RESEARCH

Swiss Finance Institute
Research Paper Series
N°22-61

Shrinking the Term Structure

Damir Filipović
Ecole Polytechnique Fédérale de Lausanne and Swiss Finance Institute

Markus Pelger
Stanford University

Ye Ye
Stanford University

Shrinking the Term Structure*

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

August 9, 2022

Abstract

We develop a conditional factor model for the term structure of treasury bonds, which unifies non-parametric curve estimation with cross-sectional asset pricing. Our factors correspond to the optimal non-parametric basis functions spanning the discount curve. They are investable portfolios estimated with cross-sectional ridge regressions and derived from economic first principles. Empirically, we show that four factors explain the discount bond excess return curve and term structure premium. Cash flows are covariances, as cash flows of coupon bonds fully explain the factor exposure. The term structure premium depends on the market complexity measured by the time-varying importance of higher order factors.

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 Michael Bauer, Darrell Duffie, Kay Giesecke, Tze Lai and Aaron Sidford for helpful comments.
†École Polytechnique Fédérale de Lausanne and Swiss Finance Institute, Email: damir.filipovic@epfl.ch
‡Stanford University, Department of Management Science & Engineering, Email: mpelger@stanford.edu.
§Stanford University, Department of Management Science & Engineering, Email: yeye@stanford.edu.

1 Introduction

The term structure of interest rates is a fundamental economic quantity for financial and macroeconomic research and applications. Risk factors form the core building block of modern asset pricing theory. This paper addresses the question of which risk factors explain the cross-section of treasury bond returns. This problem is challenging as the term structure and the underlying factors are unobserved and must be inferred from a relatively sparse set of noisy returns of treasury securities. In the context of the term structure, cross-sectional asset pricing is inherently linked to curve estimation. We provide a unifying framework that combines these two distinct literatures.

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 spans the discount bond return curve. This dual perspective is novel and important for two reasons. First, any cross-sectional asset pricing statement is subject to the limitations of the curve estimation method. The standard approach for modeling term structure factors assumes a panel of discount bond returns estimated using some ad-hoc method. However, we show that a too simplistic curve estimation results in an erroneously simplistic factor structure. Our extremely precise curve estimations provide a better description of the cross-sectional factor structure, which is reflected in the out-of-sample investment implications. Second, our dual perspective allows us to explain why certain factors occur.

We introduce a robust, flexible, and easy-to-implement method to estimate a low dimensional factor model for treasury securities. 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. Our method optimally learns a sparse set of basis functions in a reproducing kernel Hilbert space, which is derived from economic first principles, and provides the link between non-parametric curve estimation and factor modeling. 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. Setting up the objective function in terms of the aggregated return error and a smoothness measure uniquely determines 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. As a result, our solution is fully data driven.

The estimated discount bond return curve is given in closed form as a simple cross-sectional ridge regression of treasury bond returns on their normalized 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. The factors are portfolios of traded treasury securities, and therefore tradable themselves. Importantly, this implies feasible trading strategies to hedge against interest rate changes and immunize any default-free fixed income claim.

Our factor model can be viewed as a Barra model for treasury securities. In equity modeling, firm characteristics capture risk exposure. 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 smoothness of the discount bond return curve.

Our paper is the first to provide a conditional cross-sectional factor model for treasuries. The cash flows of treasuries play the same role as firm characteristics in equity modeling. We show that the factor loadings of the coupon bonds are functions of their cash flows and therefore time-varying. Two fixed income assets with the same cash flows have the same factor risk exposure. The loadings are directly related to conditional covariances with the factors. Hence, similar to the observation that "characteristics are covariances", we show that "cash flows are covariances" when modeling treasuries. We show empirically, that we obtain the same factor loadings from the conditional cash flow information as with time-series regressions on risk factors. 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 kernel formulation determines the coupon bond loadings.

We perform an extensive empirical study on excess returns of U.S. Treasury securities from 1961 to 2020. We show that four factors provide an excellent representation of the discount bond excess return curve. The portfolio weights of these factors in terms of discount bonds take the familiar shapes of slope, curvature and polynomials of increasing higher order. These shapes have already been documented for principal component analysis (PCA) based factors in among others Litterman and Scheinkman (1991), Crump and Gospodinov (2022) and Filipović, Pelger, and Ye (2022). 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 available for any single day without the need for panel data, and we can explain why they occur. Our latent factors correspond to the optimal sparse set of non-parametric basis functions to approximate smooth curves. The structure of PCA factors is a consequence of the nature of our estimated curves. On any given day, we obtain the latent factors from a cross-sectional regression on treasury cash flows. As this factor structure holds for all days, it implies that a PCA applied to a panel recovers the same type of factors. As we work with excess returns, we have mechanically removed a level factor, which would appear if we modeled factors in returns. Hence, our four factor model in excess returns corresponds essentially to the first five PCA factors in a panel of discount bond returns. The third and fourth factors, in addition to the familiar level, slope and curvature factors, capture the more complex curvature patterns in the term structure.

We show that our four factors explain the term structure premium in a cross-section of discount bond excess returns. We study the asset pricing implications for various factor models and report the Sharpe ratios, pricing errors and explained variation. The investment statements are meaningful as our factors and discount bonds are portfolios of actually traded assets. This is in contrast to the artificial discount bond returns based on common data sets such as Fama–Bliss or Gürkaynak, Sack, and Wright (2010). We demonstrate that including the fourth factor is crucial from an asset pricing perspective. It carries a large risk premium, which lifts the monthly out-of-sample Sharpe ratio of the implied stochastic discount factor to 0.85, while it lowers the pricing errors by a factor of three. The higher order factors in our model capture more complex curvature patterns in the discount bond excess return curve, and these patterns are actually the most important for the risk premium.

We contribute to the debate of how many factors are needed to represent the cross-sectional information in treasury bonds. The answer depends crucially on the objective. First, we show that four factors are optimal for a conditional factor model that explains out-of-sample the excess returns of traded coupon bonds. Explaining observed coupon bond returns instead of artificial discount bond returns is analogous to explaining individual stock returns instead of selected portfolio sorts in equity. Second, we show that we require four factors to explain the mean excess return of a cross-section of discount bonds. Third, we confirm that we also need the same four factors to explain the variation in forward returns after removing the mechanical correlation structure in excess returns. Conventional wisdom suggests that the level, slope and curvature PCA factors explain over 97% of the variation in zero-coupon yields and hence seem to capture most panel information. As shown also by Crump and Gospodinov (2022), the eigenvalues in a panel of discount bond returns are mechanically inflated, while the eigenvalues in a panel of forward returns are more informative. Hence, explained variation in panels of discount bond returns can be misleading. 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. Our findings also relate to weak factors. Our fourth factor captures a high risk premium but explains less variation. Building on similar findings as in Lettau and Pelger (2020a,b), we confirm that weak factors explaining less variation can be crucial in explaining the first moment in the cross-section.

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. Therefore, we introduce two novel measures for the state of the bond market. 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 higher order term structure factors. Our measures of the bond market condition are strongly related to the term structure risk premium. Remarkably, the sign of the risk premium of the first (slope) factor changes from positive to negative for increasing IT-VOL and T-COM. 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, our fourth factor serves as a hedge against the risk of changing market conditions. It has a negative payoff during low IT-VOL and T-COM periods, but earns a high positive risk premium during challenging times. As a result, the implied SDF based on all four factors is largely unaffected, and the overall Sharpe ratio has only a minor dependence on the market conditions. Our novel results complement prior findings on the dependency of the term structure premium on market conditions. This shares some similarities with inverted yield curves during recessions, but at a higher frequency.

We provide publicly available and regularly updated data sets of daily discount bond returns and factors based on our precise estimates.[1] This is the only fixed income data, where the discount bonds and factors are feasible portfolios of actually traded U.S. Treasuries. 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.

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 treasuries, where the risk factor loadings are instrumented by the normalized 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 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). Both set of 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. 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 based on limits to arbitrage. 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 our companion paper Filipović, Pelger, and Ye (2022) and the literature on non-parametric yield curve estimation. Both of our 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

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

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. In Filipović, Pelger, and Ye (2022), we provide one of the most extensive comparison studies between different yield curve estimation methods, which includes our own kernel ridge (KR) method, Fama and Bliss (1987), Liu and Wu (2021), Gürkaynak, Sack, and Wright (2007), and our own implementation of the Nelson–Siegel–Svensson model. In Filipović, Pelger, and Ye (2022), we find that our 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 depend 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 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 suggests 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 coupon bonds, instead of a panel of discount bonds. 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.

Our paper is complementary to the literature on bond price and yield forecasting. The starting point of the forecasting literature is a low dimensional representation of the cross-section of treasury bonds, for example, in terms of factors or a small number of discount bonds. The future yields or returns of this low dimensional representation are predicted using prior information, which allows to test important economic statements like the spanning or expectation hypotheses. Examples of work in this regard are Cochrane and Piazzesi (2005), Cooper and Priestley (2008), Ludvigson and Ng (2009, 2010), Duffee (2013), Cieslak and Povala (2015), Greenwood and Vayanos (2014), Filipović, Larsson, and Trolle (2017), Bauer and Hamilton (2018), and many others. Our cross-sectional factor model provides a low dimensional representation of the cross-section, which can be used as a forecasting target. Such forecasts would allow to study the conditional SDF representation and conditional time-varying risk premiums implied by our cross-sectional factors. Forecasting means conditional modeling in the time dimension by estimating the conditional moments of factors using prior information sets. The focus of this paper is on conditional cross-sectional factor modeling. We describe the cross-sectional information, that is the factor betas, as a function of time-varying cross-sectional information. These two dimensions of conditional modeling complement each other, and are analogous to the two complementary approaches in equity modeling of capturing conditional cross-sectional information with firm characteristics and estimating time-varying risk premiums by predicting factors with macroeconomic time-series.

This paper is organized as follows. In Section 2, we formulate the term structure estimation problem. In Section 3, we introduce the shrinkage solution and sparse factor model. In Section 4, we conduct a comprehensive empirical analysis. Section 5 concludes. The appendix collects all the proofs and additional empirical results.

2 Term Structure Estimation and Cross-Sectional Asset Pricing

2.1 Fundamental Problem

The fundamental elements for term structure modeling are the zero-coupon treasury bonds for different maturities. These discount bonds carry the same information as the yield or forward curves. As the collection of maturities may include any future day, we represent the discount curve as a function $d_t : [0, \infty) \to \mathbb{R}$, such that $d_t(x)$ denotes the price of a discount bond with face value one and time to maturity x at day t. However, we do not observe the discount bond prices but only a relatively sparse set of traded coupon treasury bonds. Hence, term structure modeling is fundamentally a curve estimation problem.

More specifically, for each day t, we denote the discount bond prices by $D_{t,j} = d_t(x_j)$, where $0 = x_0 < x_1 < \cdots < x_N < x_{N+1}$ are some fixed times to maturity.[2] We observe the prices $P_{t,1}, ..., P_{t,M_t}$ of M_t coupon bonds with cash flows summarized in the matrix C_t. This means that $C_{t,ij}$ denotes the cash flow of coupon bond i at x_j time units after day t. The law of one price connects the fundamental value of the observed coupon bonds with the unobserved discount bonds

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

through

$$P_{t,i} = \sum_{j=1}^{N+1} C_{t,ij} D_{t,j} \qquad \text{for } i = 1, ..., M_t. \tag{1}$$

Hence, we can represent any coupon bond as a portfolio of discount bonds, where the cash flows denote the portfolio weights. Absent arbitrage this portfolio has the same price as the security.

Note that we can equivalently express the relationship (1) with the cash flow matrix from the prior business day by shifting the cash flow index, that is, $P_{t,i} = \sum_{j=1}^{N} C_{t-1,ij+1} D_{t,j}$. The total return of the coupon bonds from $t-1$ to t, including their cash flow at t, are thus related to discount bond returns through

$$\frac{P_{t,i} + C_{t-1,i1} - P_{t-1,i}}{P_{t-1,i}} = \sum_{j=0}^{N} C_{t-1,ij+1} \frac{D_{t,j} - D_{t-1,j+1}}{P_{t-1,i}} = \sum_{j=0}^{N} Z_{t-1,ij} \frac{D_{t,j} - D_{t-1,j+1}}{D_{t-1,j+1}}, \tag{2}$$

where $Z_{t-1,ij} = \frac{C_{t-1,ij+1} D_{t-1,j+1}}{P_{t-1,i}}$ denote the normalized discounted cash flows of bond i. Returns are calculated for business days and the cash flow dates are adjusted accordingly.[3]

Given the cross-sectional asset pricing focus of this paper, the object of interest are excess returns over the risk-free overnight return $R_t^f = \frac{1-D_{t-1,1}}{D_{t-1,1}}$, which is the return of the discount bond maturing on the next business day. The relationship (2) carries over to excess returns. As $\sum_{j=0}^{N} Z_{t-1,ij} = 1$, we obtain the bond excess returns

$$R_{t,i}^{\text{bond}} = \sum_{j=1}^{N} Z_{t-1,ij} R_{t,j},$$

in terms of the excess returns of the discount bonds

$$R_{t,j} = \frac{D_{t,j} - D_{t-1,j+1}}{D_{t-1,j+1}} - R_t^f.$$

In matrix notation, this reads

$$R_t^{\text{bond}} = Z_{t-1} R_t \tag{3}$$

for the M_t-vector of bond excess returns $R_t^{\text{bond}} = (R_{t,1}^{\text{bond}}, \ldots, R_{t,M_t}^{\text{bond}})^\top$, the $M_t \times N$-matrix of normalized discounted cash flows $Z_{t-1} = (Z_{t-1,ij})_{i=1,\ldots,M_t,\, j=1,\ldots,N}$, and the N-vector of discount bond excess returns $R_t = (R_{t,1}, \ldots, R_{t,N})^\top$.[4]

Hence, modeling the term structure of discount bond excess returns is equivalent to modeling

[3] More specifically, we denote by $\Delta_t = m_t/365$ the time units between business days $t-1$ and t, for some multiplier $m_t \geq 1$. For regular consecutive days, we have $m_t = 1$. If a weekend or other bank holidays fall between $t-1$ and t, then $m_t \geq 2$. In this case, we replace $C_{t-1,ij+1}$ by $C_{t-1,ij+m_t}$ and $D_{t-1,j+1}$ by $D_{t-1,j+m_t} = d_t(x_j + \Delta_t)$, accordingly.

[4] 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 $R_{t,j}$ and $Z_{t-1,ij}$ for $j = 0$ in the specification of the vector R_t and matrix Z_{t-1}.

the discount bond excess return curve $r_t : [0, \infty) \to \mathbb{R}$, given by

$$r_t(x) = \frac{d_t(x) - d_{t-1}(x + \Delta_t)}{d_{t-1}(x + \Delta_t)} - R_t^f, \tag{4}$$

where Δ_t denotes the time units between business days $t-1$ and t. The discount bond excess returns follow from this curve as $R_{t,j} = r_t(x_j)$.

2.2 Cross-Sectional Factor Model

Term structure factor modeling is the joint problem of discount bond return curve estimation and cross-sectional factor modeling. A low dimensional factor model for the cross-section of discount bond returns applies if and only if a sparse set of basis functions spans the discount bond return curve. The standard approach in the literature for modeling term structure factors neglects this relationship and takes a panel of discount bond returns as given. This is problematic for at least two reasons. First, all the potential problems of the discount curve estimation carry over to the cross-sectional factor model. In particular, a too simplistic curve estimation, for example with the parametric Nelson–Siegel–Svensson model, neglects relevant complex shapes in the curve, and as a result implies an erroneous and overly simplistic factor structure for the cross-section. Second, the cross-sections of discount bonds are often selected rather arbitrarily, which may overweight some maturities and thus not reflect the properties of the actually traded coupon bonds.

Basis functions and cross-sectional factors for the discount bond return curve are just two dual perspectives on the same problem. Formally, a factor model corresponds to a set of $n \leq N$ basis functions that approximately span the discount bond excess return curve,

$$R_t \approx \underbrace{\beta}_{N \times n} \underbrace{F_t}_{n \times 1}.$$

The basis functions captured by the columns of β can be interpreted as "loadings" and only vary in the cross-section of maturities and not over time. The coefficients F_t on these basis functions can be interpreted as factors and are time-varying. The basis functions for our model will be based on a kernel that is derived from economic first principles. We can express a panel of discount bond returns in the usual matrix notation for factor models as

$$\underbrace{R}_{N \times T} \approx \underbrace{\beta}_{N \times n} \underbrace{F}_{n \times T}.$$

Our estimation approach will provide factors F_t that are portfolios of the traded assets and are thus tradable themselves. This is not the case for most curve estimators. The fact that our factors are traded assets has important investment implications. Indeed, if the term structure can be spanned by a sparse set of traded factors, then investors only need to trade in a few factor portfolios to replicate the full term structure. These factors are sufficient to hedge against all term structure risk and immunize any default-free fixed income claim.

A factor model in discount bond excess returns implies a conditional factor model in coupon bond excess returns, in view of (3),

$$R_t^{\text{bond}} \approx \underbrace{Z_{t-1}\beta}_{\beta_{t-1}^{\text{bond}}} F_t = \underbrace{\beta_{t-1}^{\text{bond}}}_{M_t \times n} F_t.$$

The loadings of the coupon bonds, $\beta_{t-1}^{\text{bond}} = Z_{t-1}\beta$, are functions of their cash flows and therefore time-varying. In term structure modeling the cash flows play the same role as firm characteristics in equity modeling. Two fixed income assets with the same cash flows have the same factor risk exposure. Under mild assumptions, the loadings are directly related to conditional covariances with the factors. Hence, similar to the observation that "characteristics are covariances", we will show that "cash flows are covariances" when modeling treasuries. The normalized discounted cash flows Z_{t-1} can be interpreted as "cash flow instruments" or "cash flow characteristics".

