On the Factor Structure of Bond Returns*

Richard K. Crump† Nikolay Gospodinov‡

May 31, 2021

Abstract

We demonstrate that characterizing the minimal dimension of the term structure of interest rates is more challenging than currently appreciated. The highly structured polynomial patterns of the factor loadings, that are widely reported and discussed in the literature, reflect local correlations of smooth curves across maturities. We derive analytical expressions for the loadings of cross-sectionally dependent processes that tend to favor a much lower dimension than the true dimension of the underlying factor space. Numerical examples illustrate the significant economic costs of erroneously committing to a parsimoniously parameterized factor space, informed by standard metrics of goodness-of-fit. Our results apply to other assets with a finite maturity structure.

Keywords: Term structure of interest rates, bond returns, principal components, eigenvectors, factor space, local correlation

JEL Classification: C38, C58, G12

*The authors would like to thank the Co-Editor, four anonymous referees, Tobias Adrian, Anna Cieslak (discussant), Mark Fisher, Ken Garbade, Domenico Giannone, Carsten Jentsch (discussant), David Lucca, Stijn Van Nieuwerburgh, José Luis Montiel Olea, Peter Van Tassel, Paula Tkac, Erik Vogt, Desi Volker, Mark Watson, Jonathan Wright, and seminar participants at the 9th Bundesbank Term Structure Workshop, Columbia Business School, Columbia University, the Federal Reserve Bank of Atlanta, the Federal Reserve Bank of New York, the 2019 LBS Summer Finance Symposium, McGill University, the 2019 NBER Summer Institute, and Princeton University for helpful discussions, comments and suggestions. Some results in this paper were previously circulated under the title "Deconstructing the Yield Curve." Oliver Kim and Hunter Wieman provided excellent research assistance. The views expressed in this paper are those of the authors and do not necessarily represent those of the Federal Reserve Bank of Atlanta, Federal Reserve Bank of New York or the Federal Reserve System.

†Federal Reserve Bank of New York.

‡Federal Reserve Bank of Atlanta.

1 Introduction

Risk factors are the fundamental building block of modern asset pricing models. In some asset classes such as equities, there are ongoing debates about how many and what risk factors are present. In contrast, the debate in the term structure literature appears largely settled in favor of the first three principal components of bond yields or returns – commonly referred to as level, slope and curvature (see Piazzesi, 2010; Gürkaynak and Wright, 2012; and Duffee, 2013, for recent surveys). In this paper, we argue that characterizing the true factor space of the term structure of interest rates is much more challenging than is widely believed. Common indicators of goodness of fit, such as explained variation, or the interpretability of the factor loadings are less informative than currently appreciated. Instead, these quantities merely signify the strong local correlation across nearby maturities observed in the term structure.

The dominant approach in the term structure literature is to use principal component analysis (PCA) to estimate factors from bond yields. However, working with highly persistent series, such as bond yields, may obscure the true factor structure and produce spurious common variation (Onatski and Wang, 2021). Instead, we follow the seminal work on common factors in the yield curve by Litterman and Scheinkman (1991) and Garbade (1996) that recommends the use of bond returns which are only weakly serially correlated. This allows us to isolate our analysis from the influence of time-series persistence and instead exclusively focus on the role of cross-sectional dependence.

Even after removing the strong time-series persistence, we show that specific properties of returns across maturities obfuscate the true minimal dimension of the underlying data generating process. The source of this problem is the strong cross-sectional dependence across maturities that produces a polynomial pattern in the eigenvectors of the covariance matrix of returns. This cross-sectional dependence plays an instrumental role in generating these patterns, regardless of the underlying true factor space. We also show that definitional overlapping relationships in the term structure can further exacerbate this problem. We characterize analytically, in a stylized setting, the properties of principal component analyses applied to non-factor models of the term structure which feature strong local dependence. Although a factor model of excess returns can be a successful dimension reduction technique, we show that a seemingly small mischaracterization of the factor space can result in meaningful economic costs. We demonstrate through two numerical examples that high explained variation can coexist with substantial deviations from the optimal portfolio allocation and hedging efficiency.

