	-0.05	-0.03	-0.00	0.01	0.03	0.08
PCA-3	0.15***	0.09**	0.01	-0.06**	-0.06*	-0.02	0.03**	0.04	-0.00	-0.05	0.07
PCA-4	0.11***	0.01	-0.04***	-0.02	0.02	0.02	-0.01	-0.02	-0.00	0.02	0.04
PCA-5	0.08***	-0.02***	-0.03**	0.02**	0.01	-0.02**	-0.01	0.01*	0.00	-0.02	0.03
PCA-6	0.03***	-0.03***	0.01***	0.01	-0.01*	0.01	0.01	-0.01	-0.00	0.01	0.02
PCA-7	0.01*	-0.01***	0.01***	-0.01***	0.01***	0.00	-0.01	0.00**	0.00	-0.00	0.01
FF-3	0.31***	0.47***	0.61***	0.73***	0.86***	0.98***	1.07***	1.09***	1.06***	1.03*	0.86
FF-5	0.31***	0.47***	0.62***	0.74***	0.88***	1.01***	1.10***	1.14***	1.12***	1.09*	0.89
GSW discount bond returns											
KR-1	0.14***	0.11**	0.05	-0.01	-0.07	-0.12	-0.17	-0.21	-0.25	-0.27	0.16
KR-2	0.07***	-0.00	-0.05	-0.06	-0.03	0.01	0.06	0.12	0.18	0.24	0.11
KR-3	0.04*	-0.03	-0.05	-0.04	-0.01	0.02	0.06	0.09	0.12	0.15	0.08
KR-4	0.01	-0.04	-0.03	-0.01	0.02	0.05	0.08	0.09	0.11	0.13	0.07
KR-5	0.00	-0.03	-0.02	0.01	0.03	0.05	0.07	0.08	0.09	0.10	0.06
KR-6	-0.00	-0.02	-0.01	0.01	0.02	0.03	0.03	0.03	0.03	0.04	0.03
KR-7	-0.01	-0.02	-0.01	0.01	0.02	0.03	0.03	0.03	0.03	0.04	0.03
PCA-1	0.18***	0.20***	0.18**	0.14*	0.09	0.04	-0.01	-0.06**	-0.11	-0.15	0.13
PCA-2	0.14***	0.10***	0.04	-0.01	-0.04***	-0.05**	-0.04*	-0.01	0.02**	0.08*	0.07
PCA-3	0.08***	0.01	-0.03***	-0.03**	-0.01	0.01***	0.02***	0.01*	-0.00***	-0.03*	0.03
PCA-4	0.01	-0.02***	-0.00	0.01***	0.01**	0.00	-0.01***	-0.01**	-0.00	0.02**	0.01
PCA-5	0.00	-0.00	-0.00	0.00	0.00	0.00	-0.00	-0.00	-0.00	0.00	0.00
PCA-6	-0.00	0.00	0.00	-0.00	-0.00	-0.00	0.00	0.00	0.00	-0.00	0.00
PCA-7	-0.00	0.00	0.00	-0.00	-0.00	-0.00	0.00	0.00	0.00	-0.00	0.00
FF-3	0.28***	0.46***	0.60***	0.73***	0.85***	0.96***	1.05***	1.15***	1.24***	1.33***	0.93
FF-5	0.28***	0.46***	0.62***	0.76***	0.88***	0.99***	1.10***	1.20***	1.29***	1.40***	0.96

This table reports the estimated mean daily pricing errors (in basis point) from the time-series regression from excess returns of discount bonds on different factors. The panel of excess returns of discount bonds is either estimated with the full KR method or GSW yields. We consider three categories of factors: (a) KR factors, (b) PCA from the panel of discount bond excess returns and (c) Fama–French factors. The sample is from July 1963 to December 2020. Stars indicate the significance level of the time-series intercepts. The last column reports the root-mean-squared pricing errors (daily α's in basis points) of the factors. The sample is from July 1963 to December 2020.

Figure A.3: Pricing errors of GSW discount bond excess returns for different factors



(a) KR-2 **(b)** KR-3 **(c)** KR-4

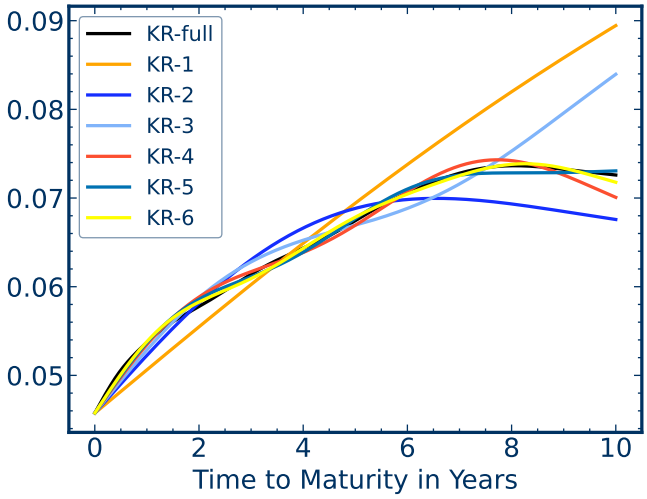
(d) PCA-2 **(e)** PCA-3 **(f)** PCA-4

(g) Mean Excess Returns **(h)** FF-3 **(i)** FF-5

This figure shows the time-series pricing errors (in basis point) from time-series regressions of 10 discount bond excess returns on different factors. The test assets are panel of 10 discount bond excess returns with yearly maturities ranging from 1 to 10 years, which are based on the estimated yields with GSW. We compare KR, PCA and Fama–French factors. The PCA factors are estimated from the GSW panel. The Fama–French 3 and 5 factors are taken from Kenneth French's website. The risk-free one-business day return is the estimate from the KR discount curve. The sampling period is from July 1963 to December 2020.

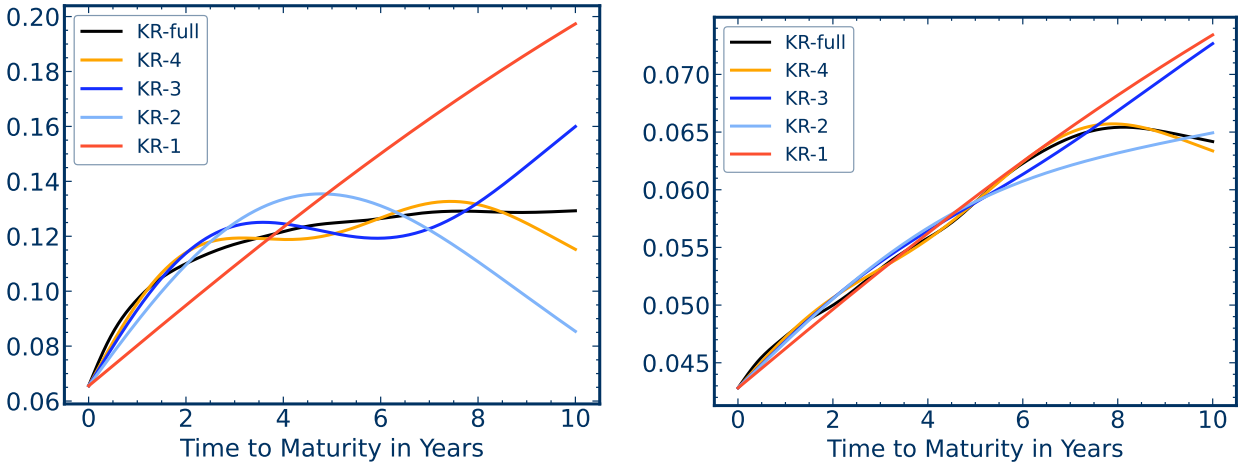
B.3 Term Structure Premium

Figure A.4: Expected returns of discount bonds for different number of factors



This figure shows the average discount bond returns given by KR models for different number of factors for the full sample. We consider the full KR model and one to six KR factors. The sample is daily data from June 1961 to December 2020. The average one-business day returns are annualized.