The analogy to equity asset pricing also carries over to test or basis assets. The panel of discount bond returns is the fixed income analogue of characteristic-sorted portfolios in equity asset pricing. Characteristic-sorted portfolios are assumed to have constant risk exposure as they model the same economic quantity over time, similar to the discount bonds. Thus, selecting factors for explaining a panel of discount bond returns shares similar issues as explaining characteristic-sorted portfolios in the equity context. Such a panel may overweight certain maturities, and hence important asset pricing factors might become weak. That is, while those factors do not explain much variation in the selected cross-section, they are important for explaining the term structure premium. Similar to the solution in the equity space, which is to explain individual stock returns instead of pre-selected portfolio sorts, we will estimate the factors that best explain the actually traded coupon bond returns, while leveraging the conditional factor structure. Hence, we combine a non-parametric estimation problem with a cross-sectional term structure perspective.

The focus of this paper is on cross-sectional asset pricing and how the risk premium of treasury bonds depends on the maturities of their cash flows. Therefore, we consider the unconditional risk premium of the panel of discount bonds. The unconditional term structure premium equals $\mathbb{E}[R_t]$ and is estimated as the sample average $\frac{1}{T}\sum_{t=1}^{T} R_t$. Under the additional assumption of a factor structure, absence of arbitrage implies that the term structure premium is explained by the term structure premium of bond factors as

$$\mathbb{E}[R_t] = \beta \mathbb{E}[F_t].$$

This is the same cross-sectional asset pricing perspective that is pursued for example in Fama and French (1993). Given the traded factors F_t, we can compute the maximal Sharpe ratio from the tangency portfolio $b_F = \text{Var}(F_t)^{-1}\mathbb{E}[F_t]$ of the mean-variance frontier that is spanned by F_t as

$$\text{SR}_F = \sqrt{\mathbb{E}[F_t]^\top \, \text{Var}(F_t)^{-1}\mathbb{E}[F_t]}.$$

As discussed, for example, in Kozak, Nagel, and Santosh (2020) and Lettau and Pelger (2020b), the implied stochastic discount factor (SDF) spanned by the factors F_t is given by

$$\mathrm{SDF}_t^F = 1 - b_F^\top \left(F_t - \mathbb{E}[F_t]\right).$$

3 General Shrinkage Solution

3.1 Discount Bond Excess Return Curve Estimation

Any asset pricing model for the term structure requires an estimate of the discount bond return curve. It is common in the literature to first estimate the discount curves d_{t-1} and d_t with parametric or non-parametric methods, and then derive the implied discount bond excess return curve given in (4). In this paper, we directly estimate the curve r_t that can best explain the observed coupon bond returns.

Due to market imperfections, observed bond excess returns R_t^{bond} deviate from their fundamental values (3), resulting in return errors ϵ_t,

$$R_t^{\text{bond}} = Z_{t-1} r_t(\boldsymbol{x}) + \epsilon_t,$$

where we write $\boldsymbol{x} = (x_1, \dots, x_N)^\top$ and $r_t(\boldsymbol{x}) = (r_t(x_1), \dots, r_t(x_N))^\top$, accordingly. Hence, the naive starting point to estimate the underlying discount bond excess return curve would be to minimize the mean squared return errors of the coupon bonds for each day t,[5]

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

However, the number of observed 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 discount bond excess 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 discount bond excess 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 assumptions on the desired smoothness of the discount curve.

The fundamental problem of estimating the discount bond excess return curve is a trade-off between fitting the returns of observed assets and rewarding its smoothness. We formulate the problem based on only these core principles without imposing any ad-hoc assumptions. The smoothness of the discount bond excess return curve is founded in economic first principles. Indeed, large sudden

[5] 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.

changes along the return curve imply trading strategies with extreme payoffs. Therefore, limits to arbitrage require a sufficiently smooth curve, and a smoothness reward can be interpreted as an arbitrage penalty. A natural measure of the smoothness of a curve $h : [0, \infty) \to \mathbb{R}$ is given by the weighted average of the second derivative

$$\|h\|_{\mathcal{H}_\alpha} = \left(\int_0^\infty h''(x)^2 \mathrm{e}^{\alpha x} dx \right)^{\frac{1}{2}} . \tag{5}$$

Accordingly, we consider the large space $\mathcal{H}_\alpha$ of all twice differentiable functions $h : [0, \infty) \to \mathbb{R}$ with $h(0) = 0$ and finite smoothness measure (5). Technically speaking, $\mathcal{H}_\alpha$ is a reproducing kernel Hilbert space, which is fully defined in Appendix A. Theorem 1 below shows that $\mathcal{H}_\alpha$ is the appropriate hypothesis space for the discount bond excess return curve. This makes our approach fully flexible and non-parametric. 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. 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.

Theorem 1

Under the absence of arbitrage and technical assumptions spelled out in the proof, the discount bond excess return curve r_t given in (4) is an element in $\mathcal{H}_\alpha$.

The curve estimation problem can now be formally expressed by the objective function

$$\min_{r_t \in \mathcal{H}_\alpha} \left\{ \underbrace{\frac{1}{M_t} \left\| R_t^{\text{bond}} - Z_{t-1} r_t(\boldsymbol{x}) \right\|_2^2}_{\text{return error}} + \underbrace{\lambda \left\| r_t \right\|_{\mathcal{H}_\alpha}^2}_{\text{smoothness}} \right\} \tag{6}$$

for the smoothness parameter $\lambda > 0$. The conceptual 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. Theorem 2 below provides the closed-form and highly tractable solution to (6). It simplifies to a kernel ridge regression (KR), which is straightforward to implement. The kernel selection is completely guided by the economic principle of a smooth discount bond excess return curve with measure (5), and the parameters of the optimal basis functions are learned from the market data. This is distinctively different from conventional kernel-smoothing methods in the literature, which select the kernel exogenously. As a result, our solution is fully data driven.

Our solution builds on insights in functional analysis and machine learning by leveraging the structure of reproducing kernel Hilbert spaces such as $\mathcal{H}_\alpha$. Reproducing kernel Hilbert spaces are particularly important in machine learning because of the celebrated representer theorem, which states that every function in a reproducing kernel Hilbert space 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.

More specifically, as shown in Filipović, Pelger, and Ye (2022), the reproducing kernel $k : [0,\infty) \times [0,\infty) \to \mathbb{R}$ of $\mathcal{H}_\alpha$, is given in closed form as

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

The general solution curve $\hat{r}_t(\cdot)$ to (6) is given as linear combination of the N kernel basis functions $k(\cdot, x_1), \dots, k(\cdot, x_N)$. 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 $\boldsymbol{K} = (k(x_i, x_j))_{1 \le i,j \le N}$, that is, the kernel function evaluated at the cash flow dates. We will interpret these coefficients as factors. In order to leverage the factor perspective, we orthogonalize the kernel basis functions. This rotation is desirable for interpreting the resulting factors. Thereto, we consider the singular value decomposition of the symmetric kernel matrix

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

with the diagonal matrix $S = \mathrm{diag}(S_1, \dots, S_N)$ of eigenvalues in descending order $S_1 \ge \dots \ge S_N$ and the corresponding orthonormal eigenvectors V such that $V^\top V = I_N$, which denotes the $N \times N$ identify matrix. We show in Appendix A that the kernel matrix $\boldsymbol{K}$ is strictly positive definite, so that all eigenvalues are strictly positive, $S_N > 0$. We can thus define the N rotated basis functions

$$u_i(\cdot) = \frac{1}{\sqrt{S_i}} \sum_{j=1}^{N} k(\cdot, x_j) V_{ji}, \tag{7}$$

which can easily be shown to be orthonormal in $\mathcal{H}_\alpha$.

Theorem 2 (Kernel–Ridge (KR) Solution)
The fundamental term structure estimation problem (6) *has a unique solution, which is given in closed form by*

$$\hat{r}_t(x) = \sum_{j=1}^{N} u_j(x) \hat{F}_{t,j} \tag{8}$$

where the factors $\hat{F}_t = (\hat{F}_{t,1}, \dots, \hat{F}_{t,N})^\top$ *are the excess returns of traded bond portfolios given by*

$$\hat{F}_t = \omega_{t-1} R_t^{bond} \tag{9}$$

with portfolio weights

$$\omega_{t-1} = \beta^\top Z_{t-1}^\top \left(Z_{t-t} \boldsymbol{K} Z_{t-1}^\top + \lambda M_t I_{M_t} \right)^{-1} \tag{10}$$

for the loadings

$$\beta = VS^{1/2}. \tag{11}$$

The vector of discount bond excess returns can be expressed as

$$\hat{R}_t = \beta \hat{F}_t. \tag{12}$$

The estimated factors $F_{t,j}$ are the excess returns of portfolios of traded coupon bonds. Hence, we obtain mimicking portfolios for discount bonds. We will estimate the risk-free overnight return R_t^f with the KR method of our companion paper Filipović, Pelger, and Ye (2022), which expresses the overnight discount bond as a portfolio of the traded coupon bonds. As a results, all discount bonds and factors are directly expressed as portfolios of traded assets. This is important as we can study the investment implications of feasible portfolios.

The factor perspective allows us to provide an equivalent formulation of problem (6). Instead of expressing it as a curve estimation problem, we can view it as conditional factor model problem in a large dimensional panel. Theorem 3 formalizes this alternative formulation and the shrinkage solution.

Theorem 3 (Conditional Factor Model Representation)
The fundamental term structure estimation problem (6) *is equivalent to the cross-sectional ridge regression*

$$\min_{F_t \in \mathbb{R}^N} \left\{ \frac{1}{M_t} \left\| R_t^{bond} - \beta_{t-1}^{bond} F_t \right\|_2^2 + \lambda \left\| F_t \right\|_2^2 \right\}, \tag{13}$$

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

$$\beta_{t-1}^{bond} = Z_{t-1}\beta$$

with $\beta = VS^{1/2}$ given in (11)*. The unique solution to* (13) *is given in closed form by the factors $\hat{F}_t = \omega_{t-1} R_t^{bond}$ in* (9)*, where the portfolio weights can be expressed in terms of β_{t-1}^{bond} by*

$$\omega_{t-1} = \beta_{t-1}^{bond\top} \left(\beta_{t-1}^{bond} \beta_{t-1}^{bond\top} + \lambda M_t I_{M_t} \right)^{-1}. \tag{14}$$

Hence, discount bond excess returns are described by an unconditional factor model $\hat{R}_t = \beta \hat{F}_t$, which implies the conditional factor model for the coupon bond excess returns $\hat{R}_t^{\mathrm{bond}} = Z_{t-1}\beta\hat{F}_t = \beta_{t-1}^{\mathrm{bond}}\hat{F}_t$. The factors are the excess returns of portfolios of discount bonds with portfolio weights given in terms of the discount bond loadings:

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

Note that $\beta^\top \beta = S$ is a diagonal matrix, and hence up to a scaling normalization the columns of β are the portfolio weights in terms of discount bonds. As we will show empirically, the loadings β and factors $\hat{F}_t$ are closely related to the principal components applied to a panel of discount bond excess returns estimated with our KR method. In fact, if the relative importance of the factors $\hat{F}_t$ stays the same over time, and they are uncorrelated, then PCA coincides with our factors and loadings. In more detail, under these conditions β are the eigenvectors of the covariance matrix of R_t. Hence, there is an intrinsic relationship between the PCA of a panel and our non-parametrically estimated basis functions, which we explain in more detail in Appendix A.

Cash flows correspond to the "characteristics" of the traded assets and determine their "covariances" with the factors. In other words, we have a cross-sectional factor regression problem, where we obtain the latent factors from cross-sectional regressions on the normalized discounted cash flows. This is the same perspective as in cross-sectional factor regressions on characteristics for stock returns. The difference is that there are no ad-hoc assumptions on the functional transformation of the characteristics, but the law of one price and the kernel formulation determines the coupon bond loadings. The cross-sectional regression is regularized with a ridge penalty, which rewards factors whose loadings are spanned by the dominating principal components in the kernel space. In other words, the ridge penalty puts larger weights on factors that provide smooth cross-sectional loadings.

3.2 A Sparse Factor Model

A sparse factor model corresponds to a sparse selection of the basis functions in Theorem 3. The fundamental challenge is to select an optimal set of sparse basis functions. A researcher might have a prior about a subset of $n \leq N$ basis functions, which should span the discount bond excess return curve. For example, this set could be the factors associated with the n largest eigenvalues of the kernel. More generally, we will consider the problem where a researcher selects a subset of n vectors from the orthogonal eigenvectors V. In this case, the optimization problem in Theorem 3 is restricted to search over the function space that is spanned by the sparse subset of basis functions corresponding to the eigenvectors V with indices in the selected set. The resulting sparse loadings $\beta_{t-1}^{n} \in \mathbb{R}^{M_t \times n}$ are based on the subset of $\beta_{t-1}^{\text{bond}}$ with the selected indices. The regularized factors, which solve the restricted problem (6), equal

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

We 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 factors, that minimize the excess return errors of traded coupon bonds, while including a no-arbitrage penalty in terms of a smoothness reward. We achieve this by adding a lasso selection penalty to the optimization problem in Theorem 3, which leads to

$$\min_{F_t \in \mathbb{R}^N} \Bigg\{ \underbrace{\frac{1}{M_t} \left\| R_t^{\text{bond}} - \beta_{t-1}^{\text{bond}} F_t \right\|_2^2}_{\text{return error}} + \underbrace{\lambda \left\| F_t \right\|_2^2}_{\text{smoothness}} + \underbrace{\lambda_1 \left\| F_t \right\|_1}_{\text{selection}} \Bigg\}. \tag{15}$$

The lasso penalty $\lambda_1 \|F_t\|_1$ in problem (15) selects a sparse solution. There is a one-to-one mapping between the values of the lasso penalty 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 lower order 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, that is, $M_t = N$, $Z_{t-1} = I_N$, and $\beta_{t-1}^{\text{bond}} = \beta = VS^{1/2}$. In this case, the optimization problem (15) has a closed-form solution given by

$$\hat{F}_{t,j} = (S_j + \lambda N)^{-1} \left(S_j^{1/2} |v_j^\top R_t^{\text{bond}}| - \lambda_1 \frac{N}{2} \right)^+ \cdot \operatorname{sign}(v_j^\top R_t^{\text{bond}}), \tag{16}$$

where $a^+ = \max\{a, 0\}$, and v_j denotes the jth eigenvector of $\boldsymbol{K}$, that is, the jth column of V. Both λ_1 and λ upweight eigenvectors corresponding to larger eigenvalues in the basis function space.[6] 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.

Appendix A provides further theoretical details and the proofs for all the theorems and statements. Equipped with our novel estimator, we perform an extensive empirical analysis in the next section. We will refer to the full KR model as the estimator given in Theorem 3 and label a model with n selected factors from the kernel space as the KR-n factor model.

4 Empirical Results

4.1 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 sampling 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, as we study the time-series of treasuries 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.[7] 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

[6] Note that the non-regularized solution in this case would be $\hat{F}_{t,j} = S_j^{-1/2} v_j^\top R_t$.

[7] 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 securities 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 overnight 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 3. As robustness test we show that our results are not affected when we use the risk-free overnight return implied by the 1-month risk-free rate of Fama and French (1993).[8] We specify the normalized discounted cash flows Z_{t-1} based on the discount curve prevailing at $t-1$ estimated in Filipović, Pelger, and Ye (2022).

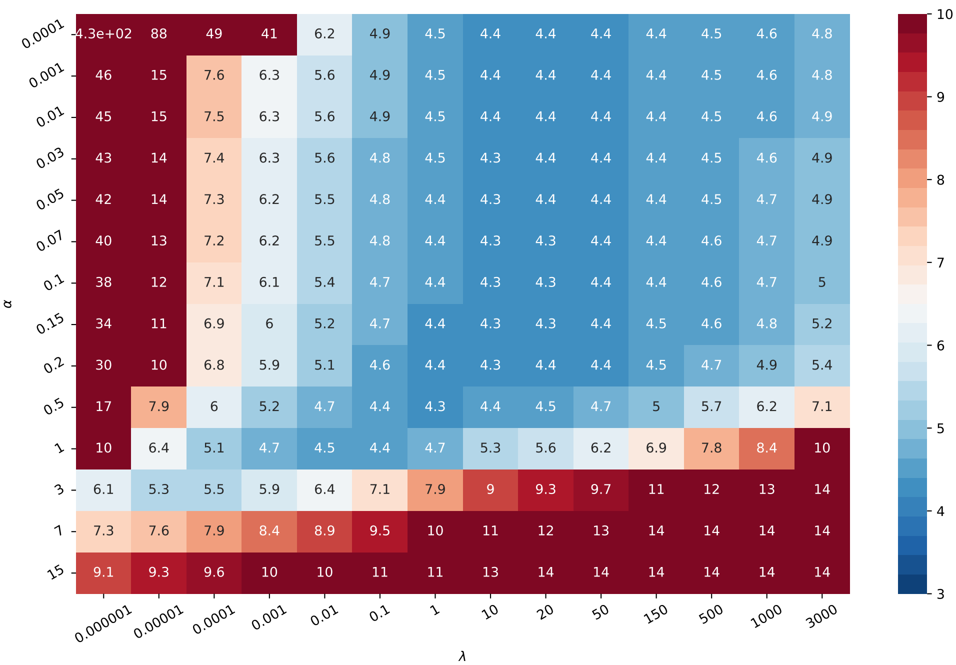
4.2 Term Structure Estimation

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 coupon 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.

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 $\lambda = 10$. Estimating the term structure of returns without a smoothness penalty results in 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

[8] The 1-month risk-free rate of Fama and French (1993) is based on the discount bond prices of Fama and Bliss (1987). As robustness test we specify the overnight return such that this daily return compounds to the 1-month return corresponding to the 1-month risk free rate.

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.