Along with highlighting the difficulties in determining the minimal dimension of the term structure, our paper also makes two additional contributions. First, in our setting, the monotonically decreasing local correlations in the spatial dimension, induced by weighted partial sums across maturities, generate highly structured polynomial patterns in the eigenvectors. This appears to hold for any smooth curve or process and thus demonstrates clear parallels with the analysis of low-frequency movements in persistent time series (Müller and Watson, 2008, 2018). Second, and relatedly, under certain realistic parameterizations, the canonical tools for selecting factors and evaluating goodness-of-fit become inappropriate and even misleading. Thus, we provide cautionary evidence of the limitations of existing data-driven procedures for factor model selection.

Our results have important implications for yield curve modeling and analysis. First, our analysis suggests that primitive processes for analyzing the true factor space should be free from the mechanical overlapping structure which is embedded in the term structure of maturity-ordered assets. From this perspective, yields appear to be a poor choice of primitive objects as they are strongly serially and cross-sectionally correlated. Thus, taking a stand on the number of factors in yields – which is the typical approach in the literature (e.g., Giglio and Kelly, 2017) – proves to be a formidable task. Second, given the challenges in characterizing the true factor space even for bond returns, it may be desirable to appeal to inference methods that are agnostic about the underlying data generating process. For example, Crump and Gospodinov (2019) propose a nonparametric, model-free bootstrap procedure which is robust to unknown factor structure.

The rest of the paper is organized as follows. Section 2 introduces the notation and the empirical characteristics of bond returns. Section 3 presents the main theoretical results. Under different assumptions about the driving process, we derive analytical expressions for the eigenvalues and eigenvectors of the covariance matrix of returns. We also demonstrate the economic costs of wrongly attributing the common variation of bond returns only to their first three principal components. Section 4 provides further discussion and conclusions. Proofs of the main results are in Appendix A while additional results and empirical evidence are provided in Appendices B–D in the Supplementary Material.

2 Setup and Notation

Before we proceed to our main results, we introduce some notation and definitions for the primary variables of interest. Let $p_t^{(n)}$ denote the time t log price of a zero-coupon bond which pays \$1 at time $t+n$, where $t = 1, \ldots, T$ and $n = 1, ..., N$. The corresponding log yield is denoted by $y_t^{(n)}$ and satisfies $p_t^{(n)} = -ny_t^{(n)}$. The log forward rate for the period $t+n-1$ to $t+n$ is defined as

$$f_t^{(n)} \equiv p_t^{(n-1)} - p_t^{(n)}. \tag{1}$$

Since $p_t^{(0)} = 0$, the one-period yield, $y_t^{(1)}$, may equivalently be written as $f_t^{(1)}$. Using the recursive (in n) nature of equation (1) and the definition of yields, we have

$$p_t^{(n)} = -\sum_{i=1}^{n} f_t^{(i)}, \qquad y_t^{(n)} = \frac{1}{n}\sum_{i=1}^{n} f_t^{(i)}. \tag{2}$$

We observe that $p_t^{(n)}$ and $y_t^{(n)}$ are partial sums and partial averages of forward rates, respectively. These formulas demonstrate that two yields, $y_t^{(n)}$ and $y_t^{(m)}$, have $\min(m, n)$ forwards in common (and the same for prices). This overlap, which arises solely from these term-structure identities, implies differential behavior in the covariance or correlation matrix of forwards relative to yields (or prices); in fact, log prices will always exhibit stronger local dependence than forwards across the same maturities.[1]

[1] This is true provided that forwards are non-negatively correlated across maturities (as is the case in practice).

The one-period holding return on a bond of maturity n from time t to $t+1$ is defined as

$$r_{t,t+1}^{(n)} \equiv p_{t+1}^{(n-1)} - p_t^{(n)}. \tag{3}$$

The corresponding excess return (in excess of the risk-free rate, $y_t^{(1)}$) is then given by

$$rx_{t,t+1}^{(n)} \equiv r_{t,t+1}^{(n)} - y_t^{(1)} = r_{t,t+1}^{(n)} - r_{t,t+1}^{(1)}. \tag{4}$$

The notation $r_{t,t+1}^{(n)}$ and $rx_{t,t+1}^{(n)}$ signifies that these returns are earned from period t to $t+1$. In the sequel, we will simplify notation to $r_{t+1}^{(n)}$ and $rx_{t+1}^{(n)}$, respectively. Since returns are defined based on changes in log prices, they have two important properties: (i) the time differencing substantially reduces the degree of time-series persistence; (ii) the overlapping nature of prices relative to forward rates is directly inherited by returns.