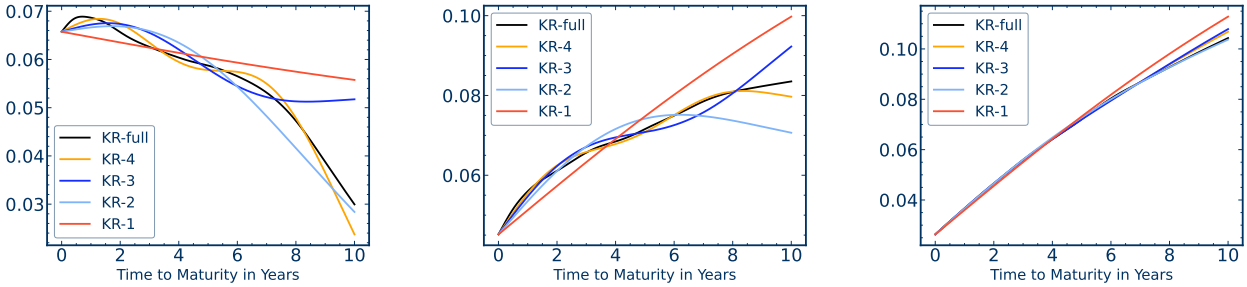
Figure A.5: Average discount bond returns conditioned on boom/recession



(a) Recession **(b)** Boom

The figure show the average returns of estimated discount bonds conditional on boom and recession dates for the full sample. We consider the full KR model and one to six KR factors. Business cycle dates are from NBER. The sample is daily data from June 1961 to December 2020. The average one-business day returns are annualized.

Figure A.6: Average discount bond returns conditioned on yield spreads

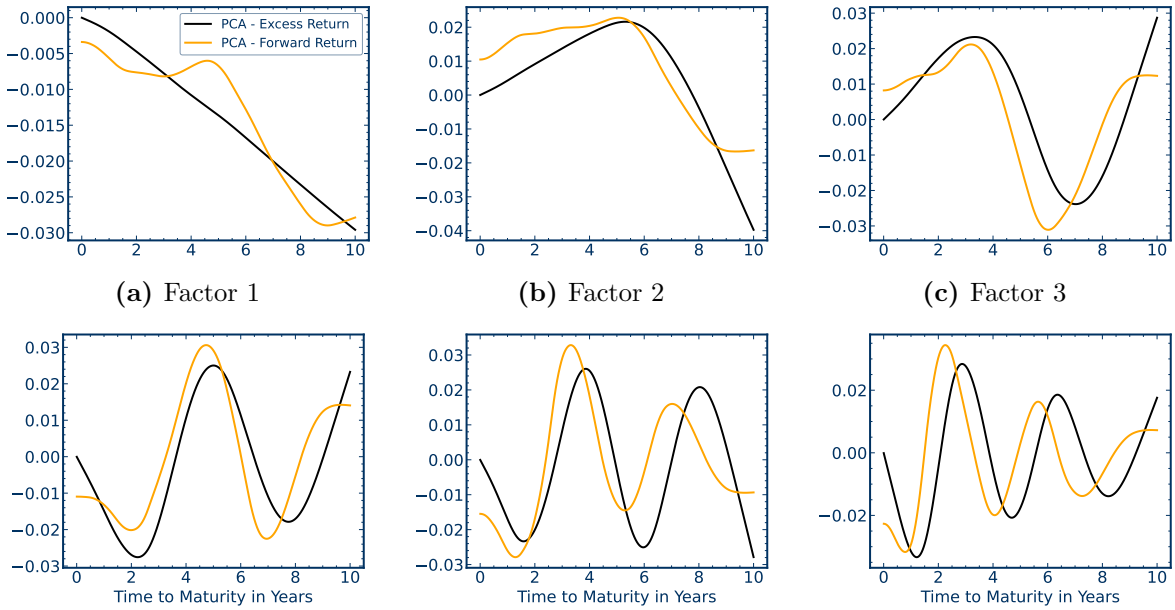


(a) low yield spread **(b)** medium yield spread **(c)** high yield spread

The figure shows the average discount bond returns conditioned on yield spreads for the full sample. The figure compares the conditional discount bond returns for KR models with different number of factors for low/medium/high yield spread dates.. The time periods are sorted into terciles based on the 10 year minus 1 year KR yield spreads. The sample is daily data from June 1961 to December 2020. The average one-business day returns are annualized.

B.4 Term Structure Factors

Figure A.7: PCA factor loadings on discount bond excess returns and forward returns

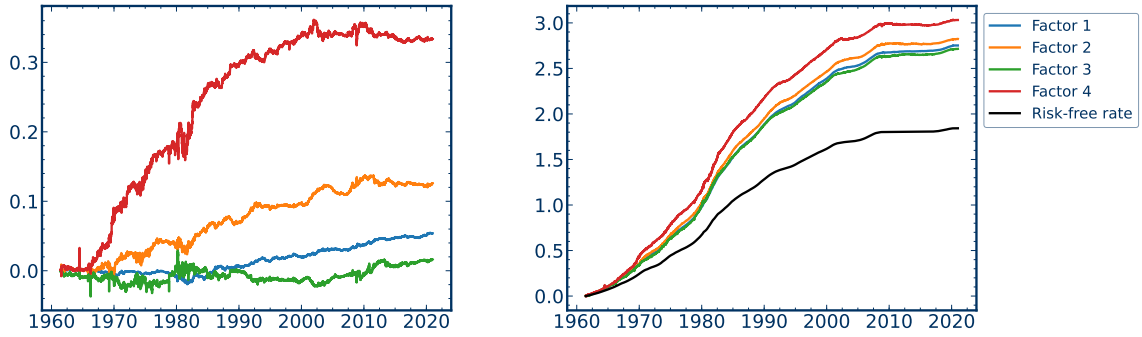


(a) Factor 1 **(b)** Factor 2 **(c)** Factor 3

(d) Factor 4 **(e)** Factor 5 **(f)** Factor 6

This figure shows the loadings of the first 6 PCA factors estimated on a panel of discount bond excess returns and and a panel of forward returns. The loadings correspond to portfolio weights of the PCA factors, either in terms of discount bond excess returns or in terms of forward returns. The loadings are estimated as the eigenvectors of the leading eigenvalues of Σ and Σ_f. The sample covariance matrices are based on a panel of discount bond excess returns estimated with the full KR model. We use daily returns with daily maturities from June 1961 to December 2020.

Figure A.8: Cumulative excess returns and returns of the first four KR factors



(a) Cumulative excess returns of factors **(b)** Cumulative returns of factors

This figure shows cumulative excess returns (left) and cumulative returns (right) of the first four KR factors. The KR factors are portfolios of tradable Treasury bonds. The risk-free one-business day return is the estimate from the KR discount curve and is also given as a tradable portfolio. The sample is daily data from June 1961 to December 2020.

Table A.3: Sharpe ratios of factors and implied SDF

n	Factor	SDF (IS)	SDF(OOS)
1	0.47	0.47	0.51
2	0.43	0.67	0.69
3	0.26	0.74	0.75
4	0.34	0.84	0.85
5	0.21	0.88	0.86
6	0.20	0.93	0.89

This table shows the annualized Sharpe ratios for different KR factors and for their implied SDF. We report the Sharpe ratio of each factor and of the SDF, when forming the mean-variance efficient portfolio by successively including KR factors. The Sharpe ratios for the individual factors and SDF are in-sample, while the SDF (OOS) shows the out-of-sample Sharpe ratios for estimating the mean-variance efficient portfolio weights on a daily updated rolling window of the previous 10-years of daily excess returns. The sample is daily data from June 1961 to December 2020.

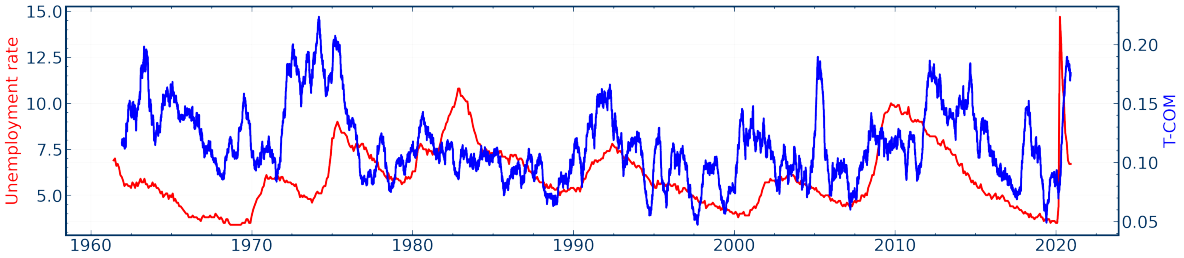
Table A.4: KR factors conditioned on economic conditions

	SR				Mean				Standard Deviation			
	F1	F2	F3	F4	F1	F2	F3	F4	F1	F2	F3	F4
Conditional on boom and recessions												
Boom	0.31	0.12	0.10	0.23	7.17	10.24	13.95	33.29	23.41	83.24	137.98	142.00
Recession	0.59	0.74	-0.31	0.79	22.45	96.39	-72.03	213.42	37.79	130.48	231.46	269.63
Conditional on yield spread terciles												
Low yield spread	-0.21	0.26	0.15	0.48	-5.96	33.04	31.04	102.37	27.92	127.19	212.01	214.24
Medium yield spread	0.53	0.38	-0.14	0.50	11.61	23.35	-15.92	65.71	22.11	61.79	110.08	132.05
High yield spread	0.81	0.11	-0.06	0.01	21.79	7.76	-6.62	1.68	26.75	68.67	115.99	132.50