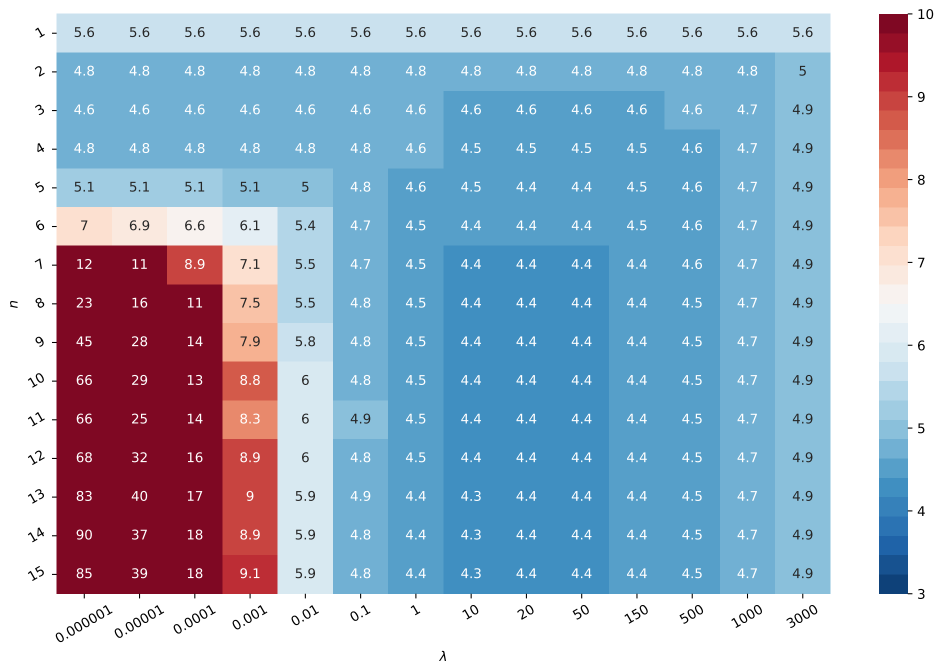
100 provide close results. We select $\alpha = 0.05$ and $\lambda = 10$ as the baseline parameters for the full KR model.

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 decomposition, that is, in increasing curvature. 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.[9]

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 sparse factor model KR-n. We study these models in more detail in the next sections.

[9] Figure A.2 in the Appendix shows the 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

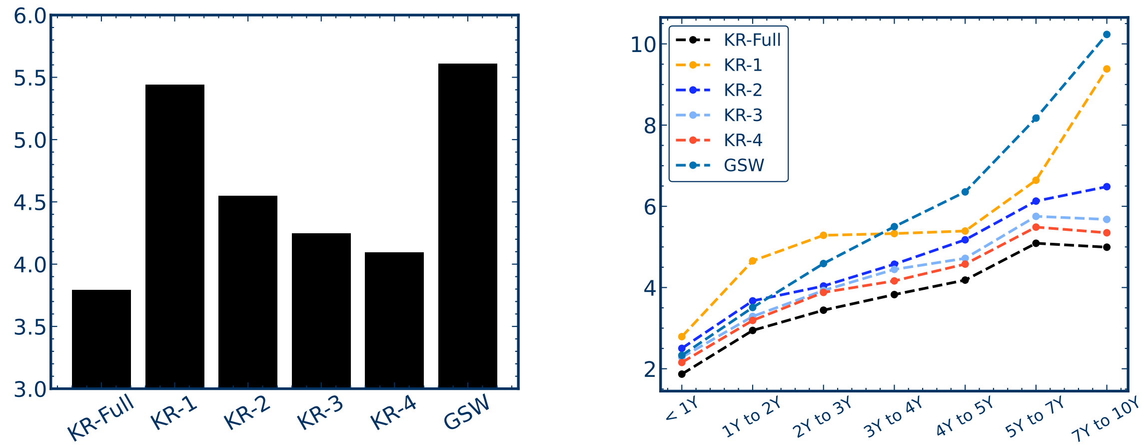


This figure shows the cross-validation excess return RMSE of the KR-n factor model in basis points (BPS) as a function of the smoothness parameter λ and the sparse number of factors n. Factors are added by the order implied by the kernel decomposition, that is, in decreasing order of the eigenvalues 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 A.3 in the Appendix 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. We are going to revisit the structure and the shapes fitted by the factors. 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.

It is sufficient to compare the KR factor model with the full KR specification. Our companion paper Filipović, Pelger, and Ye (2022) has 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 price estimates are then used to calculate the implied returns. Note that in contrast to the KR methodology the returns of alternative parametric or nonparametric

Figure 3: 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 normalized discounted cash flow matrices use the KR respectively GSW discount curves. The one-period KR risk-free rate is used. The results are calculated using daily data from June 1961 to December 2020.

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.

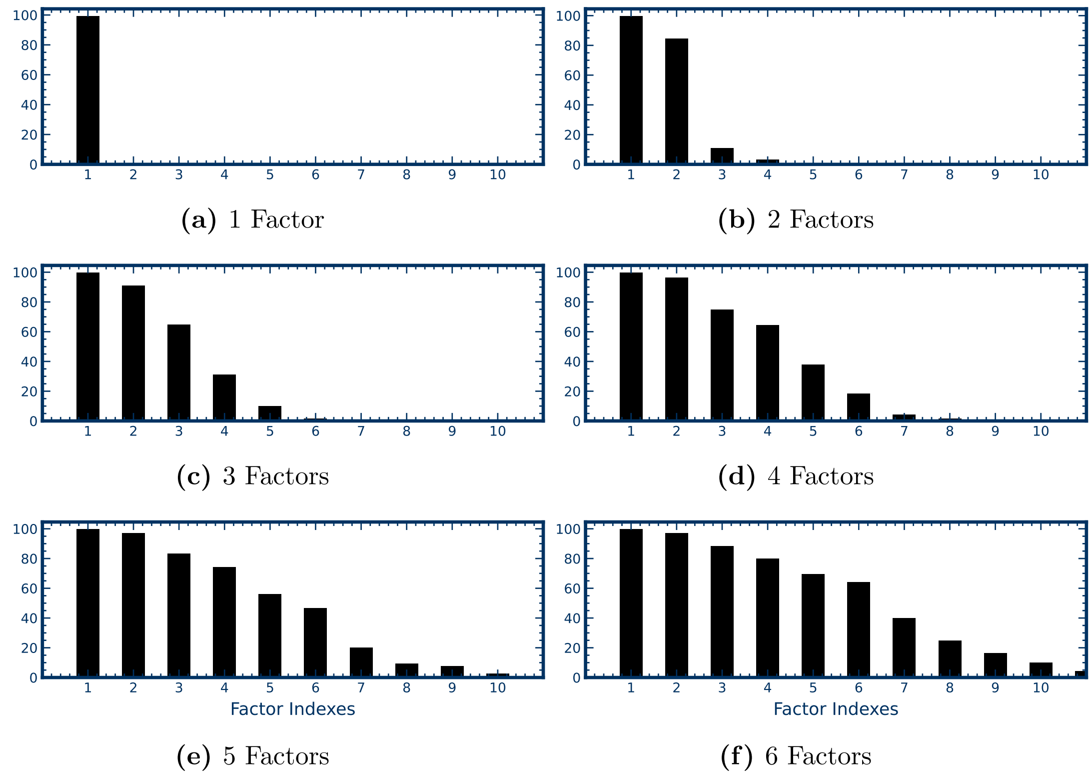
Figure 3 illustrates the small magnitude of the errors of KR method. We report the RMSE of observed securities for the full KR model, KR-n with $n = 1, ..., 4$ factors and estimates of Gürkaynak, Sack, and Wright (2010) labeled as GSW. The RMSE are calculated in-sample for all maturities or separately 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.

4.3 Selection of Term Structure Factors

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 (15) to select n factors, instead of selecting them based on the kernel order.

Figure 4 displays the fraction of times when the ith factor is selected by LASSO in an n-

Figure 4: Frequency (%) of factors selected by LASSO



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 every quarter in the sampling period, and we 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.

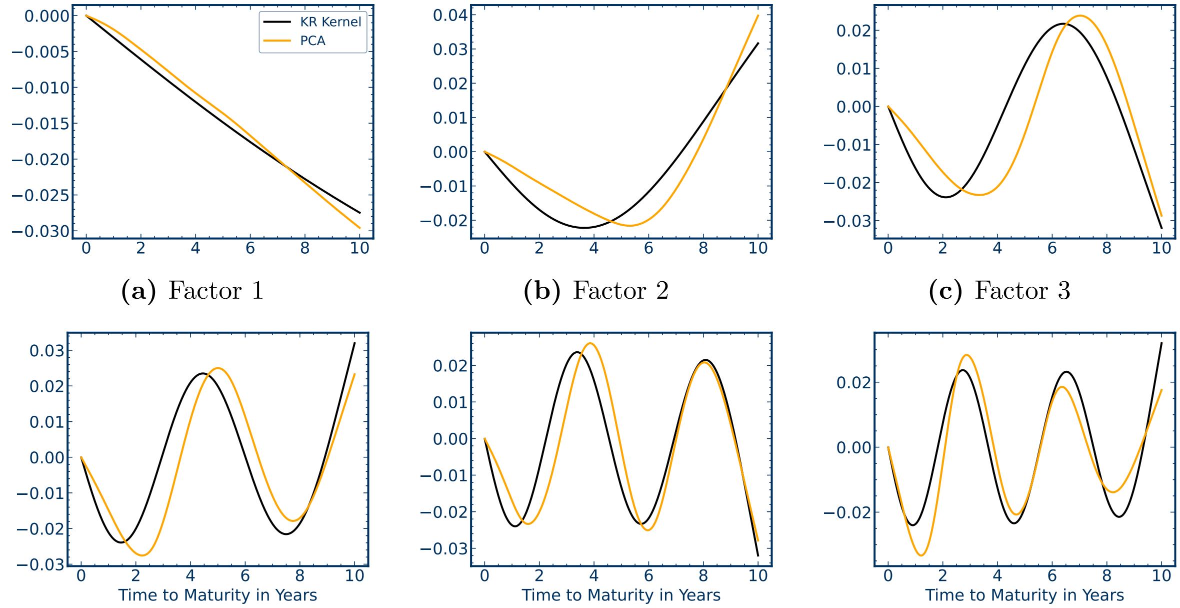
factor model. The selection is run separately on the last day in every quarter and the results are aggregated over the full sample. We observe that with a high probability an n-factor model will exactly select the first n factors based on the kernel order. For the first two factors this happens almost 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.

4.4 Factor Structure

We can interpret the structure of the KR factors from their portfolio weights. 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 5 shows the loadings β of the KR factors and the eigenvectors of PCA factors obtained from a panel of discount bond excess returns estimated with the full KR model. The loadings of the norms are normalized to one for better comparability. The eigenvectors of the PCA also correspond

Figure 5: 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

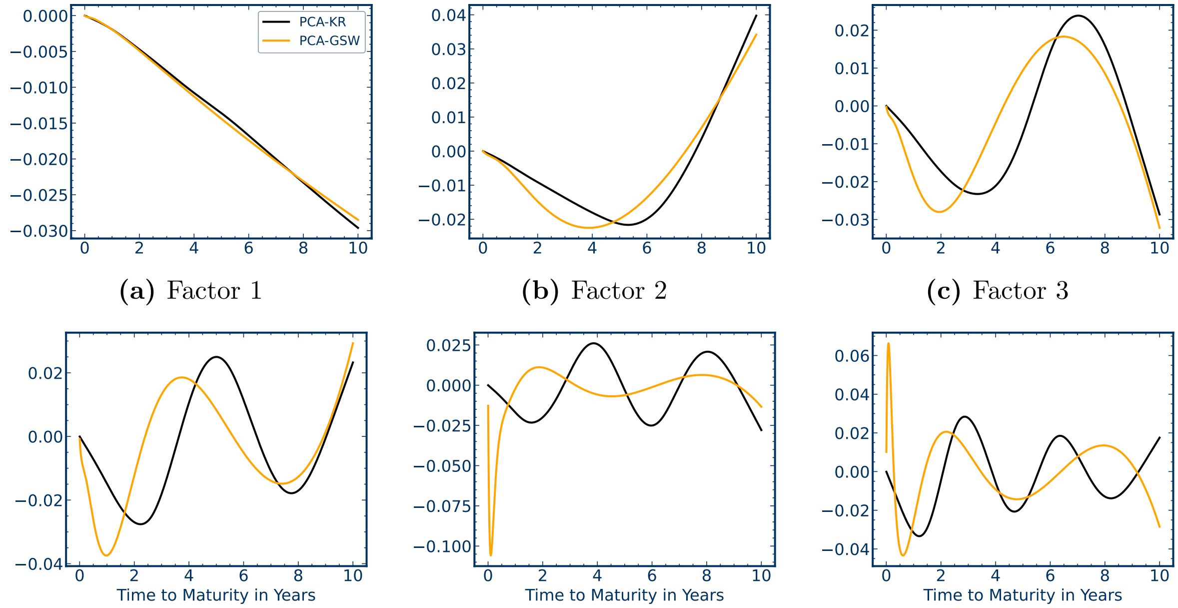
This figure shows the normalized loadings β of the first 6 KR factors and the loadings of the first 6 PCA factors on a panel of discount bond excess returns. The norm of the loadings is normalized to one. The loadings correspond to portfolio weights on discount bond excess returns to construct the KR 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.

to the loadings and portfolio weights of the PCA factors. First, we observe that the KR factors and PCA factors have very similar portfolio weights and capture 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 representation. Higher order factors have a higher curvature and capture more complex patterns. 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.

It is an important finding that KR and PCA factors represent the same factor model. As observed in Figure 5 the structure of the portfolio weights of KR and PCA factors is extremely close. Table 1 further confirms the similarity between KR and PCA factors. The explained variation of KR and PCA factors is essentially the same. Furthermore, the high generalized correlations show that the two set of factors span the same space.[10] Indeed, the average generalized correlation

[10]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

Figure 6: PCA factor loadings of KR and GSW 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 and GSW, respectively. The sample is daily data from June 1961 to December 2020.

between four KR and PCA factors is 0.94, which means that rotated PCA factors are almost perfectly correlated with KR factors.

Importantly, the PCA factors of a panel of discount bond returns estimated with the full KR are not mechanically the same as the KR factors. It is an empirical finding that the factor structure estimated with PCA represents a model that is close to the KR factor structure. First, PCA factors are by construction uncorrelated, while this does not need to hold for KR factors, which, in fact, are weakly correlated. Second, even if KR factors were uncorrelated, it is not mechanical that they have the same order as PCA factors. Hence, our findings shed new light on why we observe the slope and curvature type patterns in term structure PCA. These patterns arise naturally because these are the optimal basis functions to span the discount bond excess return curve on each day. Hence, if we aim at the best 2-factor model, we obtain the slope and curvature factors as they are the optimal basis functions with the least curvature to explain observed coupon bond excess returns on each day. As these factors optimally explain the non-parametric estimation problem on each day, they also explain best the panel of discount bond excess returns.

generalized correlation we first project out the subspace that spans the linear combination for the first generalized correlation and then determine the highest 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 $\mathrm{Cov}(F,G)\,\mathrm{Var}(F)^{-1}\,\mathrm{Cov}(F,G)\,\mathrm{Var}(G)^{-1}$. If $n = n_G = 1$, it is simply the conventional correlation. We report the average generalized correlation between n KR and PCA factors. Pelger (2020) provides a further discussion on generalized correlations.

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.

The KR factors are available for each 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.

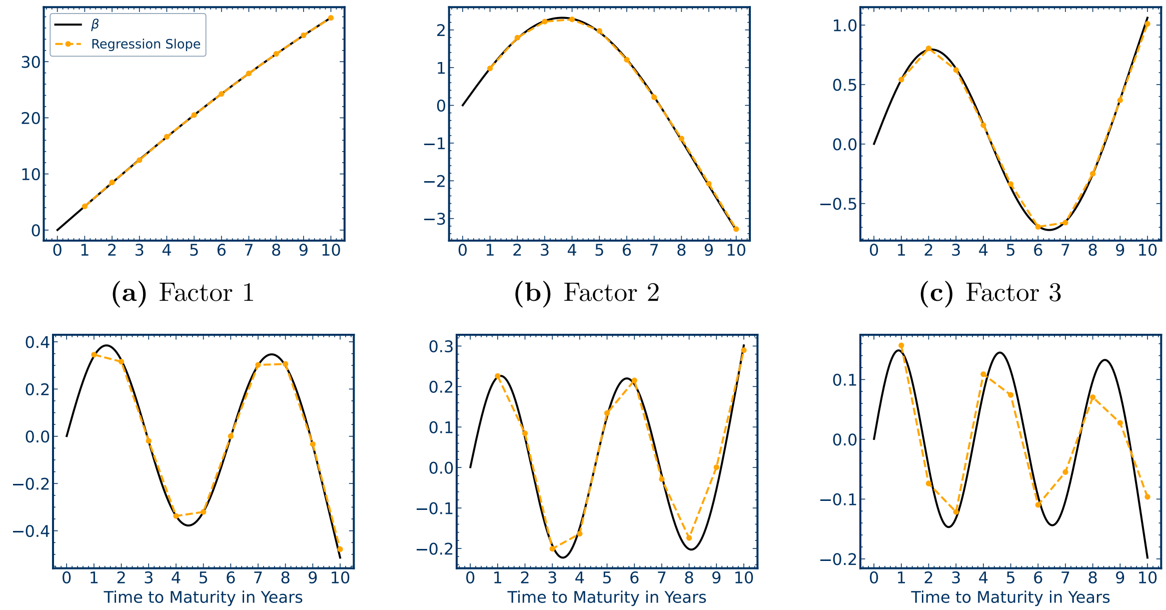
Cross-sectional term structure modeling is always a joint non-parametric curve estimation and cross-sectional asset pricing problem. The slope and curvature patterns in KR and PCA factors arise because these are the dominant eigenvectors of a kernel that explains the discount bond excess return curve well. For comparison, Figure 6 includes the PCA loadings from a panel of discount bond excess returns estimated with GSW. The first three PCA loadings have a similar shape, which capture polynomial basis patterns. However, the loadings of the higher order GSW PCA factors are distorted. A discount bond excess 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 return panels based on GSW, as the Nelson–Siegel–Svensson model omits those patterns. The cross-validation RMSE in Section 4.2 suggests that we need at least four factors to explain excess returns of observed coupon bonds. Selecting the number of factors on misspecified zero-coupon bond returns, as estimated for example with GSW, can be misleading for the correct number of factors. We will revisit the question about the number of factors later and show that a more precisely estimated discount bond excess return curve requires at least four factors.