Now define the return on the following long/short trading strategy – buying an n-maturity bond and shorting an $(n-1)$-maturity bond – as

$$dr_{t+1}^{(n)} \equiv \left(p_{t+1}^{(n-1)} - p_t^{(n)}\right) - \left(p_{t+1}^{(n-2)} - p_t^{(n-1)}\right) = r_{t+1}^{(n)} - r_{t+1}^{(n-1)} = rx_{t+1}^{(n)} - rx_{t+1}^{(n-1)}. \tag{5}$$

In light of the last equality in equation (5), we will refer to the return of this trading strategy as a difference return since we are differencing across the *maturity* dimension. Difference returns also have the interpretation of receiving the fixed leg in the forward rate (as in a forward interest rate swap) as they may be rewritten as

$$dr_{t+1}^{(n)} = \left(p_t^{(n-1)} - p_t^{(n)}\right) - \left(p_{t+1}^{(n-2)} - p_{t+1}^{(n-1)}\right) = f_t^{(n)} - f_{t+1}^{(n-1)}. \tag{6}$$

Then, it follows immediately from equation (5) that excess returns are partial sums (across maturities) of $dr_{t+1}^{(n)}$,

$$rx_{t+1}^{(n)} = \sum\nolimits_{i=2}^{n} dr_{t+1}^{(i)}. \tag{7}$$

It is instructive to contrast these definitional relationships with the current practice of estimating term structure factors by applying static PCA to yields across maturities (e.g., Duffee, 2013). It is well-known that yields are highly persistent in the time series dimension with near-unit root behavior. As demonstrated by Uhlig (2009) and Onatski and Wang (2021), principal component analysis of highly-persistent data may produce misleading results. For example, factorless, nonstationary data gives rise to a spurious inference of a small number of factors that absorb almost all of the data variation. To guard against this spurious factor problem, the data should be transformed to induce stationarity. In classical time series analysis, this is accomplished by taking first differences, $y_t^{(n)} - y_{t-1}^{(n)}$. However, in the term structure setting, we have a natural economic transformation which induces stationarity; namely, utilizing excess returns or difference returns. In fact, this was precisely the approach taken in the seminal work of the literature (Litterman and Scheinkman, 1991 and Garbade, 1996) on extracting principal components from bond data. In the

sequel, we intentionally isolate our analysis from the complexities arising from the time-series persistence by focusing on excess bond returns that are only weakly serially correlated. We show that cross-sectional dependence presents challenges in characterizing the true factor space, irrespective of the degree of time-series persistence in yields.

We start by visualizing the salient properties of bond return data that inform the analysis in this paper. Figure 1 plots excess returns (left) and difference returns (right) for maturities up to 5 years using data based on Eurodollar futures contracts at quarterly frequency and quarterly maturity intervals (see Appendix D in the Supplementary Material for a detailed description of the data). The top graphs demonstrate that both excess returns and difference returns exhibit limited time-series persistence. However, both series share a common property of disciplined co-movement across maturities similar to the appearance of sedimentary rock.