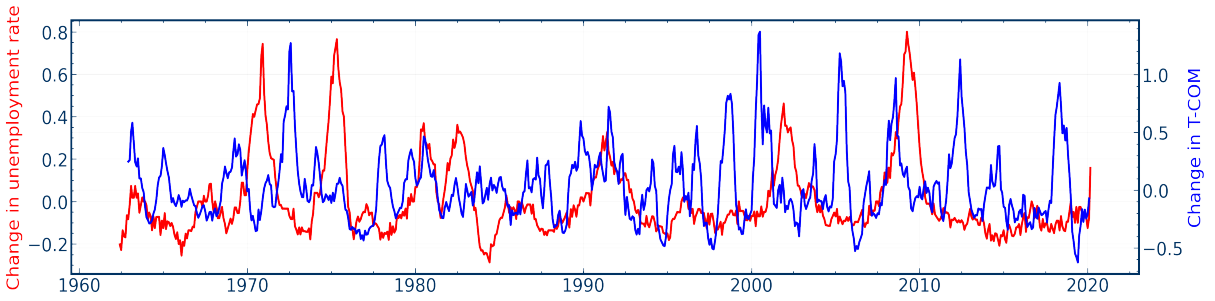
The table shows the annualized Sharpe ratio, mean and standard deviation of the four KR factors conditioned on NBER boom and recession periods and yield spread terciles. The time periods are sorted into terciles based on 10 year - 1 year KR yield spreads. Mean and standard deviation are in basis points. The sample includes daily data from June 1961 to December 2020.

B.5 Bond Market Condition Measures

Figure A.9: T-COM and unemployment rates



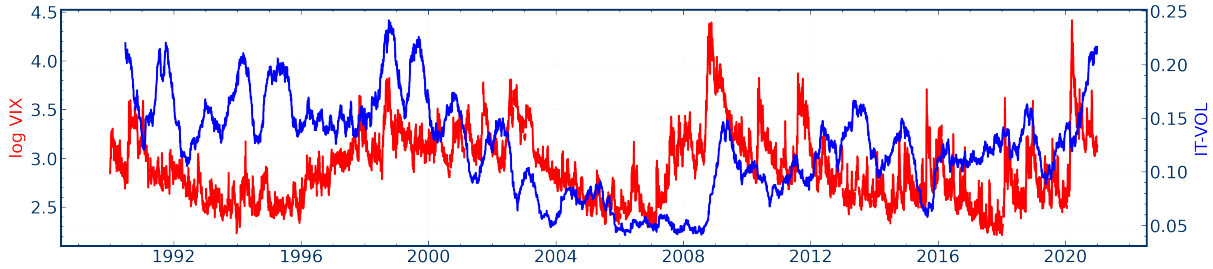
(a) Level in T-COM and monthly unemployment rate



(b) Changes in T-COM and monthly unemployment rate

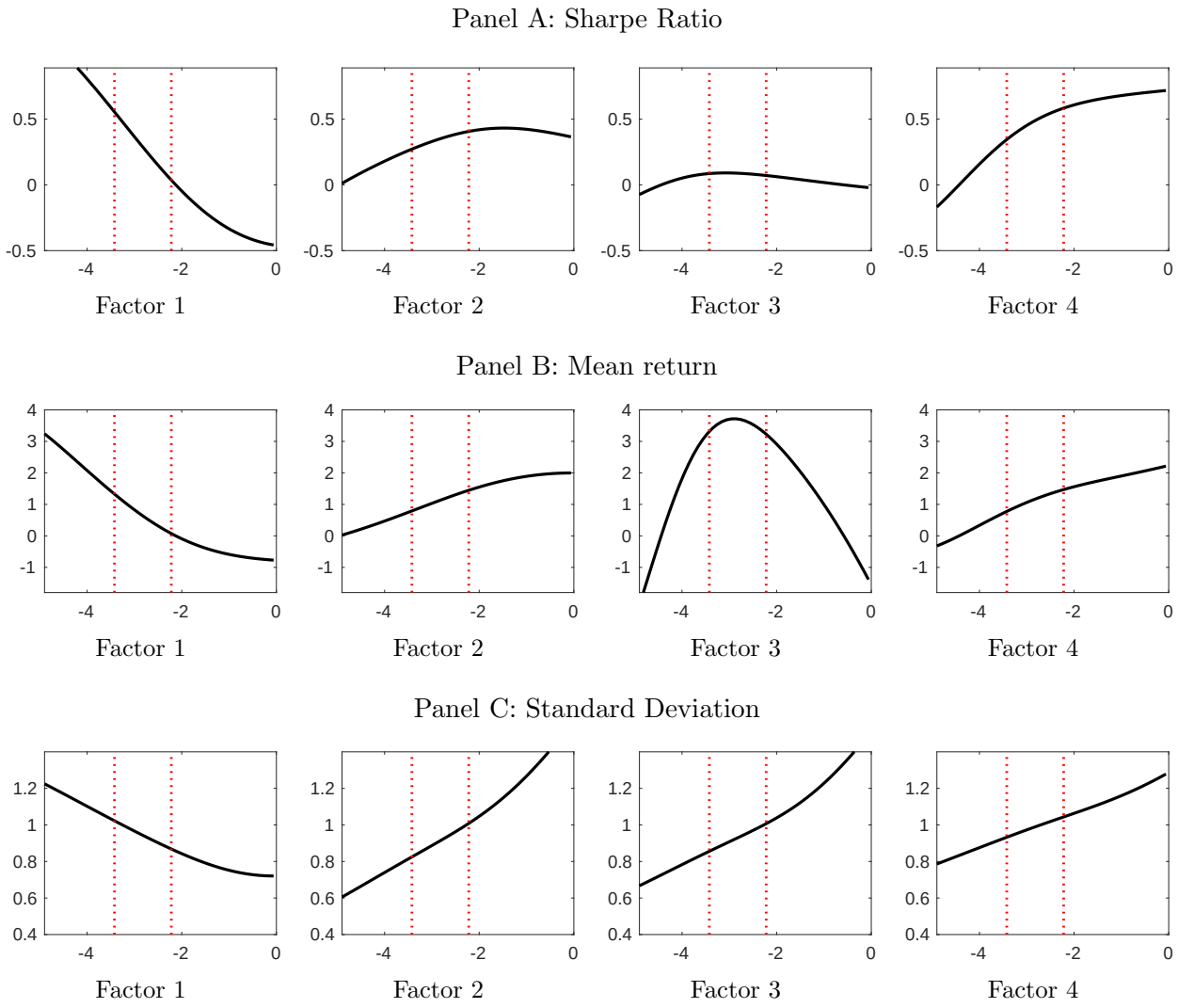
The figure compares the times-series of T-COM with unemployment rates. The top figure shows the 120-day moving average (MA) of T-COM at daily frequency and unemployment rates at monthly frequency, while the bottom figure displays the changes in the 120-day moving average of T-COM and of monthly unemployment rates relative to the previous year. The moving average uses the prior 120 business day observations of T-COM. The changes are of monthly frequency. The left axis corresponds to the scale for the unemployment rate, while the right axis is the scale for T-COM. The red line corresponds to the level/changes in unemployment rate, while the blue line is the level/change in T-COM.