4.5 Cash Flows are Covariances

The conditional KR factor loadings, based on cash flows, equal the time-series regression loadings. The loadings of KR factors are completely determined by the cash flows of the bonds without the need to estimate them. The unconditional loadings of asset pricing factors are usually estimated in time-series regressions. We show that for term structure modeling both are the same.

We start with a panel of 10 discount bond 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

Figure 7: 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

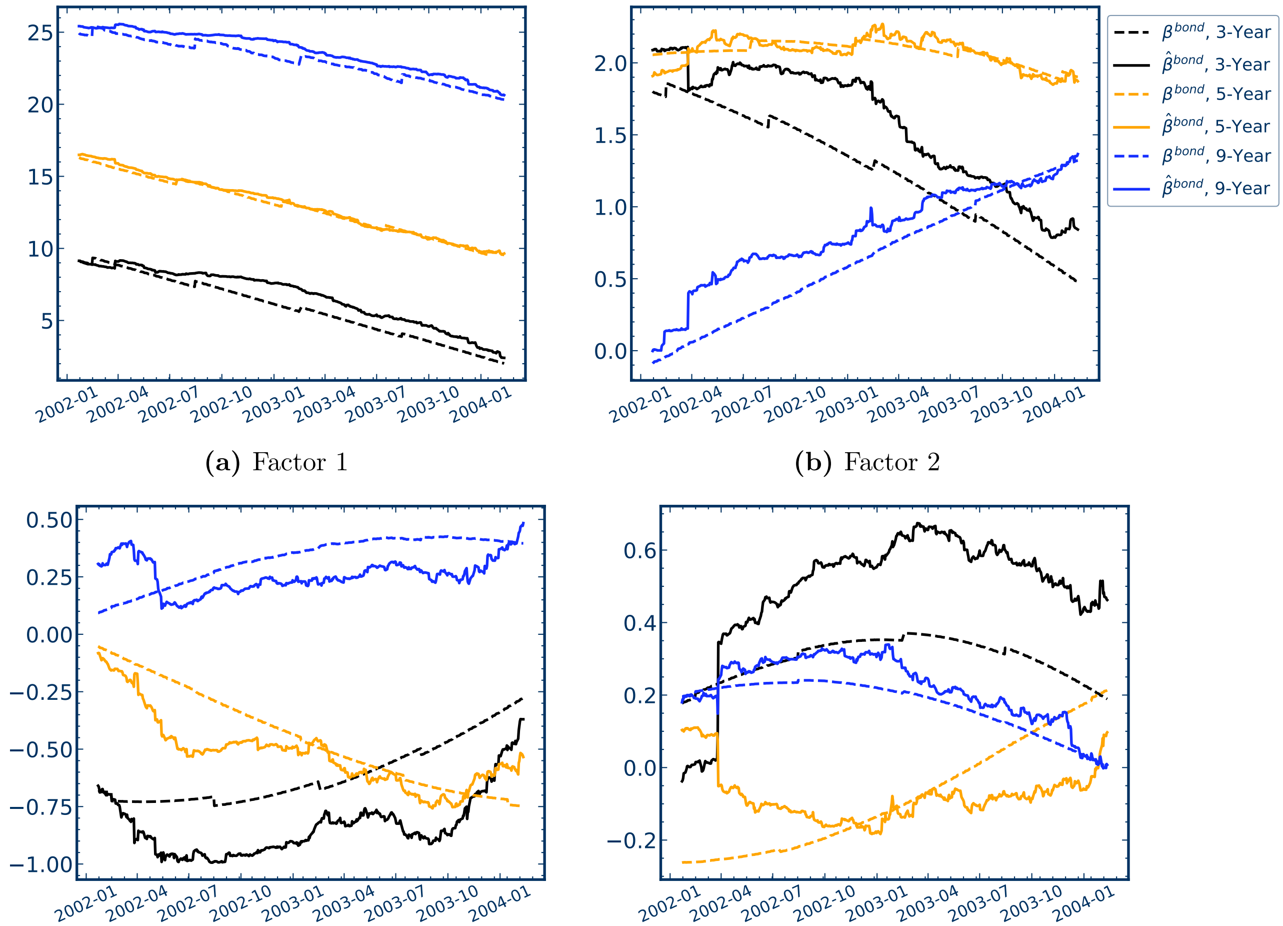
This figure shows the conditional loadings β and time-series regression slopes of KR factors on discount bond excess returns. The panel of ten discount bond excess returns with maturities ranging from 1 to 10 years is estimated with the full KR model. The time-series regression slopes are the coefficients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

structure factors. Figure 7 shows the loadings of these 10 discount bond returns on the first 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 of bond returns to obtain the loadings on discount bonds.

Coupon bonds have time-varying 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 8 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 or from the time-series.

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 show that "cash flows are covariances" for the fixed income. We obtain the same loadings for term structure factors

Figure 8: 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

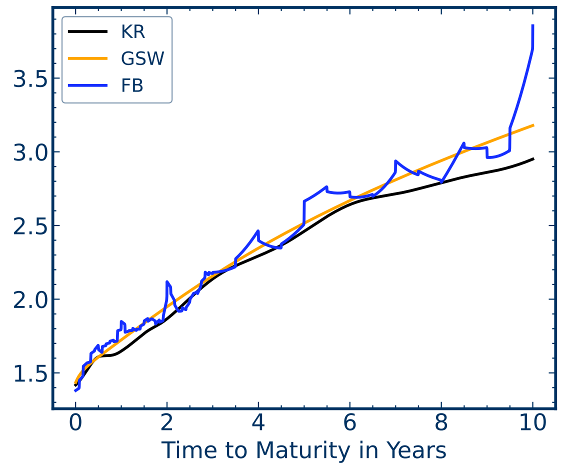
This figure shows the conditional loadings $\beta_{t-1}^{\text{bond}}$ and time-series regression slopes of the first four KR factors on three representative coupon excess returns. The dashed lines display the conditional loadings $\beta_{t-1}^{\text{bond}} = Z_{t-1}\beta$, which are determined by the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

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.

4.6 Term Structure Premium

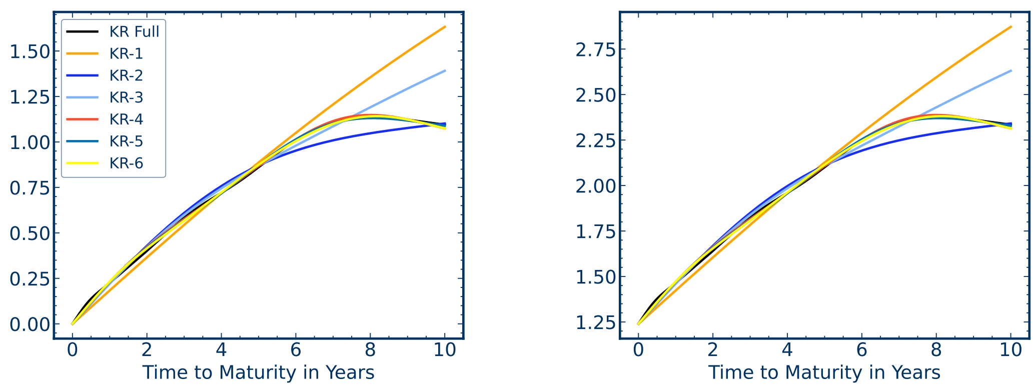
Statements about the term structure premium depend on the estimation method. Figure 9 shows the term structure premium implied by different estimators. The sample is restricted to the time-series for which the Fama and Bliss (1987) estimates are available. The figure shows the mean of the estimated discount bond returns. The term structure premium follow from subtracting the average risk-free overnight return. Overall, the different estimates share a similar shape with a

Figure 9: Term Structure Premium implied by different estimators



This figure shows expected one-day discount bond returns implied by the full KR, GSW, and FB estimates. The KR estimator directly estimates the term structure of returns, while GSW and FB estimate yields for discount bonds. The discount bond returns are averaged over time from August 1971 to December 2013, as the 10-year rate is unavailable in the FB data set until August 1971. Numbers are in basis points (BPS).

Figure 10: Expected returns and excess returns of discount bonds for different number of factors



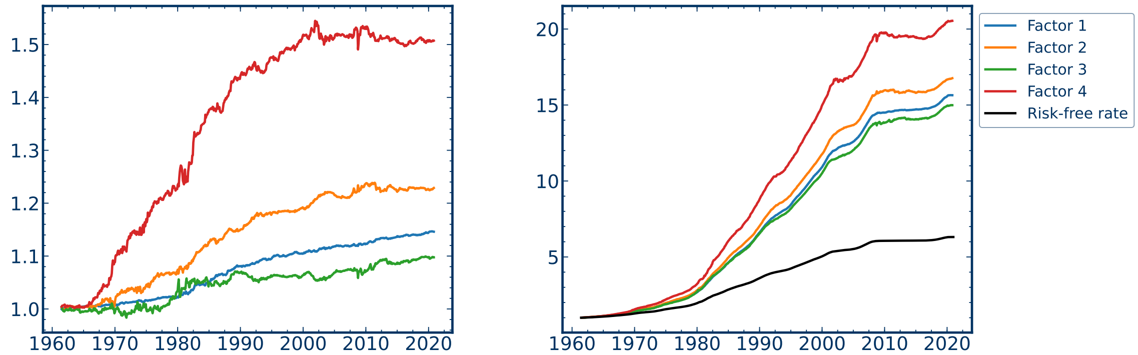
(a) Average excess returns of discount bonds **(b)** Average returns of discount bonds

This figure shows average excess returns (left) and returns (right) of discount bonds given by KR models for different number of factors. We consider the full KR model and one to six KR factors. Numbers are in basis points (BPS). The sample is daily data from June 1961 to December 2020.

premium increasing in maturity. The estimates of Fama and Bliss (1987) are not smooth, and have the largest deviations for longer maturities. This is in line with the findings in our companion paper Filipović, Pelger, and Ye (2022). The term structure premium implied by Gürkaynak, Sack, and Wright (2010) seems to be more reliable, but is restricted in the shapes that it can capture.

A small number of KR factors can explain the term structure premium. Figure 10 shows the term structure premium for the full KR model and for one to six KR factors. 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 we need at least four factors for explaining the

Figure 11: 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 securities. The risk-free overnight 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 2: Sharpe ratios of factors and implied SDF

	KR Factors			PCA Factors		
n	Factor	SDF	SDF(OOS)	Factor	SDF	SDF (OOS)
1	0.47	0.47	0.51	0.35	0.35	0.42
2	0.43	0.67	0.69	0.30	0.46	0.52
3	0.26	0.74	0.75	0.19	0.50	0.49
4	0.34	0.84	0.85	0.37	0.62	0.66
5	0.21	0.88	0.86	0.25	0.67	0.66
6	0.20	0.93	0.89	0.41	0.79	0.74

This table shows the annualized Sharpe ratios for different KR and PCA 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 or PCA 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 PCA factors are estimated on a panel estimated with the full KR model. The sample is daily data from June 1961 to December 2020.

term structure premium. Modeling excess returns is without loss of generality as we can directly revert them back to returns as shown in Figure 10(b).

The KR factors imply profitable investment opportunities. 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 Figure 11 we study the profitability of the factor portfolios and report the cumulative return and excess returns for the first four KR factors. We have observed in Figure 10 that the fourth factor is necessary to explain the term structure premium. Figure 11 shows that this fourth factor carries a large risk premium. Table 2 reports the Sharpe ratios, which are large for the fourth factor as well.

We construct the implied SDFs based on the first n KR factors and compare the Sharpe ratio of the SDF with the SDF based on the first n PCA factors. Table 2 reports the in- and out-of-sample results. The in- and out-of sample Sharpe ratio does not increase much after using four factors. However, the fourth factor boosts the Sharpe ratio out-of-sample to 0.85, emphasizing again its important asset pricing implications. While PCA factors share the same qualitative properties as KR factors, they are noisier estimates with lower Sharpe ratios. Overall, this is evidence that a four factor model is needed to explain the term structure premium. The results are robust to using the risk-free rate implied by Fama–French, as shown in Figure A.4 and Table A.2 in the Appendix.

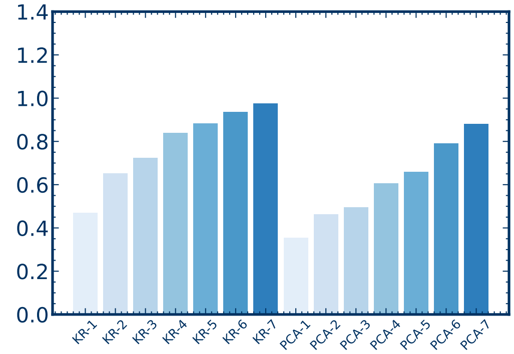
4.7 Term Structure Asset Pricing

We study the cross-sectional term structure asset pricing implications in the conventional panel setup. The exposure to risk factors is estimated from time-series regressions. Our panel consists of daily returns of 10 discount bonds estimated with the full KR methods for yearly maturities ranging from one to ten years. We compare how well different set of factors explain the risk premium and variation of the test assets. The factors include the KR factors, 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 starts in July 1963.

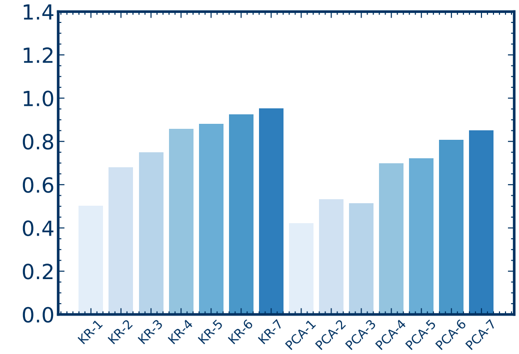
Figure 12 summarizes the asset pricing results. Panel (c) and (d) show that we require four KR factors to obtain small pricing errors in- and out-of-sample. These results are consistent with the Sharpe ratio of the factors. Similarly, four PCA factors strongly lower the pricing errors. A large part of 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 the next section, we will revisit the question of how many factors we need to explain the variation in discount bonds.

In Figure 13, we study the pricing errors for the ten discount bonds in more detail. We report the in-sample pricing errors from time-series regressions for each of the ten discount bonds. We also 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 (the level is taken out as 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

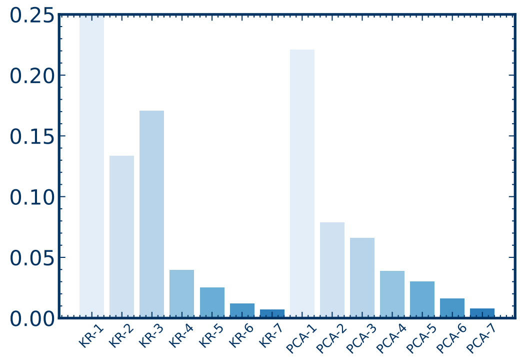
Figure 12: Asset Pricing Results



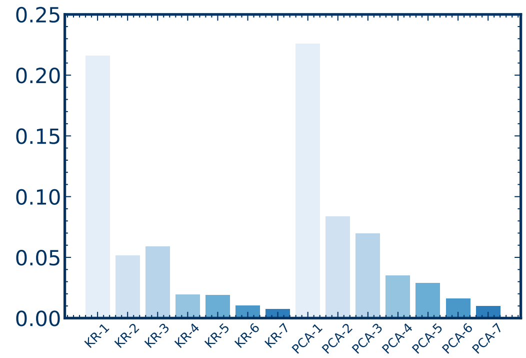
(a) SR (Full-Sample)



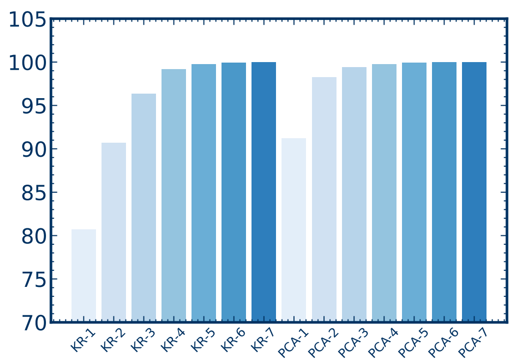
(b) SR (Out-of-Sample)



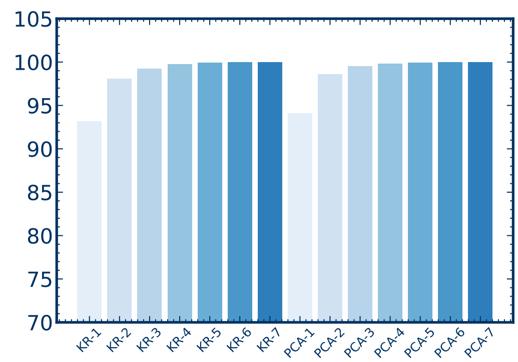
(c) RMS$_\alpha$ (Full-Sample)



(d) RMS$_\alpha$ (Out-of-Sample)