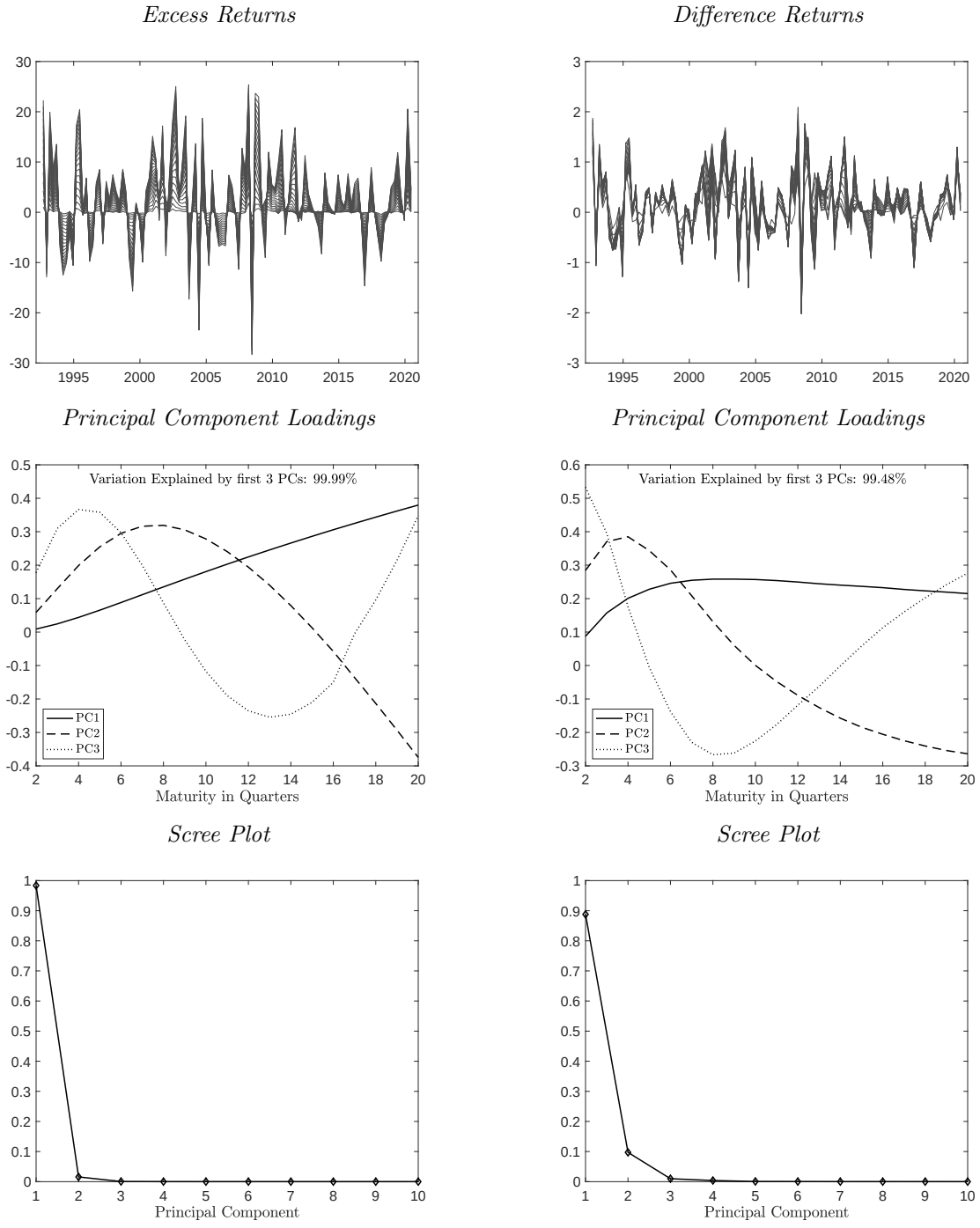
The middle left chart of Figure 1 presents the factor loadings for excess returns, based on the empirical covariance matrix, that were first documented in the term structure literature by Litterman and Scheinkman (1991) and gave rise to the nomenclature: level, slope and curvature. In this particular sample, these first three principal components explain almost 100% of the cross-sectional variation of returns with only the first principal component accounting for more than 98.3% of the variation (illustrated in the bottom left (scree) plot of Figure 1). Using the first three principal components as factors is in line with the standard approach in dynamic factor models which posits that a small number of factors drive the common variation in a large cross section of economic time series (for a comprehensive review, see Stock and Watson, 2016). In the next section, we make the argument that the highly structured principal component loadings of excess returns are guaranteed by the built-in definitional structure (cross-sectional partial sums) that links excess returns and difference returns (see equation (7)).

In the middle and bottom right charts of Figure 1, we show the first three principal component loadings and associated scree plot for difference returns. At first glance, difference returns also appear to be well approximated by a low-dimensional factor space with a possible kink in the scree plot occurring at the third component (with the first three components explaining in excess of 99.4% of the variation).[2] In the following, we will show that extracting principal components from difference returns also poses substantial challenges arising from local correlation in the forward curve.

Finally, it is important to note that the shape of the principal component loadings for excess returns will be sensitive to whether the covariance or correlation matrix is used for the eigenvalue decomposition. In contrast, the loadings for difference returns will be essentially unchanged. This is because the variability of excess returns is materially different across maturities in contrast to difference returns (which is evident from the graphs in the top row of Figure 1). In fact, the variance of excess returns is generally monotonically increasing with maturity.