Figure A.10: IT-VOL and VIX



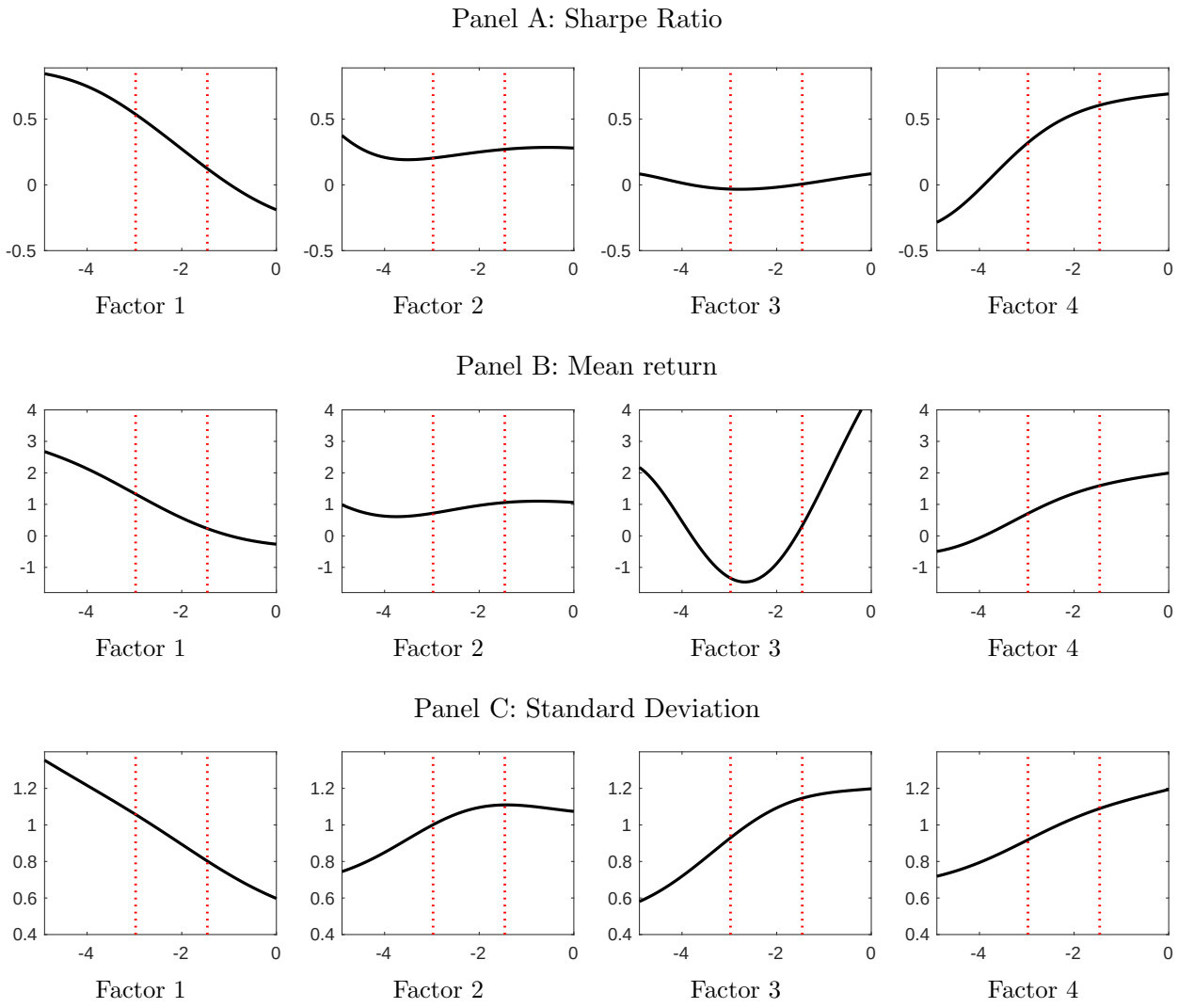
The figure compares the times-series of IT-VOL with the VIX index. We show the 120-day MA of IT-VOL and the log VIX at daily frequency. The VIX is the Chicago Board Options Exchange (CBOE) volatility index, and measures market expectation of near term volatility conveyed by stock index option prices. The moving average uses the prior 120 business day observations of IT-VOL. The left axis corresponds to the scale for log VIX, while the right axis is the scale for IT-VOL. The blue line corresponds to IT-VOL, and the red line is the VIX.

Figure A.11: Sharpe ratio, mean, and standard deviation of KR factors conditioned on T-COM



This figure shows the annualized Sharpe ratio (Panel A), normalized mean (Panel B), and normalized standard deviation (Panel C) of the four KR factors conditioned on the Treasury Market Complexity (T-COM) measure. The mean returns and standard deviation are normalized by their unconditional values. We condition on the logarithm of T-COM. The vertical dotted lines represent the tercile values of T-COM. The conditional first and second moment are estimated with kernel smoothing regressions. The sample is daily data from June 1961 to December 2020.

Figure A.12: Sharpe ratio, mean, and standard deviation of KR factors conditioned on IT-VOL



This figure shows the annualized Sharpe ratio (Panel A), normalized mean (Panel B), and normalized standard deviation (Panel C) of the four KR factors conditioned on the Idiosyncratic Treasury Volatility (IT-VOL) measure. The mean returns and standard deviation are normalized by their unconditional values. We condition on the logarithm of IT-VOL. The vertical dotted lines represent the tercile values of IT-VOL. The conditional first and second moment are estimated with kernel smoothing regressions. The sample is daily data from June 1961 to December 2020.

Internet Appendix for Shrinking the Term Structure

Damir Filipović* Markus Pelger† Ye Ye‡

This draft: August 6, 2024
First draft: September 23, 2022

Abstract

The Internet Appendix collects additional empirical results supporting the main text.

Keywords: Term structure of interest rates, bond returns, factor space, U.S. Treasury securities, non-parametric method, principal components, machine learning in finance, reproducing kernel Hilbert space
JEL classification: C14, C38, C55, G12

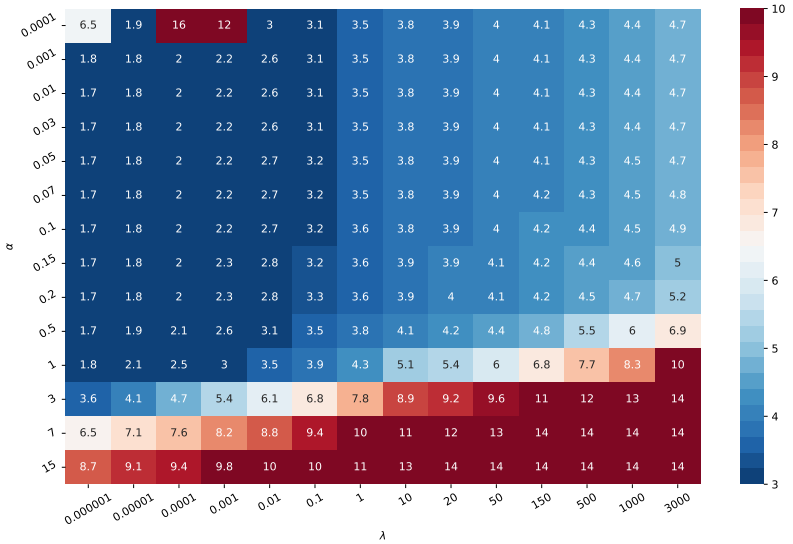
*École Polytechnique Fédérale de Lausanne and Swiss Finance Institute, Email: damir.filipovic@epfl.ch
†Stanford University, Department of Management Science & Engineering, and NBER, Email: mpelger@stanford.edu.
‡Stanford University, Department of Management Science & Engineering, Email: yeye@stanford.edu.

Contents

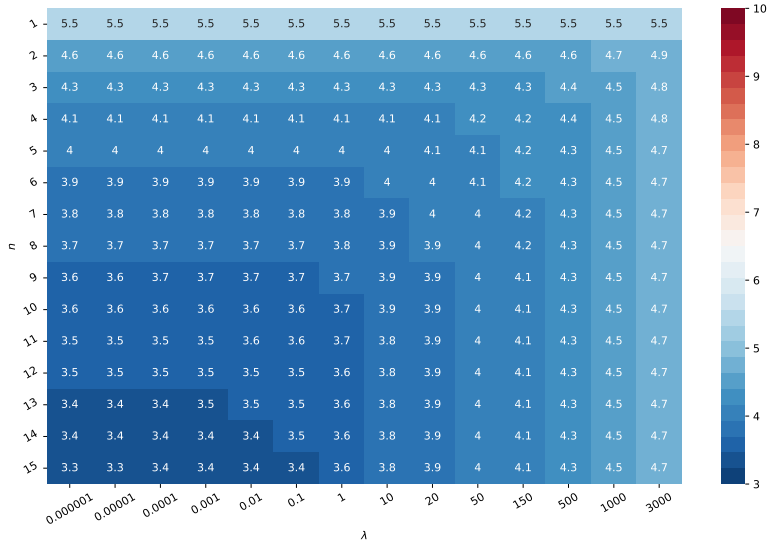
IA.A Term Structure Estimation

Figure IA.1 shows the in-sample results for the RMSE of the KR models for different choices of tuning parameters. It confirms the overall robustness of the parameter choices. As the in-sample results are prone to overfitting, a more regularized estimation with either larger smoothness penalty λ or fewer factors n has obviously higher in-sample RMSE.