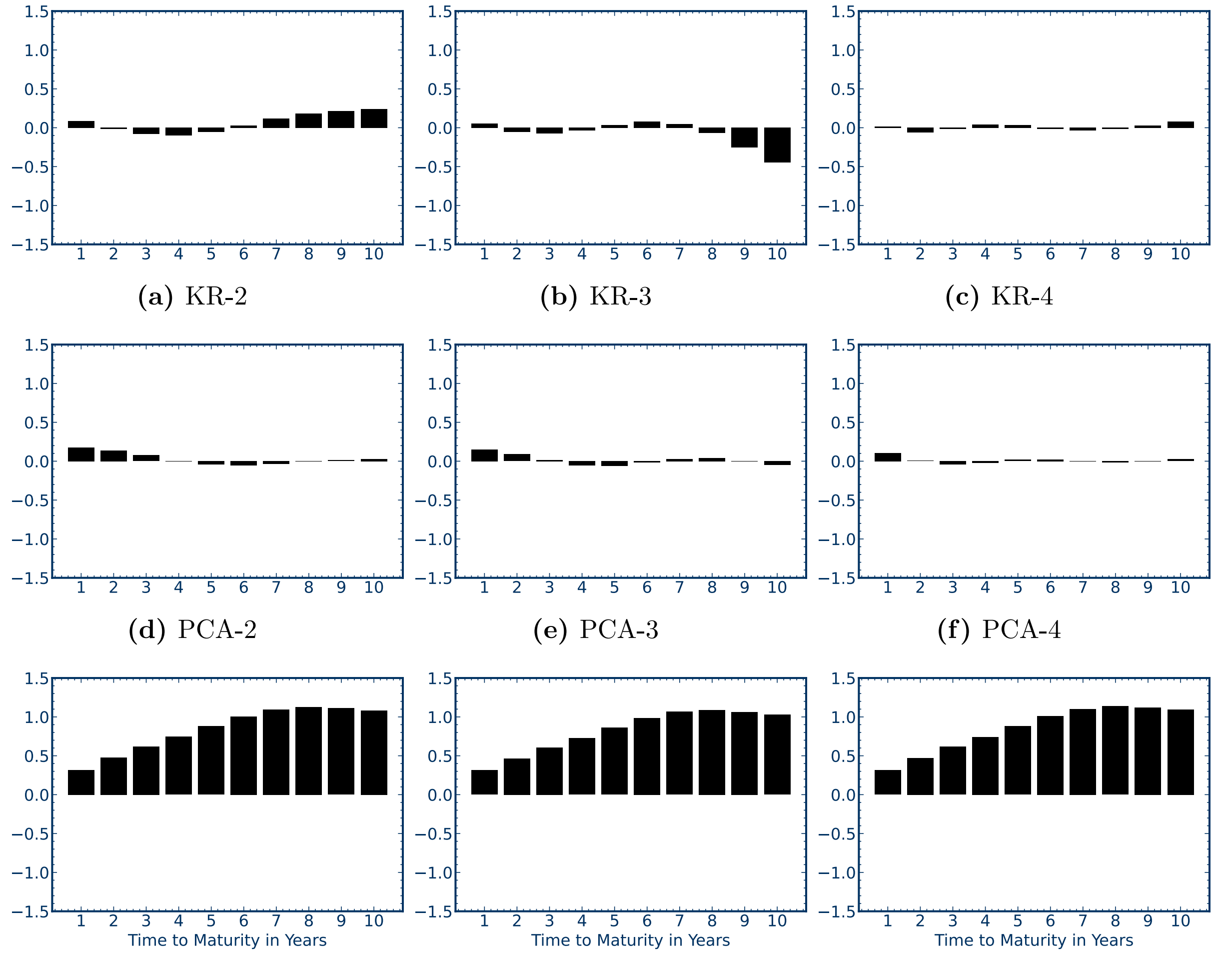
(e) Explained Variation (%, Full-Sample)



(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 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 overnight return is the estimate from the KR discount curve. The sample is daily data from July 1963 to December 2020.

Figure 13: 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 overnight return is the estimate from the KR discount curve. The sampling period is from July 1963 to December 2020.

errors are close to the mean returns of the bonds. Table A.3 in the Appendix shows that pricing errors with 4 KR factors are not only economically small, but for many 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.3 and Figure A.5 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, corresponding to the three factors level, slope and curvature in the return space, to price the complete cross-section. Table A.3 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.

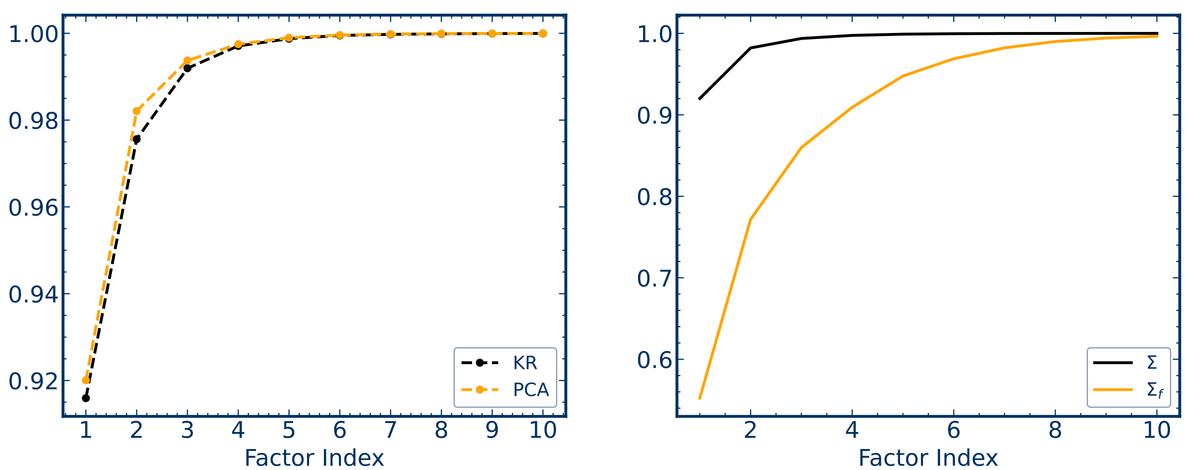
Our results are robust to the choice of the next-day risk-free rate. Figure A.6 in the Appendix shows that we obtain very similar results with the same takeaways for the Fama–French risk-free rate.

4.8 How Many Term Structure Factors?

The number of factors depends on the evaluation objective. The previous subsections have provided arguments that a good approximation of the term structure in the excess return space requires four factors. First, we have shown in the out-of-sample cross-validation analysis of Figure 2 that we need four factors to explain the observed coupon bond excess returns. Second, we have shown in Figures 10 and 12 that we also need four factors to explain the term structure premium. In this section we will argue that explaining the covariance for an appropriate panel also requires at least four factors.

The estimation of the number of latent term structure factors is often guided by explaining variation in a panel of discount bond returns. Figure 14(a) plots the explained variation in the

Figure 14: Explained variation and eigenvalue structure for different number of factors

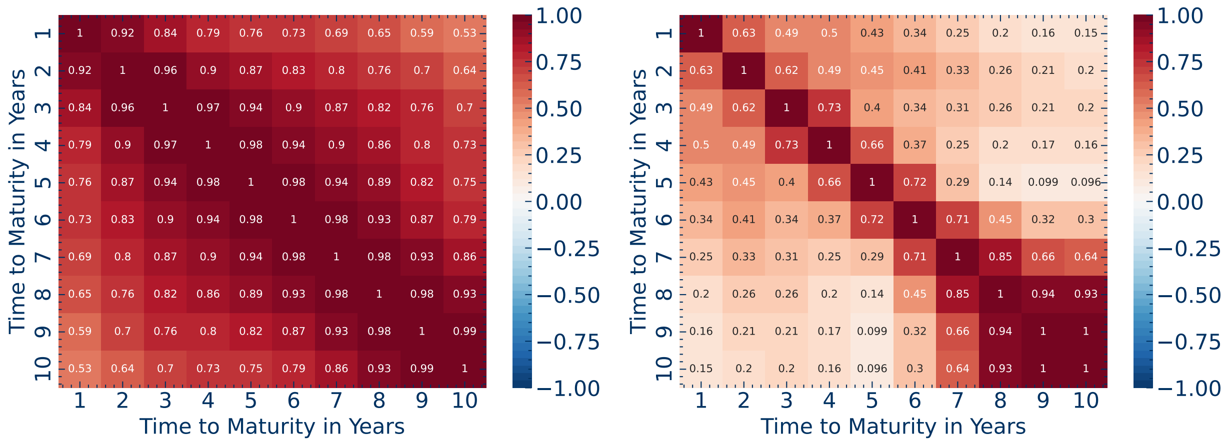


(a) Explained variation in discount bond excess returns for KR and PCA factors

(b) Normalized cumulative sum of eigenvalues 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.

Figure 15: Correlation matrices of discount bond excess returns and forward returns



(a) Correlation of excess returns **(b)** Correlation of 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.

full panel of discount bond excess returns for all daily maturities based on the full KR model for different number 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. 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 object 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, while 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 $R_{t,i}^{\text{forward}} = R_{t,i} - R_{t,i-1}$. They correspond to a forward agreement for a future one period risk-free investment. The excess 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 return 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 14(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 level, slope and curvature) would only explain around 77% of the covariance structure.

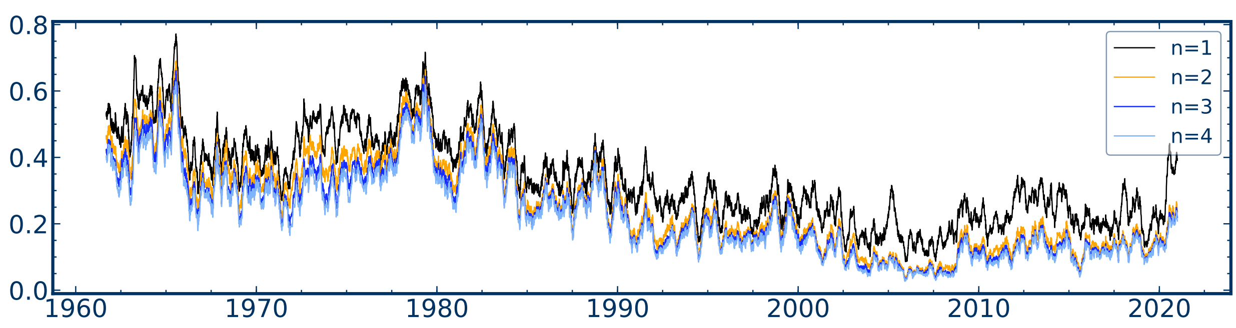
Figure 15 shows the correlation structure of a panel of discount bond excess returns and forward returns. We consider 10 assets with yearly maturities. The heatmap in Figure 15(a) illustrates the mechanical dependency in discount bond excess returns, which inflates the eigenvalues. In contrast, the correlation 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 factors to explain the correlation for the more informative forward returns.

4.9 Measures of Bond Market Conditions

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. However, 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 day. In Figure 16 we show the 3-month moving average 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.

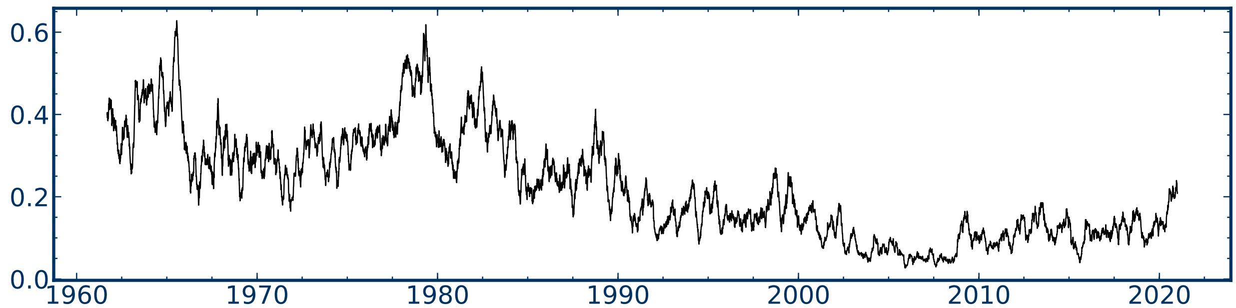
We introduce two novel measures for the state of the bond market: IT-VOL and T-COM. The

Figure 16: Unexplained variation by the first 4 KR factors

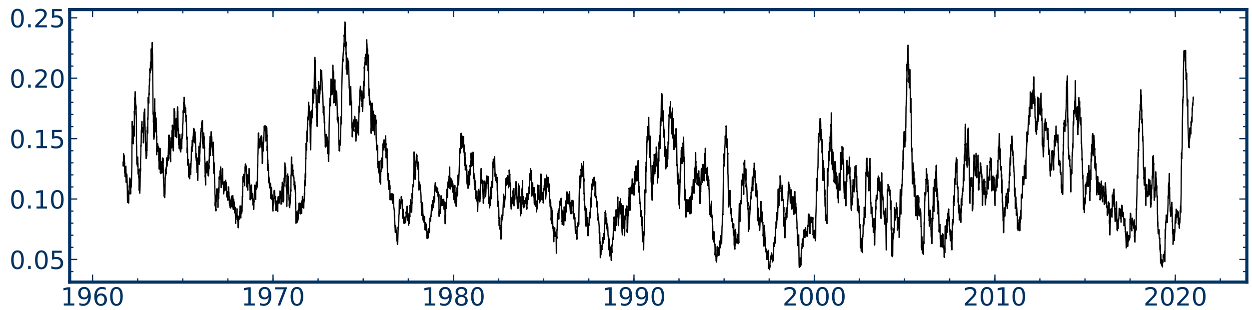


This figure shows the 3-month moving average of the unexplained cross-sectional variation for 1 to 4 KR factors. The unexplained cross-sectional variation of a factor model is normalized by the overall cross-sectional variation on that day.

Figure 17: Time-Series of Market Condition Measures



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



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

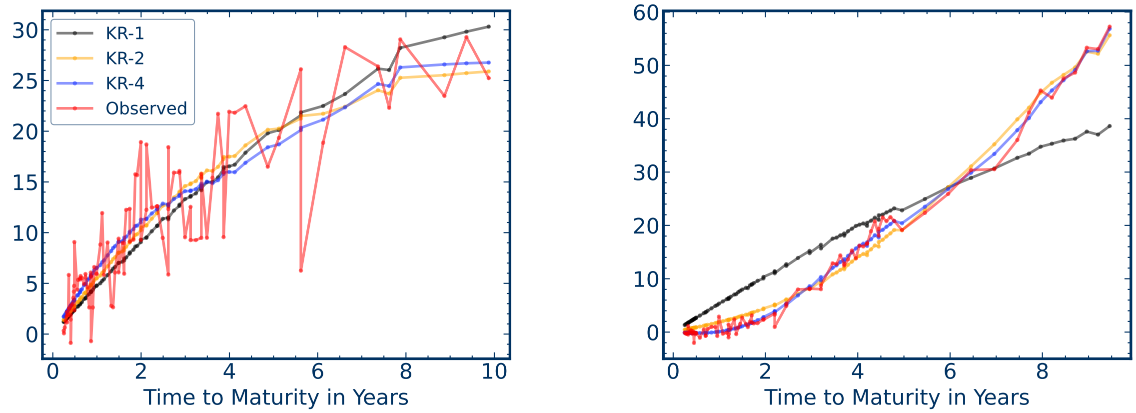
This figure shows the 3-month moving average of the two market condition measures. Subfigure (a) displays the Idiosyncratic Treasury Volatility measure (IT-VOL), which is the unexplained variation by the 4-factor KR model. Subfigure (b) presents the Treasury Market Complexity measure (T-COM), which is 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.

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 discount bond excess 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 16. This means that IT-VOL is essentially 1 minus the cross-sectional R^2 of a four factor model. T-COM is the difference between the explained variation with KR-1 and KR-4 factors, which corresponds to the difference between the black and light blue line in Figure 16,

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

Figure 17 plots the 3-month 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

Figure 18: 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.

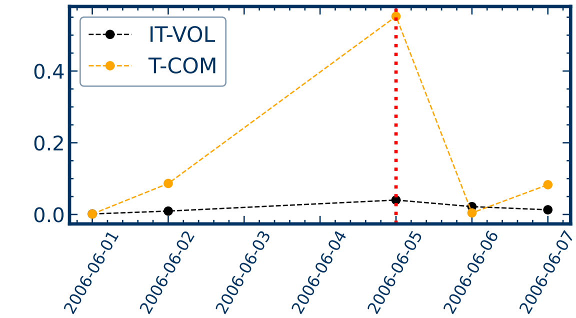
has more curvature and an estimate based on a simple slope factor would be strongly misspecified.

The examples in Figures 18 and 19 illustrate the meaning of the two measures. In Figure 18 we plot the observed and estimated coupon bond 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.

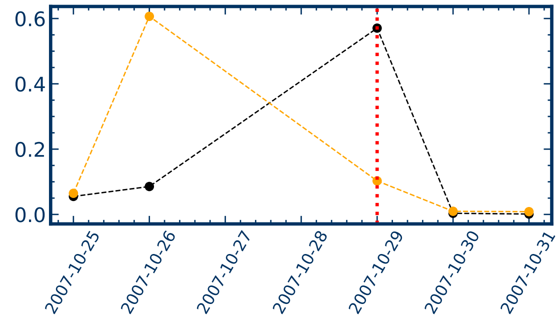
Our novel measures also detect high-frequency changes in bond markets. While Figure 17 shows the average time-series behavior of the market condition measures, Figure 19 illustrates the two measures and the fitted coupon bond excess returns for a time window around example days. In Panel (a), T-COM peaks 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 market that can be used for high frequencies.