[2]Appendix C in the Supplementary Material includes results on an additional data set using Treasury STRIPS prices, which displays more modest local dependence properties for *difference* returns. Despite this weaker correlation structure, we show that the principal component loadings and explained variation for *excess* returns are strikingly similar across the two data sets.

Figure 1. Properties of the Term Structure of Bond Returns. The top row of this figure shows the time series dynamics of excess bond returns (left column) and difference returns (right column) for maturities up to 20-quarters (in percent, annualized). The second and third rows show the principal component (PC) loadings and scree plot based on the corresponding empirical covariance matrix using data from Eurodollar futures. The sample period is 1992Q3–2020Q2.



3 Analytical Factor Loadings

To provide more concrete intuition for the underlying mechanisms behind the patterns documented in Figure 1, we characterize analytically the factor loadings under different assumptions about the

driving process for difference returns. The setup we maintain is purposefully simple; we abstract from such generalities as measurement error, conditional heteroskedasticity, yield curve smoothing and so on. The simplicity of the setup allows us to obtain analytical approximations to build intuition and ensure transparency of the results. More specifically, we show how local correlation across maturities gives rise to a polynomial structure in the principal component loadings – expressed as perturbations of sine functions. The familiar level, slope and curvature characterizations are the first three polynomials in this representation.

3.1 Role of Overlapping Maturities in the Cross-Section

Consider equation (7) where holding-period excess returns of an n-maturity bond are expressed as the sum of difference returns from maturity 2 to n. This implies that excess returns of maturities n_1 and n_2 will have $\min(n_1, n_2) - 1$ difference returns in common. The commonality of these difference returns then generates strong cross-sectional dependence of excess returns regardless of the underlying dependence structure of $dr_t^{(n)}$ across the curve. A simple analogy to this is a moving average of a time-series process which produces a smoother, and hence more locally correlated, series than the underlying process.

We can write equation (7) in matrix form as

$$R = C_1 D,$$

where R is an $N \times 1$ vector of excess returns $rx_t^{(n)}$, D is an $N \times 1$ vector of difference returns, and C_1 is an $N \times N$ lower triangular matrix of ones.[3] Thus, the covariance matrix of R is $V_R = C_1 V_D C_1'$, where V_D is the covariance matrix of D. The loadings from a principal component analysis of this matrix are given by the eigenvectors of V_R with order of importance based on their associated eigenvalues.

The next lemma aims at isolating the effect of partial summation on the loadings. For this purpose, we assume that $dr_t^{(n)}$ are mutually uncorrelated with a variance matrix equal to $\sigma^2 I_N$, where I_N denotes the $N \times N$ identity matrix, so that $V_R = \sigma^2 C_1 C_1'$.[4] We acknowledge that this is an unrealistic case but it is instructive in understanding the effect of the definitional maturity relationship between $rx_t^{(n)}$ and $dr_t^{(n)}$ on principal component analyses. Moreover, it enables a simple characterization of the eigenvectors and eigenvalues of V_R provided in the following lemma.

Lemma 1. *Suppose $V_R = \sigma^2 C_1 C_1'$ with $\sigma^2 \in (0, \infty)$. Let λ_n^0 denote the eigenvalues of the matrix V_R with corresponding eigenvectors $\psi_n = (\psi_{1,n}, \ldots, \psi_{N,n})'$. Then,*

$$\lambda_n^0 = \frac{\sigma^2}{2 - 2\cos\left(\frac{(2n-1)\pi}{2N+1}\right)}, \qquad \lambda_1^0 > \lambda_2^0 > \cdots > \lambda_{N-1}^0 > \lambda_N^0, \tag{8}$$

[3] For simplicity of presentation, in our theoretical results we define the maximum maturity to be $N + 1$ so that $n = 2, \ldots, N + 1$.

[4] See also Lekkos (2000) for a numerical investigation of a similar setup for yields.

where $\sum_{n=1}^{N} \lambda_n^0 = \frac{\sigma^2}{2}(N+1)N$, *and*

$$\psi_{h,n} = \frac{2}{\sqrt{2N+1}} \sin\left(\frac{h(2n-1)\pi}{2N+1}\right) \tag{9}$$

for $h = 1, \ldots, N$ *and* $n = 1, \ldots, N$. *Finally,* $\|\psi_n\| = 1$, *where* $\|\cdot\|$ *denotes the Euclidean norm.*

Proof. See Appendix A.