Figure IA.1: Average in-sample excess return RMSE



(a) RMSE for λ and α



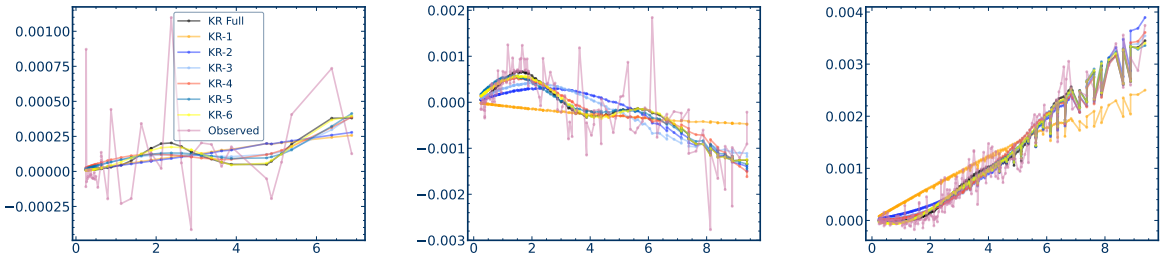
(b) RMSE for λ and n

This figure shows the average in-sample excess return RMSE of the full KR model in basis points (BPS) as a function of the smoothness parameter λ, the maturity weight α and the number of factors n. Subplot (a) displays the RMSE for λ and α, while subplot (b) reports the RMSE for λ and n. Factors are added by the order implied by the kernel decomposition, that is, in decreasing order of the eigenvalues of the kernel matrix $\boldsymbol{K}$. Results are calculated using quarterly data from June 1961 to December 2020.

Figure IA.2 illustrates the estimated discount bond excess return curve on three representative example days. Including more factors allows the estimated curves to capture more complex patterns and move closer to the full KR curves. A KR factor model with four to six factors approximates the full KR curve relatively well. The first KR factor seems to capture a linear slope pattern. The lower panels of the figures apply the sparse basis functions to yield estimation instead of return estimation. They illustrate that the yield estimates of Filipović, Pelger, and Ye (2022) can also be approximated well by a small number of yield factors.

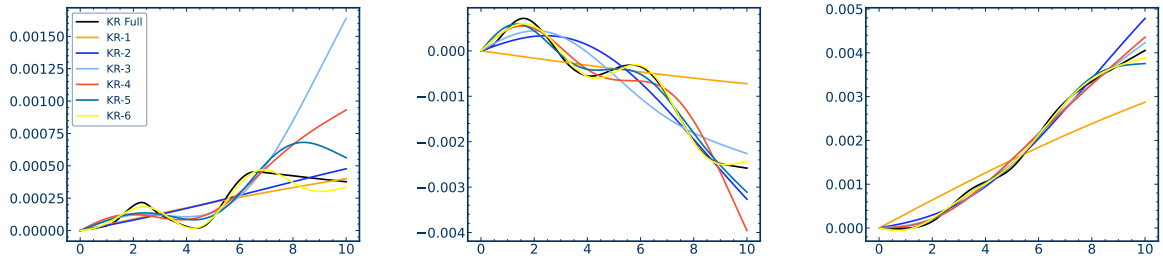
Figure IA.2: Illustration of KR estimator

Panel A: Observed and Fitted Excess Returns of Securities



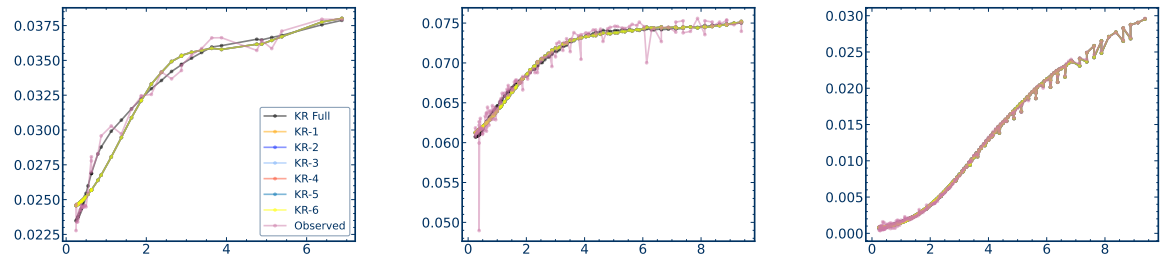
(a) 1961-06-30 **(b)** 1986-06-30 **(c)** 2013-12-30

Panel B: Fitted Excess Returns of Discount Bonds



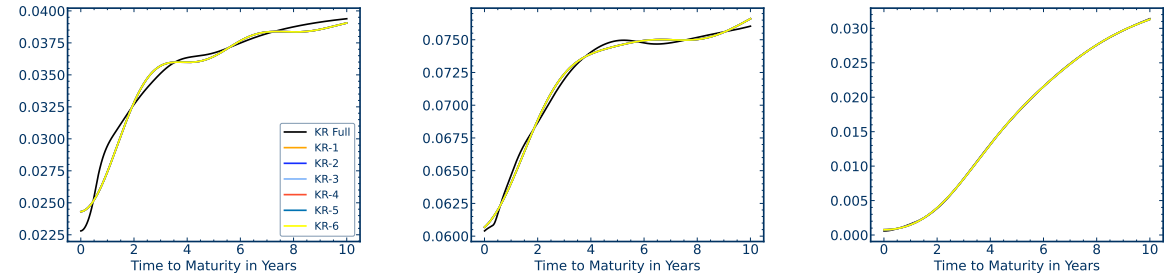
(d) 1961-06-30 **(e)** 1986-06-30 **(f)** 2013-12-30

Panel C: Observed and Fitted YTM of Securities



(g) 1961-06-30 **(h)** 1986-06-30 **(i)** 2013-12-30

Panel D: Fitted Yield Curves



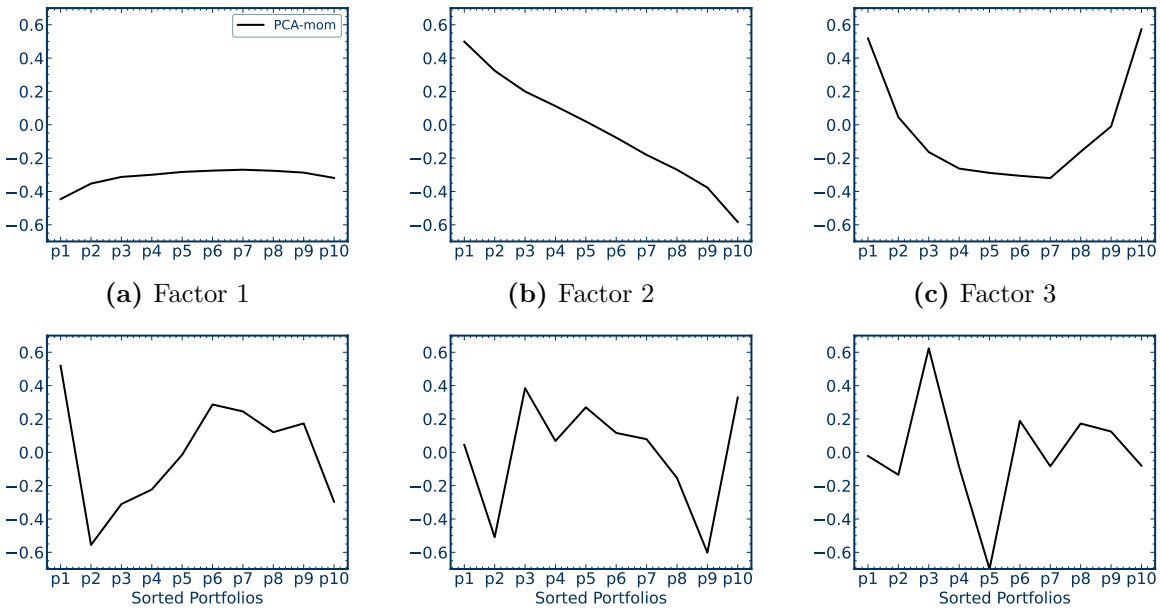
(j) 1961-06-30 **(k)** 1986-06-30 **(l)** 2013-12-30

This figure illustrates the KR estimates on three example dates. The panel A plots observed and fitted excess returns of securities. The panel B displays fitted excess returns of discount bonds. Panel C plots observed and fitted YTM of securities, and Panel D shows fitted yield curves.

IA.B Explaining Slope and Curvature by Smoothness for Equities

The relevance of our findings extends beyond term structure modeling. Slope and curvature patterns in PCAs of panels of asset returns have been demonstrated for a variety of asset classes, including equities and foreign exchange markets. In equities it is common to use portfolio sorts to reflect the conditional impact of firm characteristics. For example, for book-to-market ratios or momentum, one can obtain a panel of decile sorts for the different quantiles of the characteristics. These basis assets are the analogue to our discount bonds as they have close to constant exposure to the sorting variable in contrast to individual stocks. As shown in Figures IA.3 and IA.4, a PCA on the decile sorts leads to the same slope and curvature patterns as for "maturity sorted" discount bonds. Our findings suggest that these type of factors simply arise because the functional dependency of stock returns on momentum or book-to-market is inherently smooth. This non-parametric perspective is novel, and we are able to make it because our approach unifies non-parametric curve estimation with cross-sectional factor modeling.