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 IT-VOL and T-COM. The risk premium of the first (slope)

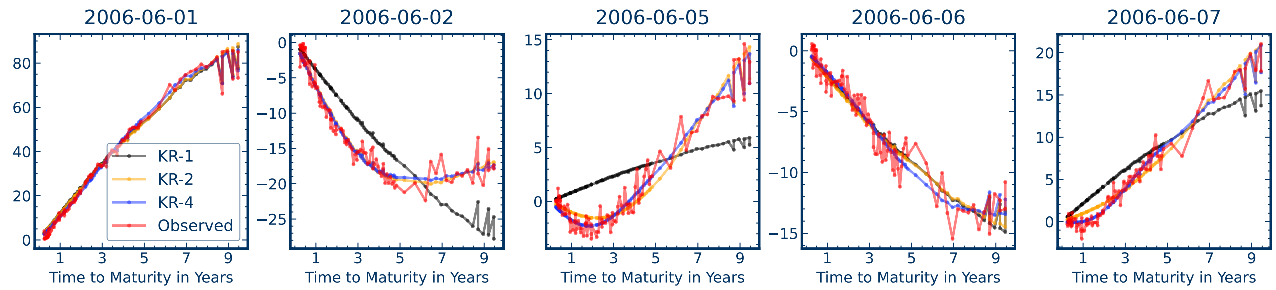
Figure 19: The Idiosyncratic Treasury Volatility (IT-VOL) measure and the Treasury Market Complexity (T-COM) measure on example dates over time



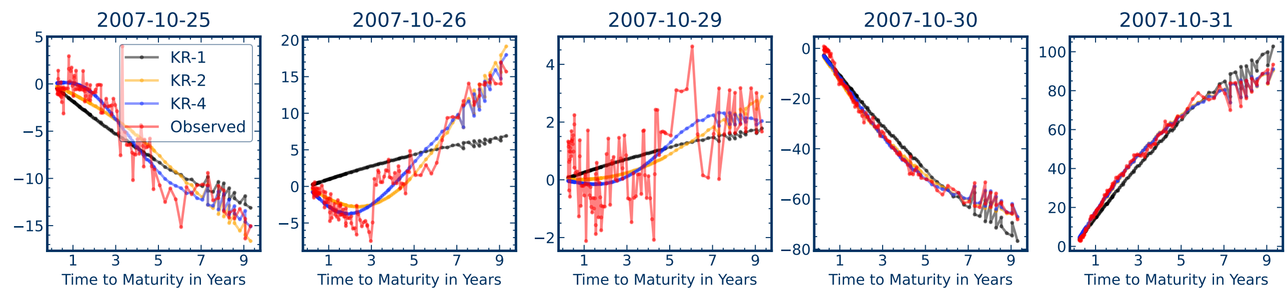
(a) IT-VOL and T-COM centered on 2006-06-05



(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.

factor is strongly affected by IT-VOL and T-COM. 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 IT-VOL and T-COM periods, but earns a high positive risk premium during challenging times. As a result, the implied SDF based on all four factors is largely unaffected, and the overall Sharpe ratio has

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

	SR					Mean				Standard Deviation			
	F_1	F_2	F_3	F_4	SDF	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	0.93	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	1.04	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	0.94	-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	1.14	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	0.75	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	0.95	-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.

only a minor dependence on the market conditions.

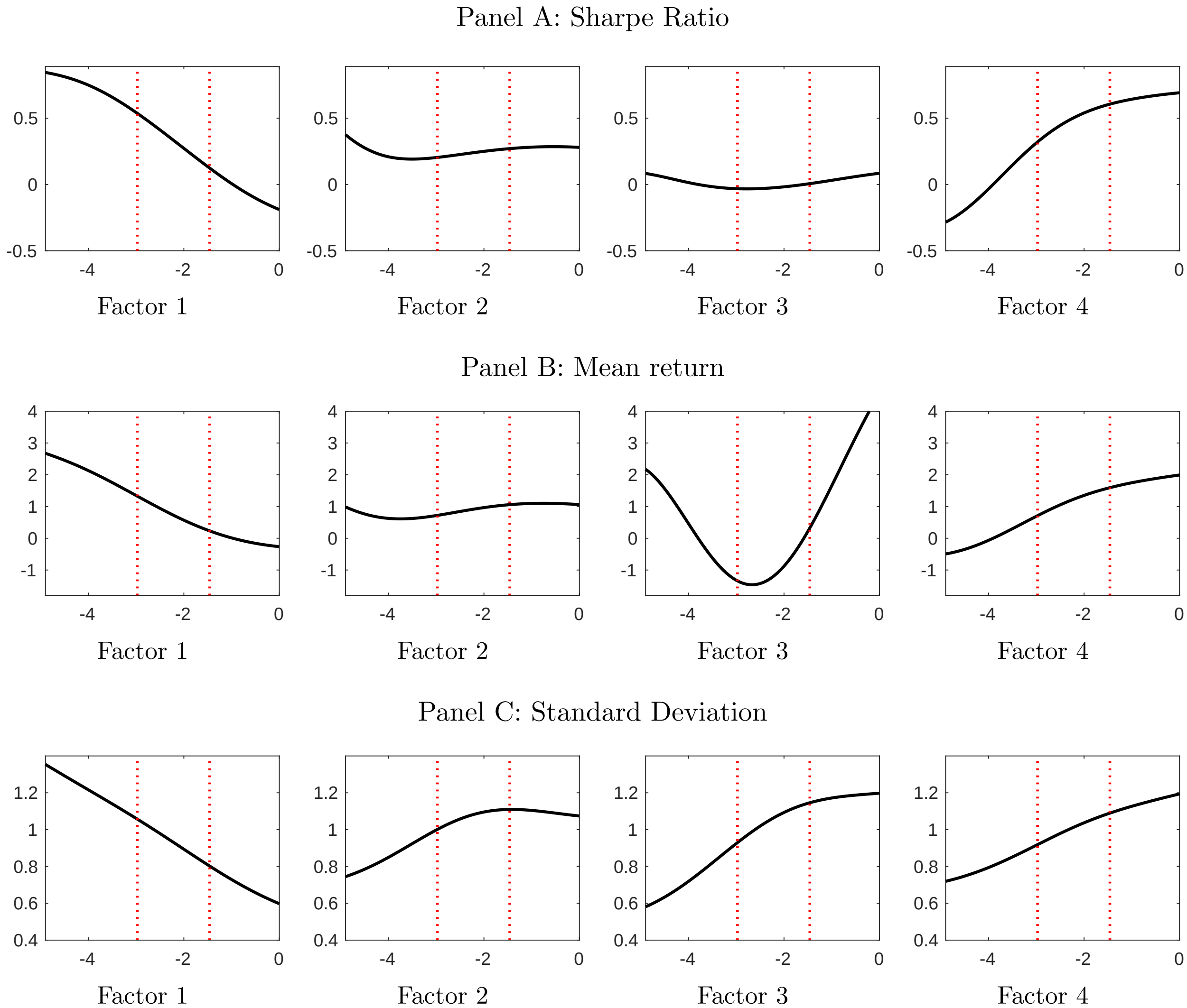
The first and fourth KR factors exhibit the largest conditional and unconditional Sharpe ratios. The second and fourth factors carry only a small risk premium as reflected by the lower Sharpe ratios. We emphasize that the dependency of the mean and Sharpe ratio of the KR factors on T-COM and IT-VOL is not mechanical and is evidence that these measures capture a source of risk in the bond market. The standard deviation of the first factor is decreasing in these risk measures, while the second to fourth factors exhibit an increasing variation. This makes sense, as for given loadings, factors with higher variance should explain more variation.

Figures 20 and 21 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.[11] 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.

The results are robust to alternative formulations of T-COM and IT-VOL. Table A.4 and Figures A.9 and A.10 in the Appendix 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

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

Figure 20: Sharpe ratio, mean, and standard deviation of KR factors conditioned on IT-VOL

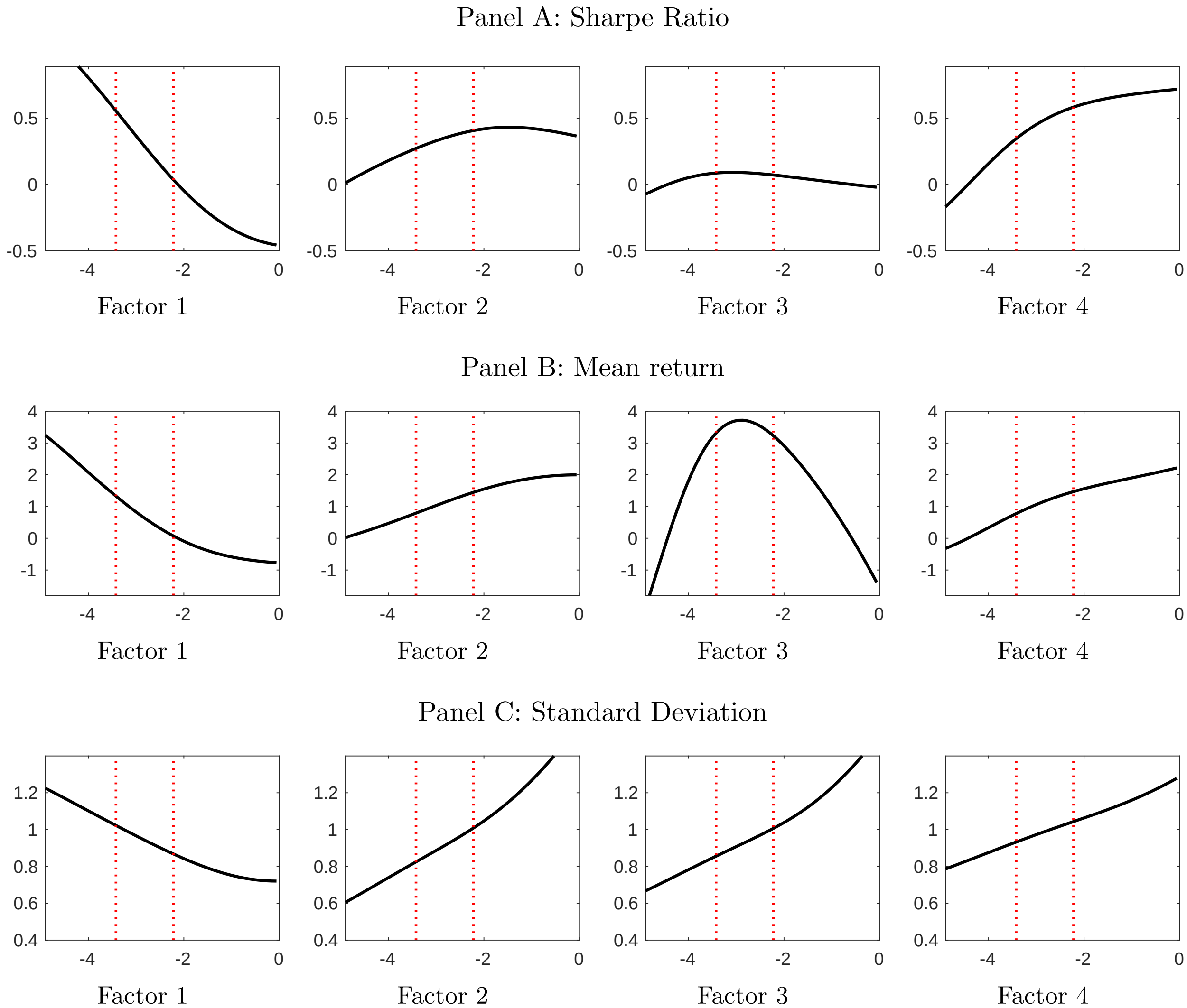


This figure shows the annualized Sharpe ratio (Panel A), normalized mean (Panel B), and normalized standard deviation (Panel C) of the four KR factors conditioned on the Idiosyncratic Treasury Volatility (IT-VOL) measure. The mean returns and standard deviation are normalized by their unconditional values. 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.

results are virtually identical, showing that our findings depend not on the choice of the number of factors.

Our novel results complement prior findings on the dependency of the term structure premium on 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 term structure curves during recessions, but at a higher frequency. Importantly, we find that the inversion of the slope can be at least partly hedged by the convexity effects of higher

Figure 21: 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.

order risk factors. This opens up a new research perspective and suggests that it is crucial to consider all relevant term structure factors when studying the impact of market conditions on the term structure curve.

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 discount bond excess return curve directly from observed returns of treasury securities. This curve lies in a reproducing kernel Hilbert space, which is derived from 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 discount bond excess 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. Treasuries, we show that the discount bond excess 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 available for any single day without the need for panel data, and we can explain why they occur. Our latent factors correspond to the optimal sparse set of non-parametric basis functions 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.

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. Treasuries. 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.

Our work makes important conceptional contributions. First, 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 connect these two problems. Second, 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, treasuries and equity can be studied with similar tools and interpreted in similar ways after identifying the analogy between cash flows and firm characteristics. Last but not least, we show the importance and how to use shrinkage in the term structure space. The smoothness of the discount curve, which follows from limits to arbitrage, naturally implies ridge and lasso shrinkage. Our novel methodology shows how a successful machine learning solution needs to be tailored to an economic problem.

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.

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

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., 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.

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 Reproducing Kernel Hilbert Space $\mathcal{H}_\alpha$

We first recap a the general definition of a reproducing kernel Hilbert space. This 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). 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 $x_1, \dots, x_n \in E$ the $n \times n$ matrix with elements $k(x_i, x_j)$ is symmetric and positive semi-definite. Thanks to the powerful property called the representer theorem, many kernel-based machine learning problems boil down to finite-dimensional standard convex optimization. Our results are a variant thereof.

Our hypothesis space $\mathcal{H}_\alpha$ is the reproducing kernel Hilbert space endowed with norm (5) and containing all twice continuously differentiable functions $h : [0, \infty) \to \mathbb{R}$ with $h(0) = 0$ and $\lim_{x\to\infty} h'(x) = 0$. For a full detailed technical discussion we refer to Filipović, Pelger, and Ye (2022), where we also derive a Bayesian distribution theory, which directly carries over to this setup and provides valid confidence intervals for our curve estimators.

A.2 Proof of Theorem 1

Now we show that the function r_t given in (4) is an element in $\mathcal{H}_\alpha$. First, economic first principles imply that r_t is twice continuously differentiable. Second, by definition, we have $r_t(0) = 0$. Third, it remains to show that $\lim_{x\to\infty} r_t'(x) = 0$. For this, we denote by f_t the forward curve such that $d_t(x) = \mathrm{e}^{-\int_0^x f_t(s)\,ds}$. We assume that

(i) $f_t : [0,\infty) \to \mathbb{R}$ is continuous and the limit $f_t(\infty) = \lim_{x\to\infty} f_t(x)$ exists in $\mathbb{R}$,

(ii) the difference function $\phi_t(x) = f_t(x) - f_t(\infty)$ is integrable such that $\int_0^\infty |\phi_t(x)|\,dx < \infty$.

Accordingly, we represent the discount bond excess return curve as

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

The absence of arbitrage implies that long forward rates can never fall, according to Dybvig, Ingersoll, and Ross (1996), that is, $f_{t-1}(\infty) \le f_t(\infty)$. Hence, we see that $\lim_{x\to\infty} r_t(x)$ exists in $\mathbb{R}$. Moreover, differentiating gives

$$\begin{aligned} r_t'(x) &= \frac{1}{d_{t-1}(\Delta_t)}\mathrm{e}^{\int_0^x (\phi_{t-1}(s+\Delta_t) - \phi_t(s))\,ds + (f_{t-1}(\infty) - f_t(\infty))x} \\ &\quad \times \left(\underbrace{\phi_{t-1}(x+\Delta_t) - \phi_t(x)}_{\to 0} + \underbrace{f_{t-1}(\infty) - f_t(\infty)}_{\le 0}\right) \end{aligned}$$

Again, by the absence of arbitrage, we see that $\lim_{x\to\infty} r_t'(x) = 0$, as desired. This completes the proof of Theorem 1.

A.3 Proof that the Reproducing Kernel of $\mathcal{H}_\alpha$ Is Strictly Positive Definite

This property is a consequence of the following lemma.

Lemma 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. Assume by contradiction that there exists a linear combination $h = \sum_{j=1}^N c_j k(\cdot, x_j) = 0$ 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 absurd. Hence the first claim follows. The second part follows, because $\boldsymbol{K}$ is the Gram matrix of the linearly independent kernel functions. □

A.4 Proof that the Functions in (7) Are Orthonormal

By the reproducing kernel property of k, we obtain $\langle u_i, u_j\rangle_{\mathcal{H}_\alpha} = \frac{1}{\sqrt{S_i}\sqrt{S_j}} \sum_{s,t=1}^N V_{si} k(x_s, x_t) V_{tj} = 1$ if $i = j$ and zero otherwise, as desired.

A.5 Proof of Theorem 2

(Filipović, Pelger, and Ye, 2022, Theorem A.1) show that the general solution of (6) is of the form $\hat{r}_t(\cdot) = k(\cdot, \boldsymbol{x})^\top \tilde{F}_t$ for the coefficients $\tilde{F}_t = Z_{t-t}^\top \left(Z_{t-t}\boldsymbol{K} Z_{t-1}^\top + \lambda M_t I_{M_t}\right)^{-1} R_t^{\text{bond}}$. Note that we have $\hat{F}_t = S^{1/2} V^\top \tilde{F}_t$, for $\hat{F}_t$ given in (9)–(11). By definition (7), we have $u(\cdot) = (u_1(\cdot), \ldots, u_N(\cdot))^\top = S^{-1/2} V^\top k(\cdot, \boldsymbol{x})$, 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 proves the first part of the theorem. Equation (12) now follows because $(u_1(\boldsymbol{x}), \ldots, u_N(\boldsymbol{x})) = \boldsymbol{K} V S^{-1/2} = V S^{1/2} = \beta$.

A.6 Proof of Theorem 3

The first part of the theorem follows from (8), noting that the functions u_j are orthonormal in $\mathcal{H}_\alpha$ and hence $\|\hat{r}_t\|_{\mathcal{H}_\alpha} = \|\hat{F}_t\|_2$. Expression (14) follows from (10)–(11) and the definition of $\beta_{t-1}^{\text{bond}}$.

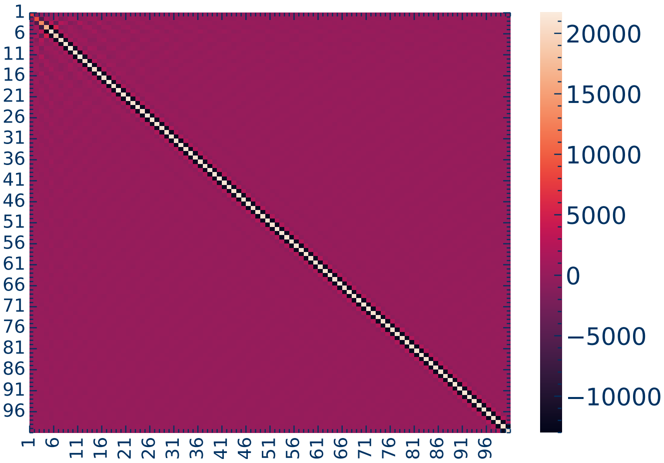
A.7 Intrinsic Relationship between PCA and KR Factors

In view of (11) and (12), the discount bond excess return vector R_t and the KR factors F_t are generically related such that $\operatorname{Var}(R_t) = V S^{1/2} \operatorname{Var}(F_t) S^{1/2} V^\top$. Therefore, $S^{1/2} F_t$ are the PCA factors, and V are the normalized PCA loadings, of R_t if and only if the factors F_t are uncorrelated, that is, $\operatorname{Var}(F_t)$ is diagonal. In this case, the order of the PCA factors matches those of F_t if the scaled variances $S_j \operatorname{Var}(F_{t,j})$ are descending in j. We show empirically that both conditions are approximately satisfied.