The expressions in Lemma 1 are related to the Karhunen-Loève (KL) approximation of a Wiener process,[5] as N is allowed to go to infinity, as well as the basis functions (the discrete cosine transform) in Müller and Watson (2008, 2018):

$$\varpi_{h,n} \propto \cos\left(\frac{(2h-1)n\pi}{2N}\right).$$

There are two distinct features of the expansion implied by Lemma 1. The first is that the basis functions are similar, but not the same as those of the asymptotic expansions mentioned above. The second is that this is an exact representation for the discrete-time random walk with homoskedastic Gaussian errors.[6] The important point is that the cross-sectional telescoping sum of uncorrelated difference returns induces an eigenfunction structure which is fully characterized only by the time to maturity.

Figure 2. Principal Component Output: Excess Returns. The figure shows the analytical output (based on Lemma 1) from principal component analysis applied to the covariance matrix of excess returns, where difference returns are assumed to have a covariance matrix proportional to an identity matrix. The maximum maturity is $N = 20$. The left chart shows the principal component loadings for the first three principal components; the right chart shows the scree plot for the first ten principal components with horizontal line at $1/N$.

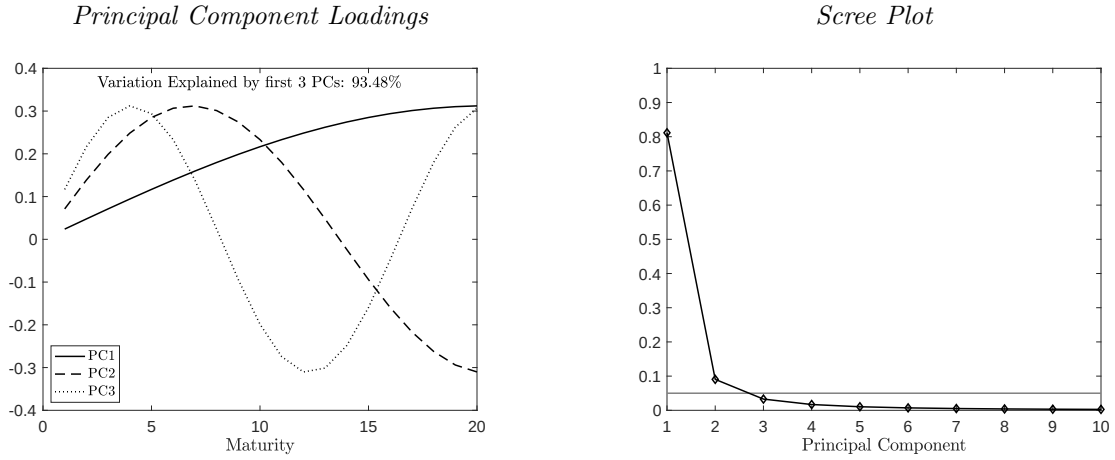


Figure 2 illustrates the analytical principal component loadings along with the scree plot for the setting of Lemma 1. We observe that the factor loadings for holding-period returns take a

[5] See, for example, Shorack and Wellner (1986) for a textbook treatment and Phillips (1998) for an early application of the KL approximation to the econometrics literature.

[6] See the discussion in Tanaka (2017). It follows immediately that a sum of an independent random walk and white noise process has a similar exact expansion with identical eigenvectors and eigenvalues of the form $\lambda_n^0 + \sigma_e^2$, where σ_e^2 is the variance of the additive white noise.

highly structured and parsimonious polynomial pattern. More importantly, the first three principal components explain the bulk (in excess of 93%) of the cross-sectional variation of bond returns. This holds despite the fact that the true process is described by N idiosyncratic factors in $dr_t^{(n)}$. To emphasize this, the scree plot includes a line at $1/N$ which is the share of variance explained by each factor when applied to the covariance matrix for difference returns. The shape of the loadings is nearly indistinguishable from that of the empirical loadings in Figure 1 despite the fact that Lemma 1 makes the empirically implausible assumption of mutually uncorrelated difference returns. The scree plot differs from its empirical counterpart in magnitude but not in shape. This deficiency can be rectified by accommodating local correlation in difference returns as we do in the next section.