Figure IA.3: PCA factor loadings of momentum sorted decile portfolios

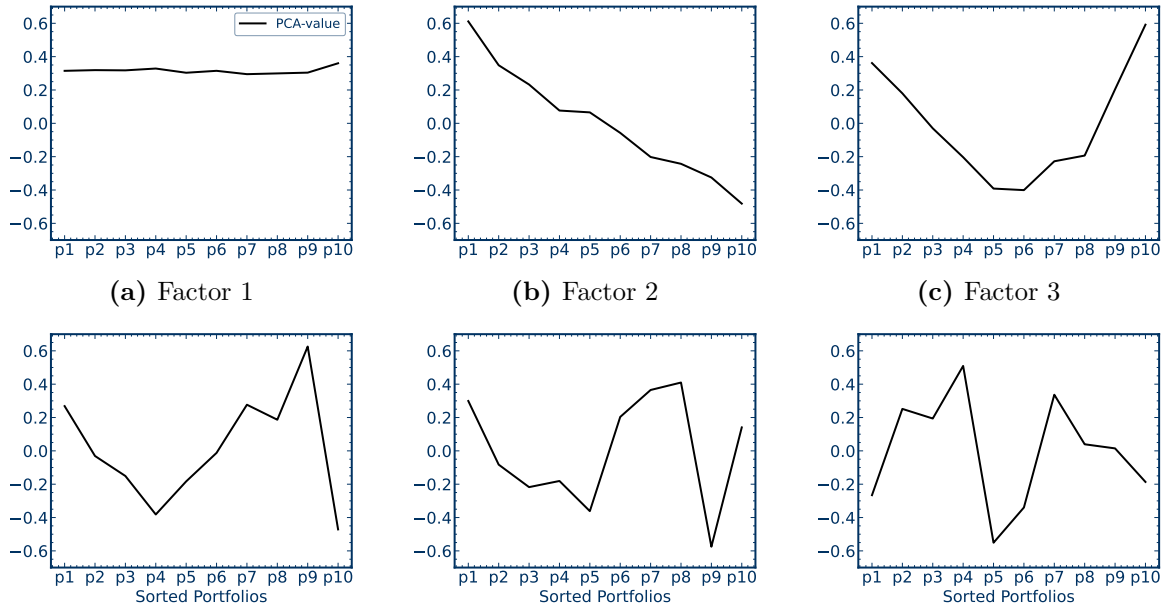


(a) Factor 1 **(b)** Factor 2 **(c)** Factor 3

(d) Factor 4 **(e)** Factor 5 **(f)** Factor 6

This figure shows the normalized loadings of the first 6 PCA factors from a panel of monthly excess returns of decile sorted portfolios based on momentum. The data is from Lettau and Pelger (2020). The panel of ten portfolios starts in November 1963 and ends in December 2017 (T=650). Note that PCA factors are not identified up to a sign, and without loss of generality the sign of the loadings can be changed.

Figure IA.4: PCA factor loadings of value sorted decile portfolios



(a) Factor 1 **(b)** Factor 2 **(c)** Factor 3

(d) Factor 4 **(e)** Factor 5 **(f)** Factor 6

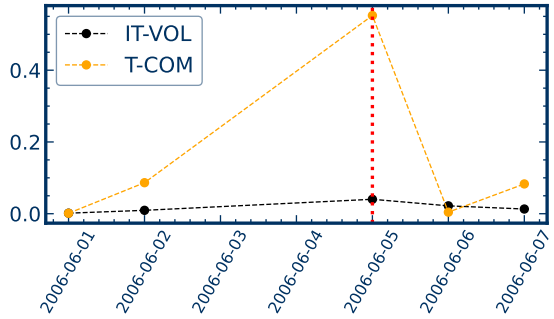
This figure shows the normalized loadings of the first 6 PCA factors from a panel of monthly excess returns of decile sorted portfolios based on book-to-market ratios. The data is from Lettau and Pelger (2020). The panel of ten portfolios starts in November 1963 and ends in December 2017 (T=650). Note that PCA factors are not identified up to a sign, and without loss of generality the sign of the loadings can be changed.

IA.C Interpretation of Bond Market Conditions

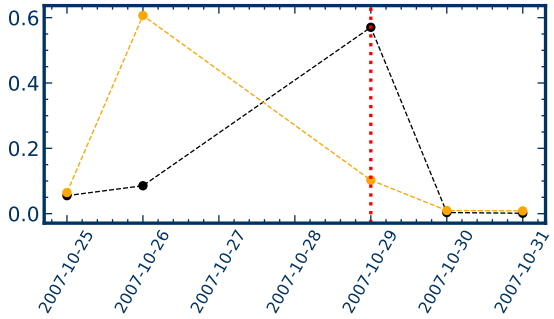
IA.C.1 High-Frequency Changes in Bond Markets

Our novel measures also detect high-frequency changes in bond markets. Figure IA.5 illustrates the two measures and the fitted coupon bond excess returns for a time window around example days. In Panel (a), T-COM peaks on 2006-06-05 while IT-VOL remains low. In Panel (b), IT-VOL peaks on 2007-10-25, while T-COM stays low. The start and end date for the windows around the example days indicate a "simple" term structure, that is a simple linear model provides a reasonable fit, while the noise level is comparatively low. These examples indicate that bond market conditions can change quickly. In summary, we provide two refined measures of the state of the Treasury bond market that can be used for high frequencies.

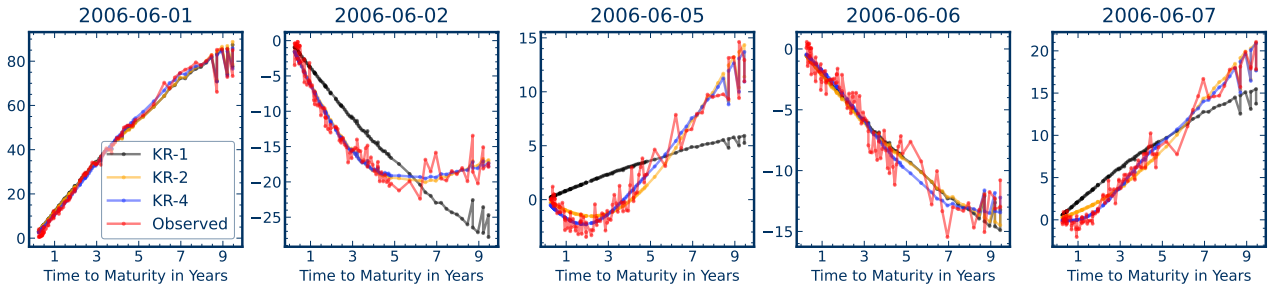
Figure IA.5: The Idiosyncratic Treasury Volatility (IT-VOL) measure and the Treasury Market Complexity (T-COM) measure on example dates over time



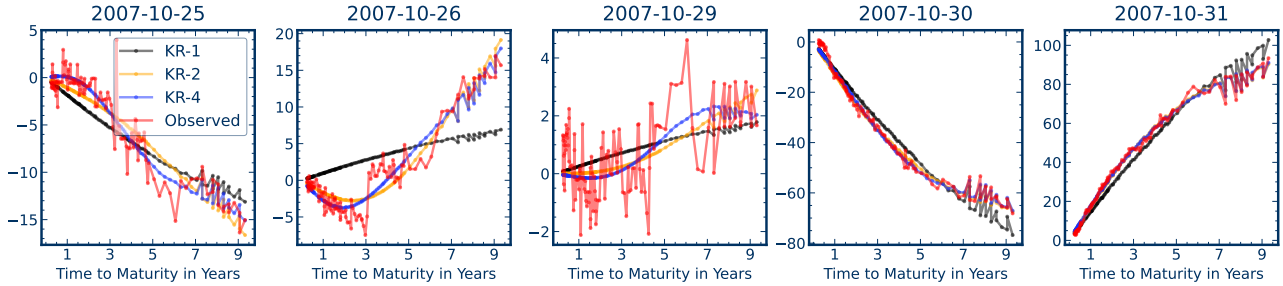
(a) IT-VOL and T-COM centered on 2006-06-05



(b) IT-VOL and T-COM centered on 2007-10-29



(c) Fits of coupon bond excess returns centered on 2006-06-05



(d) Fits of coupon bond excess returns centered on 2007-10-29

Panels (a) and (b) show the Idiosyncratic Treasury Volatility (IT-VOL) measure and the Treasury Market Complexity measure (T-COM) measure centered on two example dates: 2006-06-05 for (a) and 2007-10-29 for (b). Example dates are marked with red vertical lines. Panels (c) and (d) display fits of coupon bond excess returns on dates adjacent to the example dates. Numbers are in basis points. In Panel (a), T-COM peaks on 2006-06-05 while IT-VOL remains low. Panel (b), IT-VOL peaks on 2007-10-25, while T-COM stays low.