A.8 Alternative Factors

One could also consider the rotated factors $G_t = V F_t = \boldsymbol{K}^{-1/2} R_t$, for the positive definite inverse square root of the kernel matrix $\boldsymbol{K}^{-1/2} = V S^{-1/2} V^\top$. Figure A.1 shows that $\mathbf{K}^{-1/2}$ behaves approximately like a second-order difference operator. Thus, the rotated factors G_t are a proxy for the second order differences of the discount bond return vector. This provides another heuristic illustration for why the penalty term $\|r_t\|_{\mathcal{H}_\alpha}$ in the objective function (6) can be replaced by the ℓ_2-norm of the factors $\|F_t\|_2 = \|G_t\|_2$ in the equivalent problem (13). We prefer F_t over G_t, because they behave approximately like the PCA factors of the discount bond excess return curve.

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.

A.9 Proof of Expression (16)

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 (15) splits into $\Phi(F_t) = \sum_{j=1}^N \Phi_j(F_{t,j})$, with components

$$\Phi_j(F_{t,j}) = \frac{1}{N}(v_j^\top R_t^{\text{bond}} - S_j^{1/2}F_{t,j})^2 + \lambda F_{t,j}^2 + \lambda_1|F_{t,j}|.$$

The function $\Phi_j(F_{t,j})$ is differentiable in all $F_{t,j} \neq 0$, with derivative

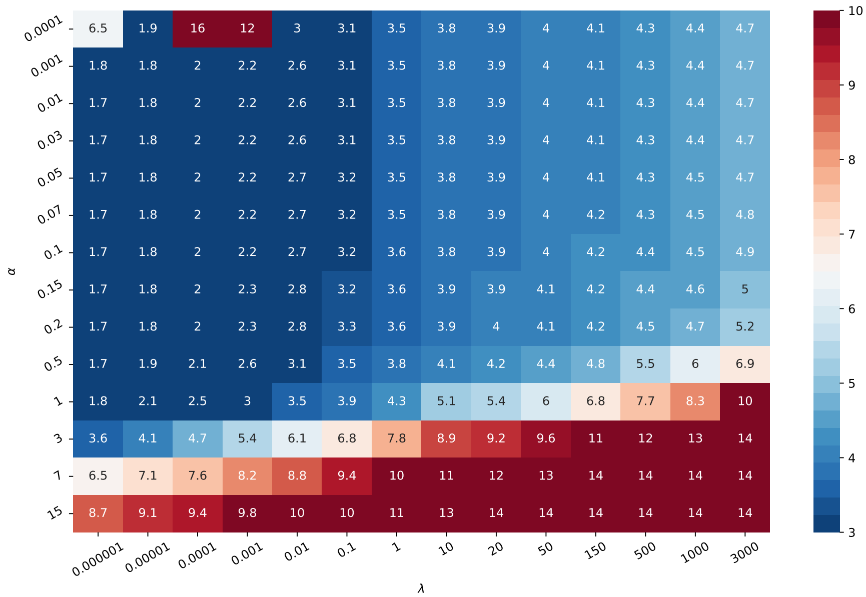
$$\Phi_j'(F_{t,j}) = -\frac{2}{N}S_j^{1/2}v_j^\top R_t^{\text{bond}} + \frac{2}{N}S_jF_{t,j} + 2\lambda F_{t,j} + \lambda_1\text{sign}(F_{t,j}).$$

It follows that $F_{t,j} = 0$ is optimal, all else equal, if and only if $-\lambda_1\frac{N}{2} \leq S_j^{1/2}v_j^\top R_t^{\text{bond}} \leq \lambda_1\frac{N}{2}$, and thus (16). This completes the proof.

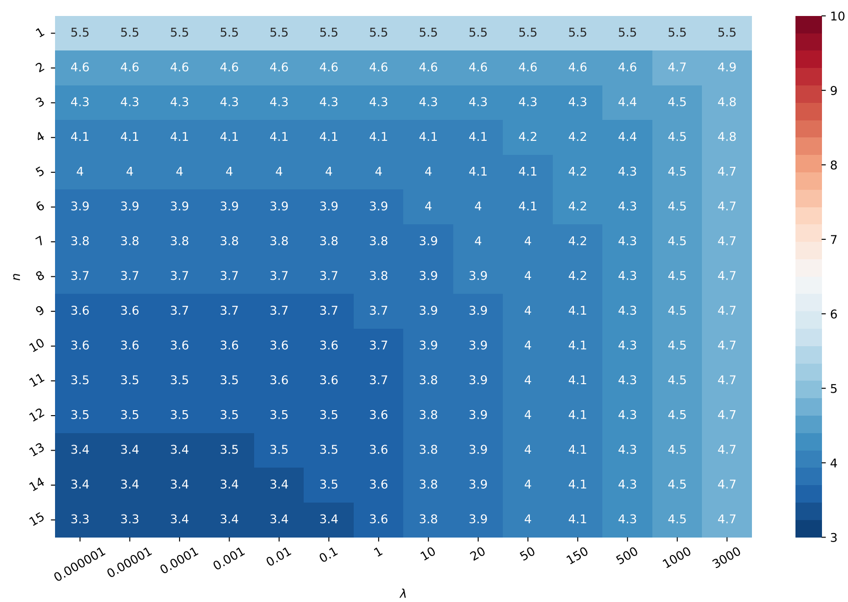
B Additional Empirical Results

B.1 Term Structure Estimation

Figure A.2: 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.

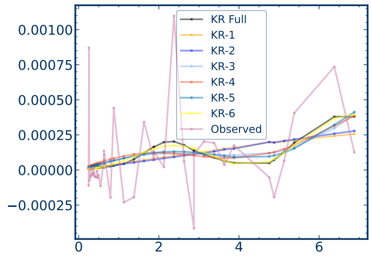
Table A.1: Excess return fitting RMSE by maturity bucket

	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	3.9	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

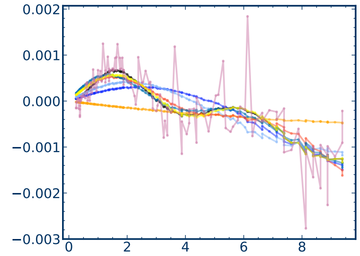
This table shows the average RMSE in fitting excess returns of observed securities. RMSE is calculated separately on each date and then averaged over time. The first column displays RMSE aggregated over all maturities. The rest of the columns show RMSE broken up by the six maturity buckets. GSW discount 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. One-period risk-free rate by KR is used. Results are calculated using daily data from June 1961 to December 2020. Numbers are in basis points (BPS).

Figure A.3: Illustration of KR estimator

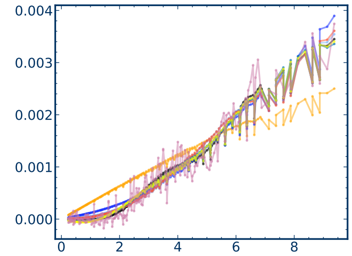
Panel A: Observed and Fitted Excess Returns of Securities



(a) 1961-06-30

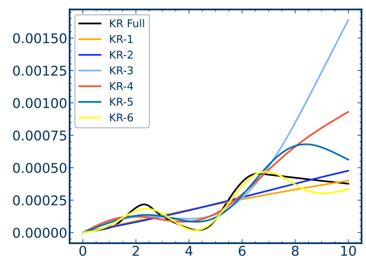


(b) 1986-06-30

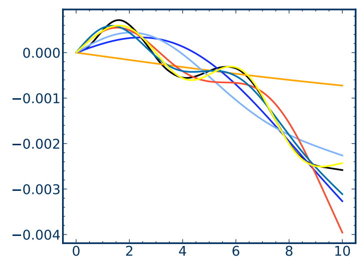


(c) 2013-12-30

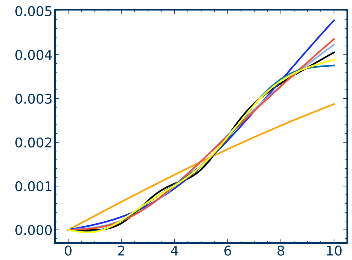
Panel B: Fitted Excess Returns of Discount Bonds



(d) 1961-06-30

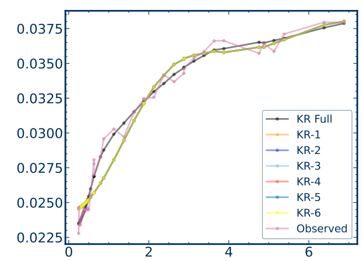


(e) 1986-06-30

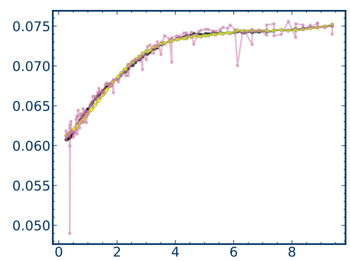


(f) 2013-12-30

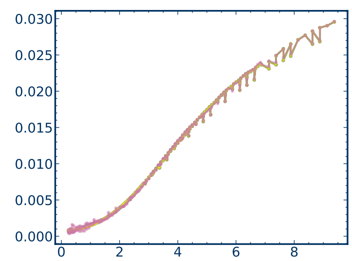
Panel C: Observed and Fitted YTM of Securities



(g) 1961-06-30

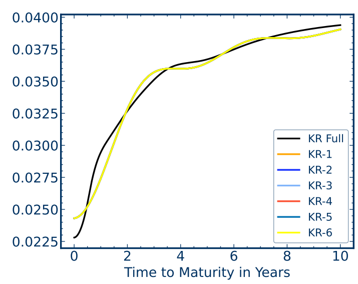


(h) 1986-06-30

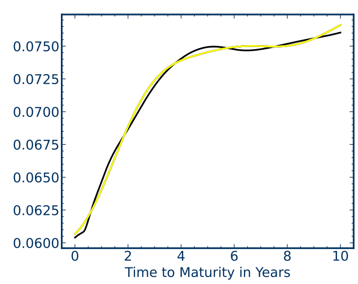


(i) 2013-12-30

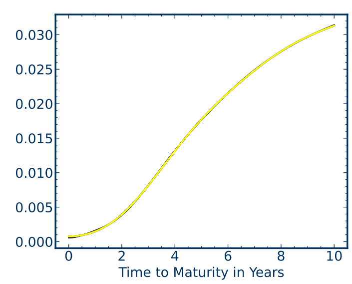
Panel D: Fitted Yield Curves



(j) 1961-06-30



(k) 1986-06-30

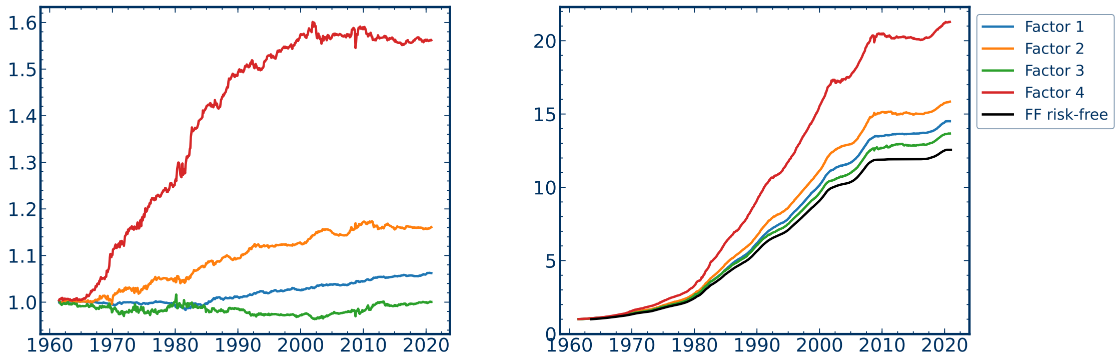


(l) 2013-12-30

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.

B.2 Term Structure Premium

Figure A.4: 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 securities. The risk-free overnight return is based on the Fama–French monthly risk-free rate. The sampling period is from June 1961 to December 2017.

Table A.2: Sharpe Ratios of Factors and Implied SDF with FF Risk-Free Rate

	KR Factors			PCA Factors		
n	Factor	SDF	SDF (OOS)	Factor	SDF	SDF (OOS)
1	0.51	0.51	0.54	0.37	0.37	0.43
2	0.53	0.77	0.78	0.33	0.50	0.55
3	0.37	0.89	0.88	0.24	0.55	0.54
4	0.46	1.03	1.02	0.45	0.71	0.75
5	0.34	1.10	1.05	0.32	0.78	0.77
6	0.32	1.18	1.10	0.55	0.95	0.90

This table shows the annualized Sharpe ratios for different KR and PCA 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 or PCA 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 PCA factors are estimated on a panel estimated with the full KR model. The risk-free overnight return is based on the Fama–French monthly risk-free rate. The sample is daily data from June 1961 to December 2020.

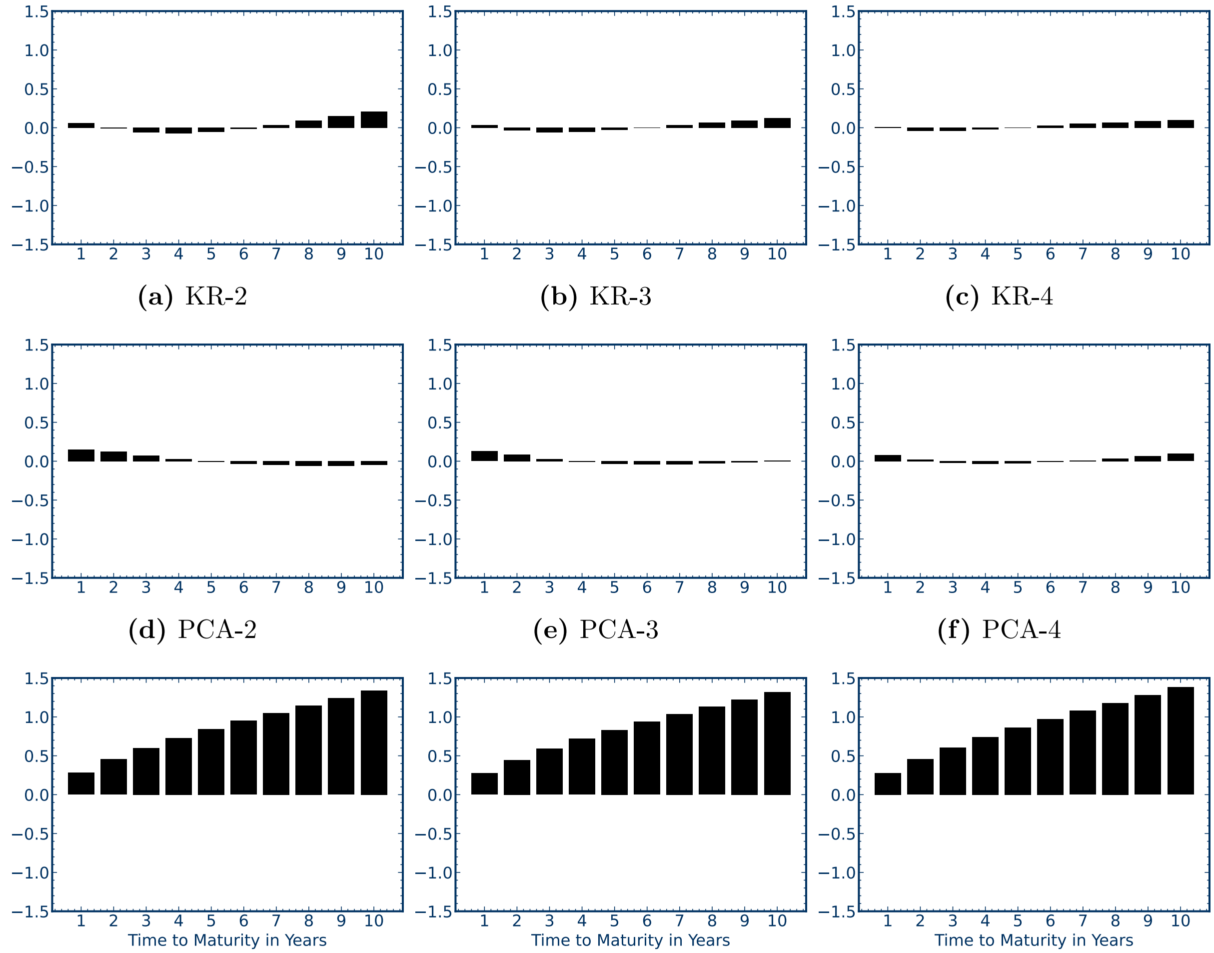
B.3 Term Structure Asset Pricing

Table A.3: Cross-Sectional Asset Pricing Results

	1Y	2Y	3Y	4Y	5Y	6Y	7Y	8Y	9Y	10Y	RMS$_\alpha$
					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.5: Pricing errors of GSW discount bond excess returns for different factors



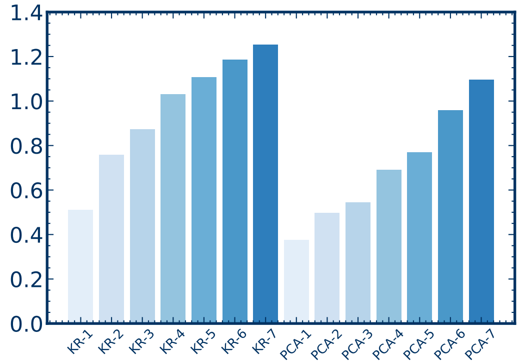
(a) KR-2 **(b)** KR-3 **(c)** KR-4

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

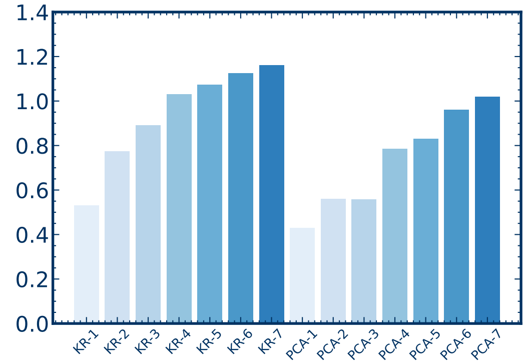
(g) Mean Excess Returns **(h)** FF-3 **(i)** FF-5

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 overnight return is the estimate from the KR discount curve. The sampling period is from July 1963 to December 2020.