Finally, observe that the ratio of adjacent eigenvalues $\lambda_j^0/\lambda_{j+1}^0$, which has been used to form an estimator of the number of factors in approximate factor models (Ahn and Horenstein, 2013), may be written explicitly. This ratio will be monotonically decreasing in j with $\lambda_j^0/\lambda_{j+1}^0 \approx 1$ for large j. In contrast, in a reduced-rank setting, the ratio of adjacent eigenvalues will diverge when j is equal to the rank of the matrix.

3.2 Local Correlation in Difference Returns

The assumption in the previous section that difference returns are uncorrelated, idiosyncratic processes is unrealistic as it does not produce a smooth forward curve. In order to generate a more realistic covariance structure, consider a discrete-time analog (Gáll, Pap, and van Zuijlen, 2006) of the random field model in Kennedy (1997), Goldstein (2000), and Santa-Clara and Sornette (2001). Let $\mathbb{Z}_+$ be the set of positive integers and $\triangle S_{t+1}^{(n)} = S_{t+1}^{(n)} - S_t^{(n)}$, where $\{S_t^{(n)}\}_{t,n\in\mathbb{Z}_+}$ is a two-dimensional random field across time and maturity with $\{S_t^{(n)}\}_{t\in\mathbb{Z}_+}$ adapted to some filtration $\{\mathcal{F}_t\}_{t\in\mathbb{Z}_+}$ for each $n \in \mathbb{Z}_+$. Next, suppose that forward rates evolve over time as

$$f_{t+1}^{(n)} = f_t^{(n)} + \alpha_t^{(n)} + \sigma \triangle S_{t+1}^{(n)}, \tag{10}$$

where $\alpha_t^{(n)} \in \mathcal{F}_t$ is a maturity-specific, time-varying drift and forward rates of different maturities are each driven by the conditionally mean-zero processes $\triangle S_{t+1}^{(n)}$.[7] To ensure smoothness of the forward curve, $S_t^{(n)}$ is assumed to satisfy

$$S_{t+1}^{(n)} = S_t^{(n)} + \rho S_{t+1}^{(n-1)} - \rho S_t^{(n-1)} + \epsilon_{t+1}^{(n)},$$

where $|\rho| < 1$, $S_t^{(0)} = 0\ \forall t \in \mathbb{Z}_+$, $S_0^{(n)} = 0\ \forall n \in \mathbb{Z}_+$, and $S_0^{(0)} = 0$. The cross-sectional structure of $\triangle S_{t+1}^{(n)}$ is thus modeled as a spatial AR(1) process,

$$\triangle S_{t+1}^{(n)} = \rho \triangle S_{t+1}^{(n-1)} + \epsilon_{t+1}^{(n)}.$$

[7] When $\triangle S_{t+1}^{(n)}$ is replaced by $S_t - S_{t-1}$, the model coincides with that of Heath, Jarrow, and Morton (1992), where the forward rates are driven by a single stochastic process.

For tractability, we make several simplifying assumptions. First, assume that $\{\epsilon_t^{(n)}\}$ are jointly normally distributed and mutually independent with $\mathbb{E}[\epsilon_t^{(n)}] = 0$ and $\mathbb{V}[\epsilon_t^{(n)}] = 1$ for all $n \in \mathbb{Z}_+$, where $\mathbb{E}[\cdot]$ and $\mathbb{V}[\cdot]$ denote the expectation and variance operator, respectively. Second, we consider an exponentially affine pricing kernel with constant prices of risk. We denote the price of risk associated with $\epsilon_t^{(j)}$ by $\phi^{(j)}$ which satisfies $\left|\sum_{j=1}^{\infty} \phi^{(j)}\rho^j\right| < \infty$. To ensure that there is an absence of arbitrage opportunities, restrictions need to be placed on $\{\alpha_t^{(n)}\}$ which, following Gáll, Pap, and van Zuijlen (2006, Corollary 1), imply

$$f_t^{(n+1)} - f_{t+1}^{(n)} = dr_{t+1}^{(n+1)} = \kappa^{(n)} - \sigma \sum_{i=1}^{n} \rho^{n-i}\epsilon_{t+1}^{(i)}, \tag{11}$$

where $\kappa^{(n)} = \frac{1}{1-\rho^2}\sum_{k=1}^{\infty} \phi^{(k)} \left(\rho^{|n-k|} - \rho^{n+k}\right) - \frac{\sigma}{2}\frac{(1-\rho^n)^2}{(1-\rho)^2}$.