IA.C.2 T-COM and Unemployment Rates

The measures of bond market conditions directly relate to real economic conditions. We show that the bond market complexity is correlated with future unemployment rates.

Changes in T-COM predict changes in unemployment rate one year ahead. Figure IA.6 shows the correlation of changes in T-COM and the unemployment rate relative to last year for different lags. The correlation is computed by shifting the time series of changes in the unemployment rate one month forward at a time. The largest correlation of 0.21 occurs at the 14-month lag. This

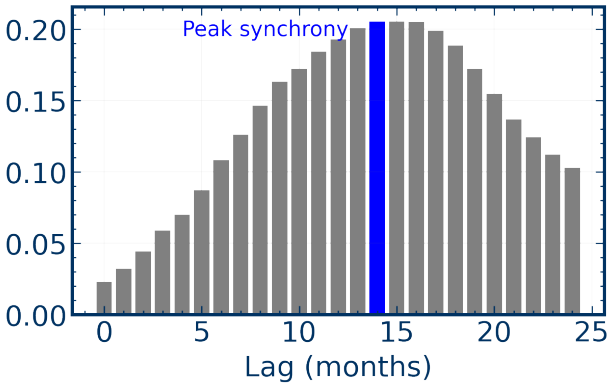
means that that changes in T-COM have the highest correlations with changes in unemployment rates roughly one year in the future.

This prediction is statistically significant. First, we confirm that the two time-series of changes do not have unit-roots. Second, we run a Granger causality test. We find that changes in T-COM Granger cause changes in unemployment rates with a p-value of 0.01. Lastly, we estimate a simple regression model using the 14-month lagged change in T-COM to predict changes in unemployment rates:

$$\text{unemployment}_t = 0.01 + 0.14 \cdot \text{T-COM}_{t-14} + \epsilon_t \tag{1}$$

The regression coefficient on T-COM is highly significant with a t-statistic of 4.7 and an $R^2 = 0.05$. We conclude that more complex shapes of the term structure are predictive for future bad economic conditions.

Figure IA.6: Correlation of changes in T-COM and unemployment rate

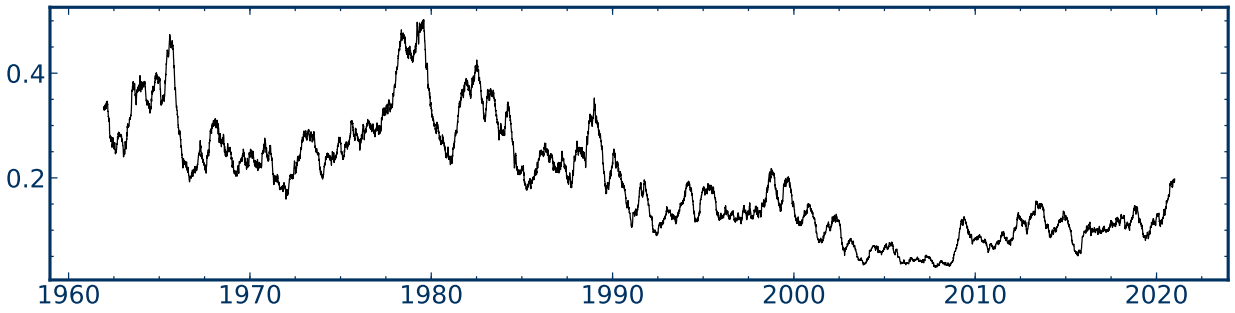


The figure shows the correlation of changes in T-COM and the unemployment rate relative to last year for different lags. The correlation is computed by shifting the time series of changes in unemployment rate one month forward at a time. Changes of T-COM are calculated using the moving average of the prior 120 business day observations of T-COM. The largest correlation of 0.21 occurs at the 14-month lag.

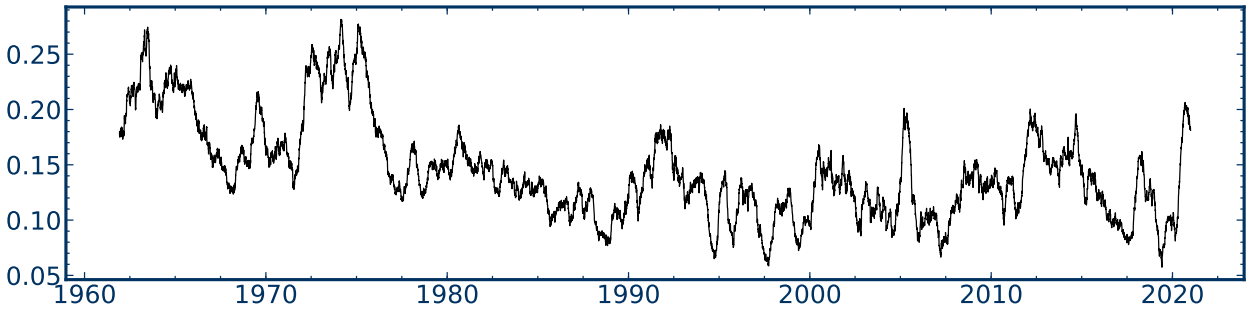
IA.D Alternative Bond Market Condition Measures

The results in the main text are robust to alternative formulations of T-COM and IT-VOL. Table IA.1 and Figures IA.7, IA.8, and IA.9 consider an alternative formulation based on the full KR model. There, the IT-VOL and T-COM measures are constructed with the full KR model, that is, the errors for the unexplained variation are based on the full KR model. Similarly, the alternative complexity measure is based on the normalized difference in cross-sectionally explained variation between the first KR factor and the full KR model instead of only the first four KR factors. The results are virtually identical, showing that our findings do not depend on the choice of the number of factors.

Figure IA.7: Time-series of market condition measures based on full KR model



(a) Idiosyncratic Treasury Volatility measure (IT-VOL)



(b) Treasury Market Complexity measure (T-COM)

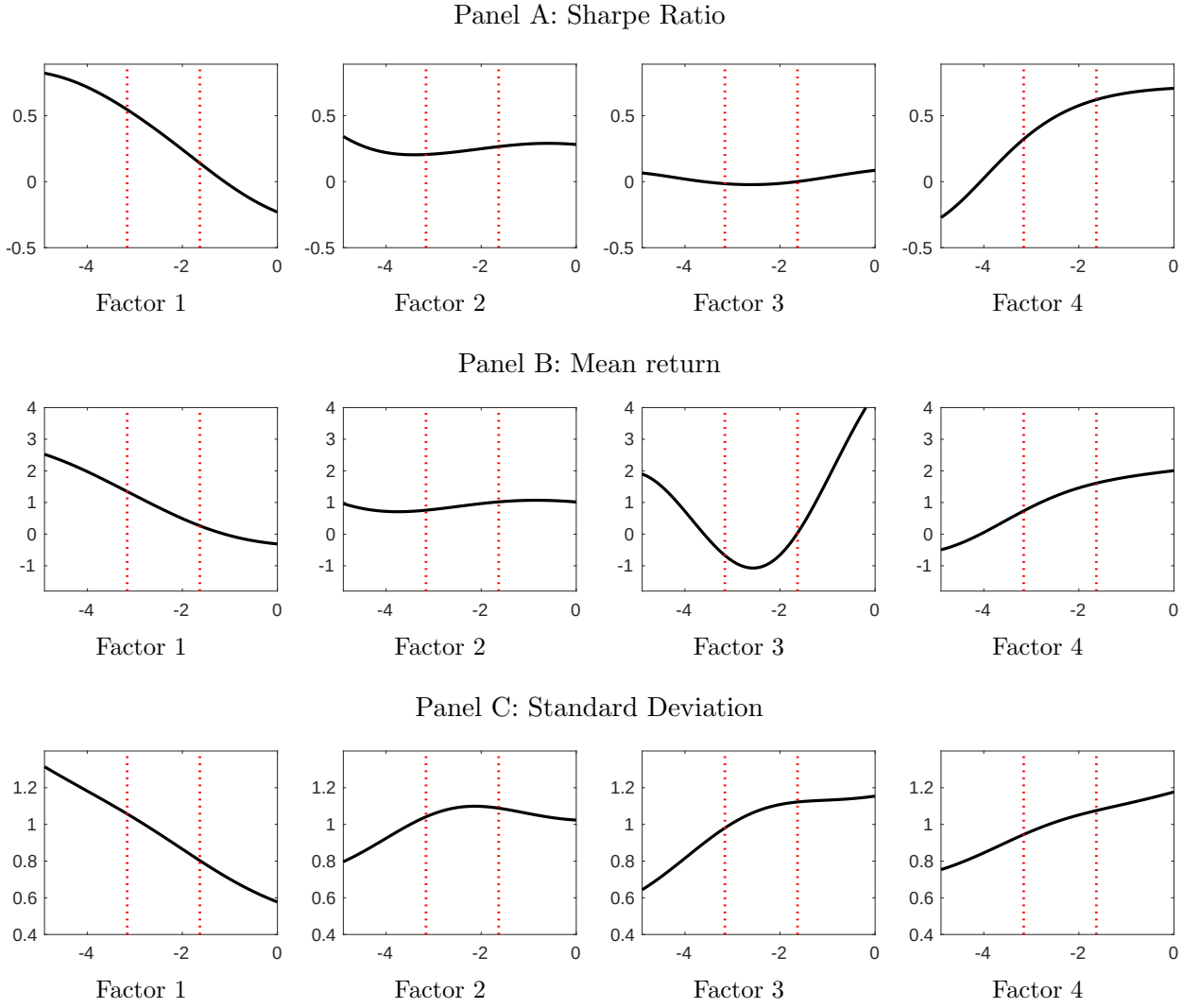
This figure shows the 120-day moving average of the two market condition measures. Subfigure (a) displays the Idiosyncratic Treasury Volatility measure (IT-VOL) as the unexplained variation by the full KR model. Subfigure (b) presents the Treasury Market Complexity measure (T-COM) as the difference between unexplained variation by the 1-factor and the full KR model. The unexplained cross-sectional variation of a factor model is normalized by the overall cross-sectional variation on that day.