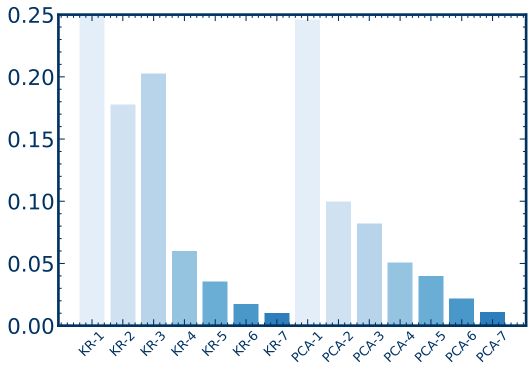
Figure A.6: Asset Pricing Results for FF one-day risk-free rate



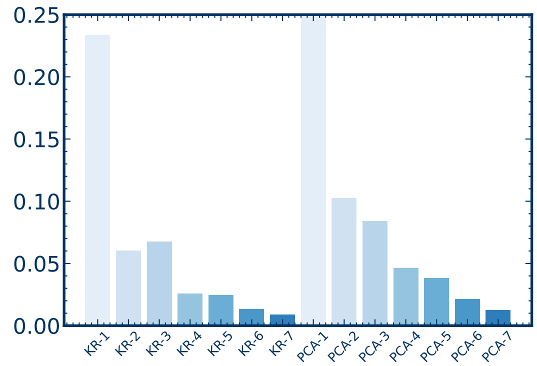
(a) SR (Full-Sample)



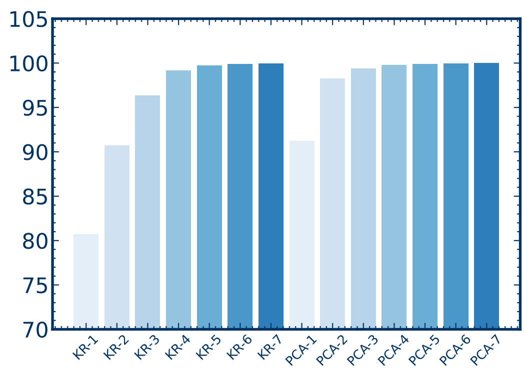
(b) SR (Out-of-Sample)



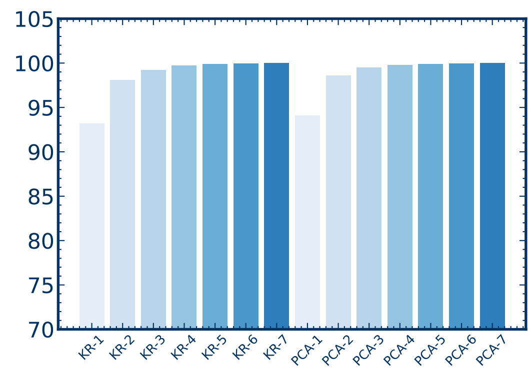
(c) RMS$_\alpha$ (Full-Sample)



(d) RMS$_\alpha$ (Out-of-Sample)



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

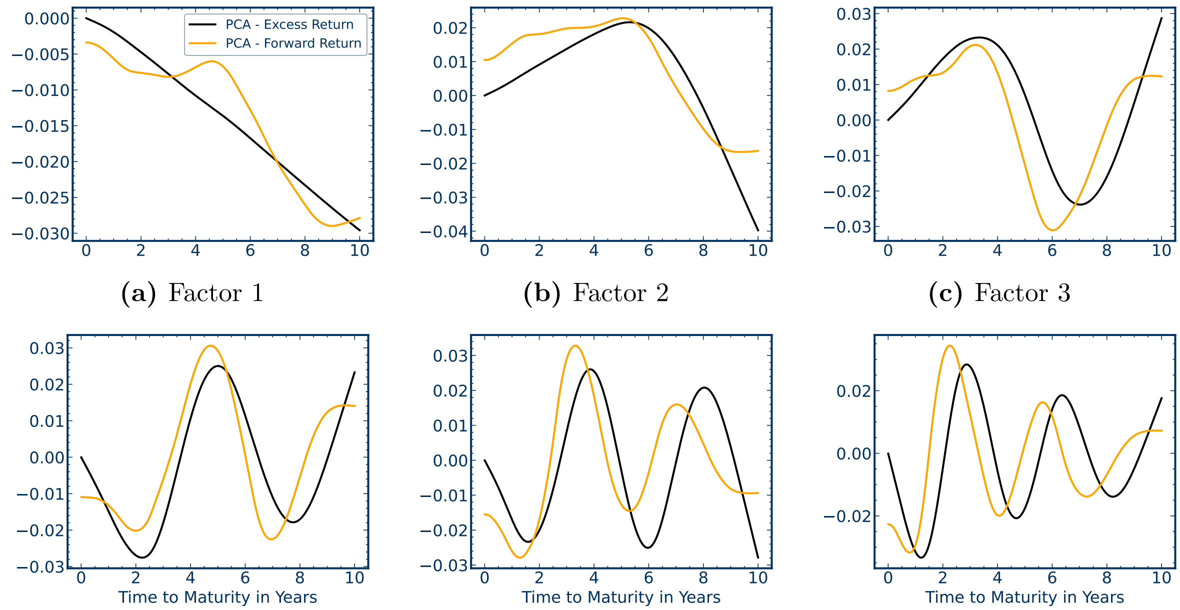


(f) Explained Variation (%, Out-of-Sample)

This figure summarizes the asset pricing results for different terms structure factors on a panel of discount bond excess returns. The risk-free overnight return is based on the Fama–French monthly risk-free rate. 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 panel of 10 discount bond excess returns estimated with KR with yearly maturities ranging from 1 to 10 years. The annualized Sharpe ratios of the implied SDF are based on the mean-variance efficient combination of successively adding factors. The 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 overnight return is the estimate from the KR discount curve. The sampling period is from July 1963 to December 2020.

B.4 Number of Term Structure Factors

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



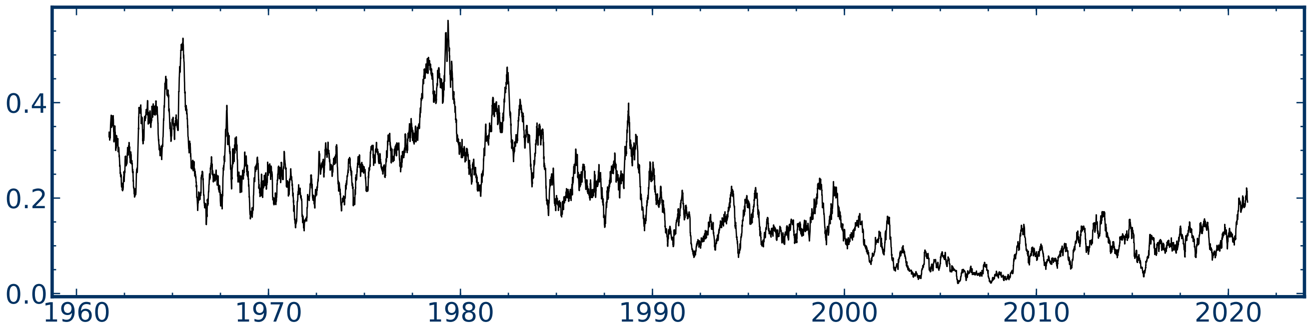
(a) Factor 1 **(b)** Factor 2 **(c)** Factor 3

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

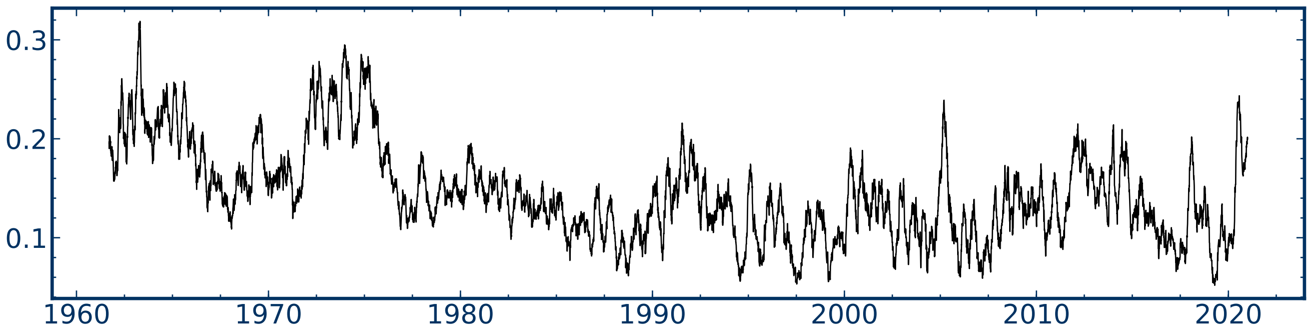
This figure shows the loadings of the first 6 PCA factors estimated on a panel of discount bond excess returns and and a panel of forward returns. The loadings correspond to portfolio weights of the PCA factors, either in terms of discount bond excess returns or in terms of forward returns. The loadings 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.

B.5 Market Condition Measures

Figure A.8: 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 3-month 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.

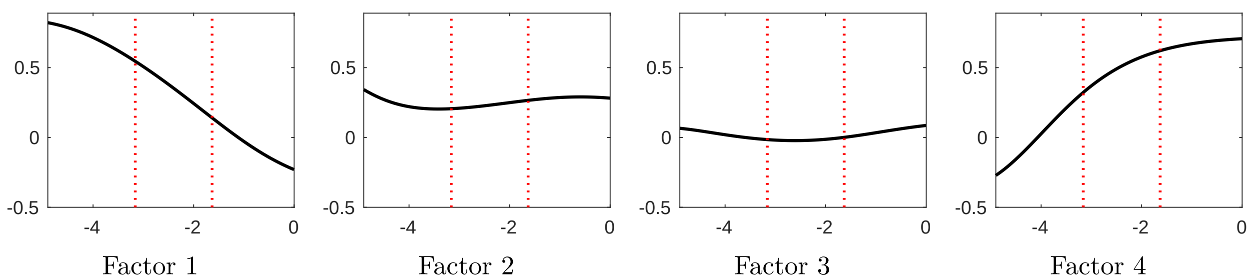
Table A.4: Sharpe ratios, means, and standard deviation of KR factors conditioned on IT-VOL and T-COM based on full KR model

	SR					Mean (BPS)				Standard Deviation (BPS)			
	F_1	F_2	F_3	F_4	SDF	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	0.85	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	1.09	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	0.98	-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	1.2	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	0.94	-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	0.92	-0.37	2.99	0.56	6.85	15.63	128.58	216.06	209.06

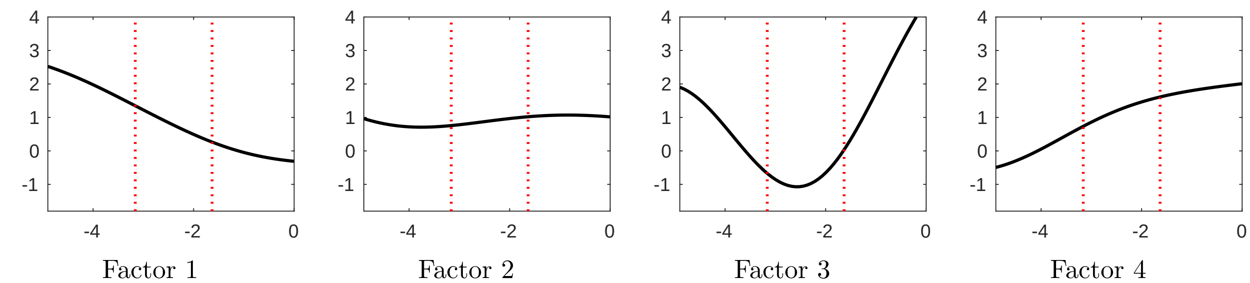
This table shows the annualized Sharpe ratio, mean, and standard deviation of the four KR factors conditioned on terciles of the Idiosyncratic Treasury Volatility (IT-VOL) measure and the Treasury Market Complexity (T-COM) measure. The IT-VOL and T-COM measures are constructed with the full KR model, 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 A.9: Sharpe ratio, mean, and standard deviation of KR factors conditioned on IT-VOL based on full KR model

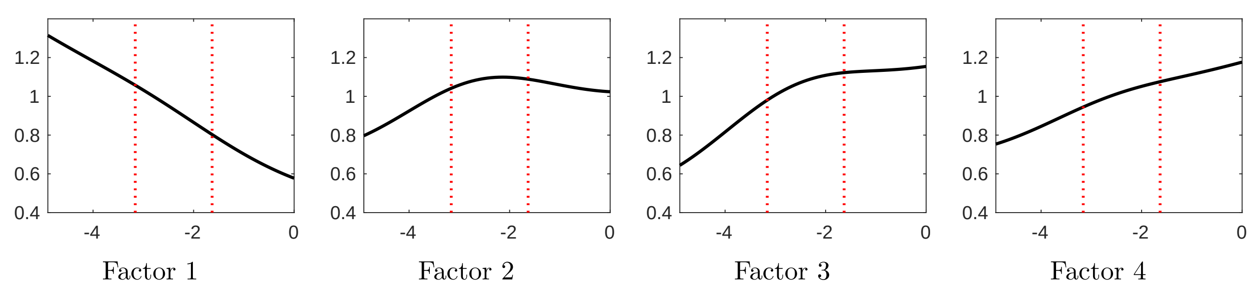
Panel A: Sharpe Ratio



Panel B: Mean return

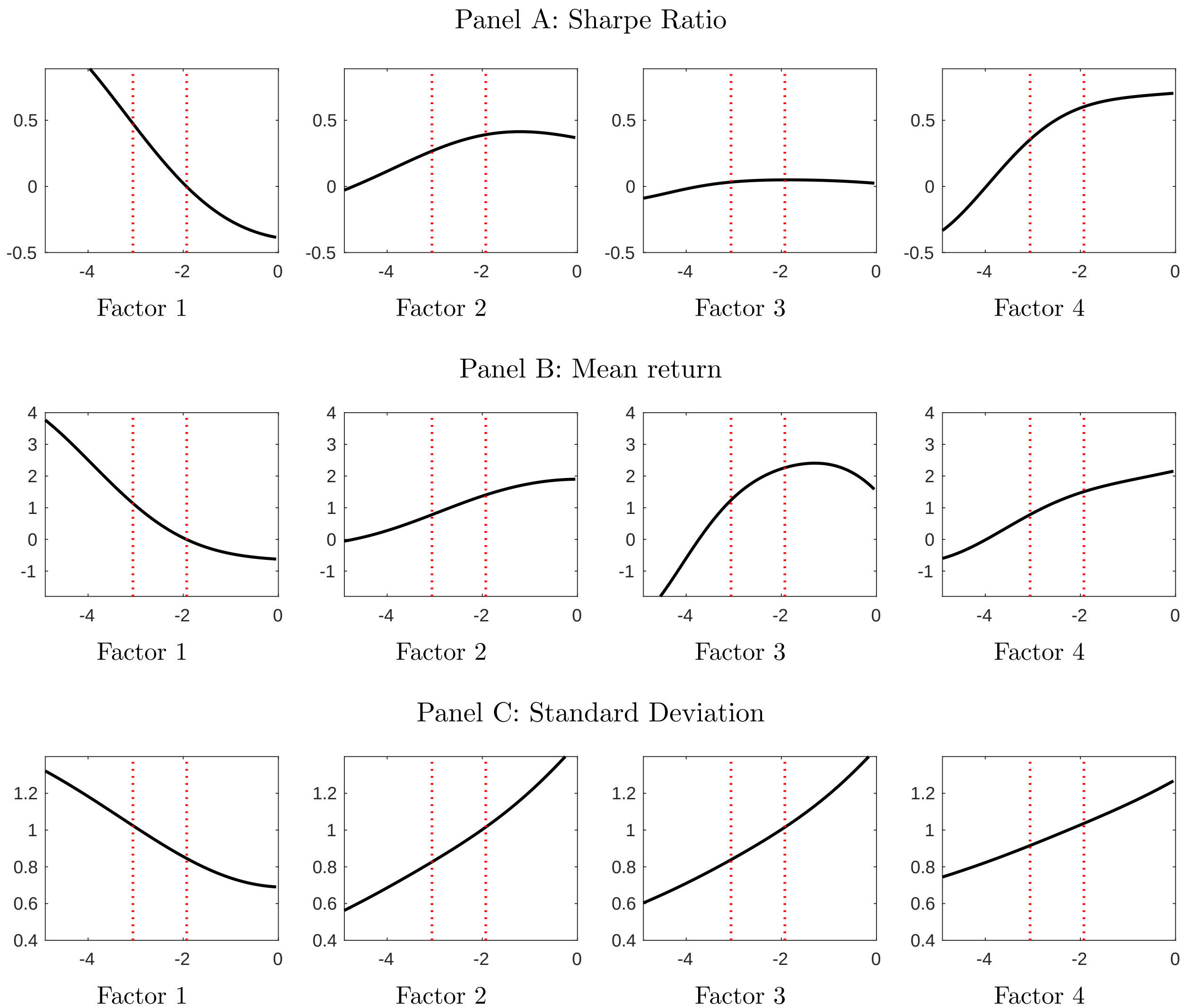


Panel C: Standard Deviation



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 A.10: 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.

Swiss Finance Institute

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

swiss:finance:institute

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