Table IA.1: Sharpe ratios, means, and standard deviation of KR factors conditioned on IT-VOL and T-COM based on full KR model

	SR				Mean (BPS)				Standard Deviation (BPS)			
	F_1	F_2	F_3	F_4	F_1	F_2	F_3	F_4	F_1	F_2	F_3	F_4
IT-VOL												
Tercile 1	0.64	0.26	0.02	-0.22	22.22	17.52	1.70	-25.83	34.55	66.39	83.73	117.04
Tercile 2	0.49	0.13	-0.12	0.41	12.43	14.64	-22.64	72.25	25.16	114.45	188.51	177.83
Tercile 3	-0.58	0.38	0.18	0.65	-7.21	32.00	29.44	123.34	12.53	84.88	167.24	188.53
T-COM												
Tercile 1	1.13	-0.25	-0.14	-0.48	2.47	-0.75	-0.77	-3.57	34.70	47.07	86.71	117.32
Tercile 2	-0.25	0.37	0.09	0.77	-0.37	1.80	0.75	7.41	23.06	77.17	127.74	152.86
Tercile 3	-0.38	0.37	0.04	0.52	-0.37	2.99	0.56	6.85	15.63	128.58	216.06	209.06

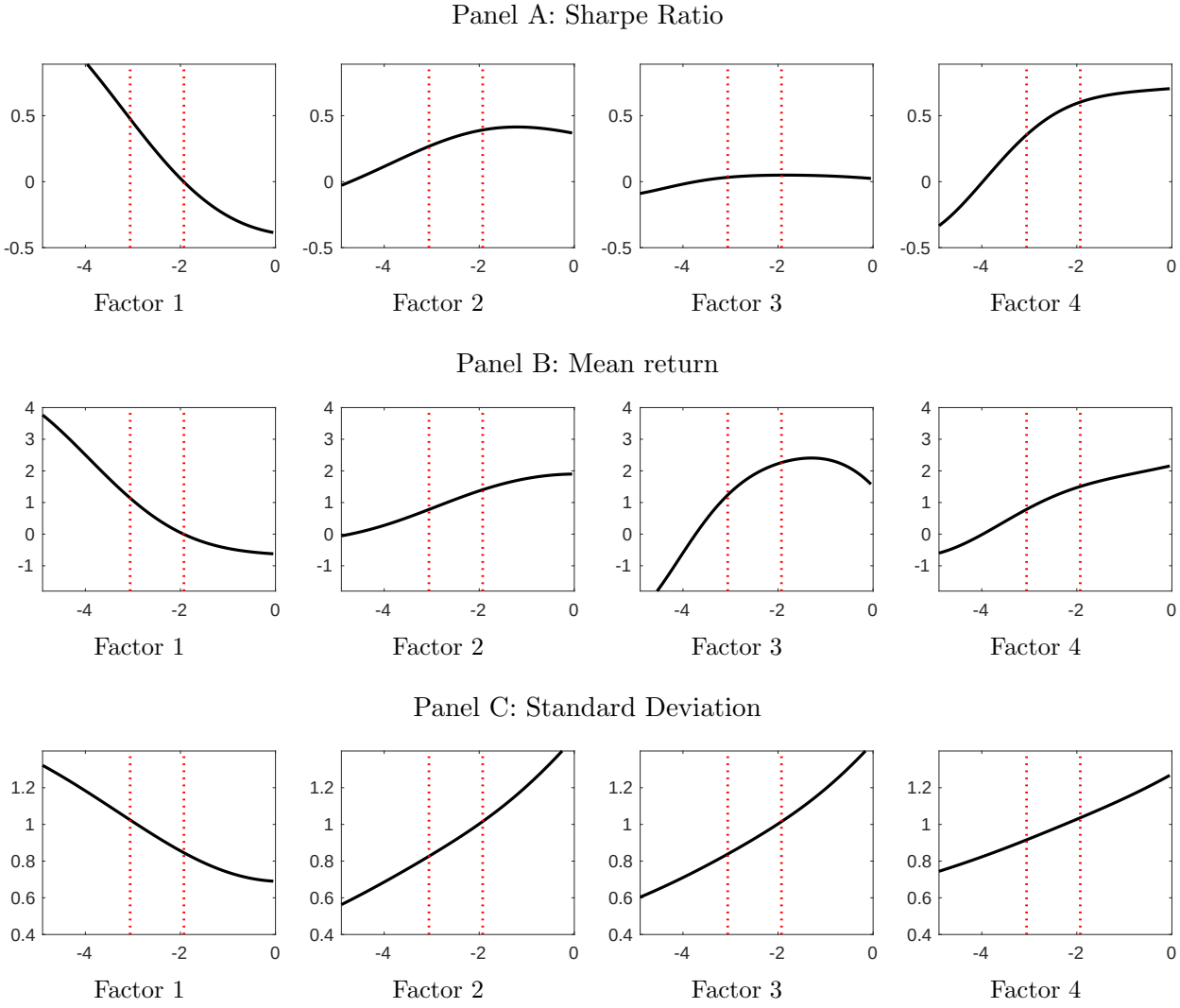
This table shows the annualized Sharpe ratio, mean, and standard deviation of the four KR factors conditioned on terciles of the Idiosyncratic Treasury Volatility (IT-VOL) measure and the Treasury Market Complexity (T-COM) measure. The IT-VOL and T-COM measures are constructed with the full KR model, that is, the cross-sectional unexplained variation with respect to the full KR model, respectively, the difference in cross-sectionally explained variation between the first KR factor and the full KR model (instead of only the first four KR factors). Mean and standard deviation are annualized and in basis points. The sample is from June 1961 to December 2020.

Figure IA.8: Sharpe ratio, mean, and standard deviation of KR factors conditioned on IT-VOL based on full KR model



This figure shows the annualized Sharpe ratio (Panel A), normalized mean (Panel B), and normalized standard deviation (Panel C) of the four KR factors conditioned on the Idiosyncratic Treasury Volatility (IT-VOL) measure. The IT-VOL measure is constructed with the full KR model, that is, the cross-sectional unexplained variation with respect to the full KR model (instead of only the first four KR factors). The mean returns and standard deviation are normalized by their unconditional values. We condition on the logarithm of IT-VOL. The vertical dotted lines represent the tercile values of IT-VOL. The conditional first and second moment are estimated with kernel smoothing regressions. The sample is from June 1961 to December 2020.

Figure IA.9: Sharpe ratio, mean, and standard deviation of KR factors conditioned on T-COM based on full KR model



This figure shows the annualized Sharpe ratio (Panel A), normalized mean (Panel B), and normalized standard deviation (Panel C) of the four KR factors conditioned on the Treasury Market Complexity (T-COM) measure. The T-COM measure is constructed with the full KR model, that is, the difference in cross-sectionally explained variation between the first KR factor and the full KR model (instead of only the first four KR factors). The mean returns and standard deviation are normalized by their unconditional values. We condition on the logarithm of T-COM. The vertical dotted lines represent the tercile values of T-COM. The conditional first and second moment are estimated with kernel smoothing regressions. The sample is from June 1961 to December 2020.

References

Filipović, D., M. Pelger, and Y. Ye (2022): "Stripping the Discount Curve – a Robust Machine Learning Approach," *Working paper*.

Lettau, M., and M. Pelger (2020): "Factors that Fit the Time-Series and Cross-Section of Stock Returns," *Review of Financial Studies*, 33(5), 2274–2325.