Thus, the parameter ρ controls the local correlation of difference returns and the covariance matrix of difference returns V_D has (i,j) element

$$\mathbb{C}\left(dr_t^{(i+1)}, dr_t^{(j+1)}\right) = \frac{\sigma^2(\rho^{|i-j|} - \rho^{i+j})}{1-\rho^2},$$

where $\mathbb{C}(\cdot,\cdot)$ is the covariance operator. Thus, V_D can be written as

$$V_D = \frac{\sigma^2}{1-\rho^2}\left(\mathcal{T}(\rho) - \rho^2 \cdot \mathcal{H}(\rho)\right), \tag{12}$$

where $\mathcal{T}(\rho)$ and $\mathcal{H}(\rho)$ are Toeplitz and Hankel matrices each with first row equal to $(1, \rho, \rho^2, \ldots, \rho^{N-1})$ and $\mathcal{H}(\rho)$ has last row equal to $(\rho^{N-1}, \rho^N, \ldots, \rho^{2N-2})$.

It appears that V_D does not have an eigendecomposition which is available in closed form; however, we may write V_D as a rank-one perturbation of an alternative matrix which has an analytical eigendecomposition. Define $\tilde{V}_D^{-1}$ as the $N \times N$ tridiagonal matrix with off-diagonal elements all equal to $-\rho/\sigma^2$, first $N-1$ diagonal elements equal to $(1+\rho^2)/\sigma^2$, and final diagonal element equal to $(1-\rho+\rho^2)/\sigma^2$. Then,

$$V_D^{-1} = \tilde{V}_D^{-1} + xx', \tag{13}$$

where $x = (0, \ldots, 0, \sqrt{\rho(1-\rho)/\sigma^2})'$. Both V_D^{-1} and $\tilde{V}_D^{-1}$ are tridiagonal matrices and differ only by their (N,N) element. Notice that as $\rho \to 1$, then $\tilde{V}_D^{-1} \to V_D^{-1}$.[8] In the next lemma, we provide explicit forms of the eigenvalues and eigenvectors of $\tilde{V}_D$.

Lemma 2. *Assume that $0 < \rho < 1$ and $\sigma^2 \in (0,\infty)$. Let V_D be as in equation (12) and $\tilde{V}_D$ be defined as above. Also, let λ_n denote the eigenvalues of $\tilde{V}_D$ and ψ_n be the eigenvectors defined in Lemma 1. Then, the eigenvalue decomposition of $\tilde{V}_D$ is $\tilde{V}_D = \Psi\Lambda\Psi'$, where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$,*

[8] When $\rho < 0$, we would use an alternative proxy matrix, say $\breve{V}_D^{-1}$, which replaces the (N,N) element of $\tilde{V}_D^{-1}$ with $(1+\rho+\rho^2)/\sigma^2$. In this case, as $\rho \to -1$, we would have $\breve{V}_D^{-1} \to V_D^{-1}$. All of our results could be straightforwardly modified to deal with this case.

with

$$\lambda_n = \frac{\sigma^2}{1+\rho^2-2\rho\cos\left(\frac{(2n-1)\pi}{2N+1}\right)}, \qquad \lambda_1 > \lambda_2 > \cdots > \lambda_{N-1} > \lambda_N,$$

$\sum_{n=1}^{N}\lambda_n = \frac{\sigma^2}{2}\frac{(2N+1)(1-\rho^{2N+1})(1+\rho)-(1+\rho^{2N+1})(1-\rho)}{(1+\rho^{2N+1})(1-\rho^2)(1+\rho)}$ *and* $\Psi = [\psi_1 \cdots \psi_N]$. *Moreover,* $V_D = \tilde{V}_D - \varkappa\varkappa'$, *where*

$$\varkappa = \frac{\sqrt{\rho(1-\rho)}\sum_{n=1}^{N}\lambda_n\psi_n\psi_{N,n}}{\sqrt{\sigma^2+\rho(1-\rho)\sum_{n=1}^{N}\lambda_n\psi_{N,n}^2}}.$$

Define $z = \Psi'\varkappa$ *with* i*-th element* z_i. *Then, the eigenvalue decomposition of* V_D *is* $V_D = \Gamma\Phi\Gamma'$, *where the eigenvalues* $(\phi_1, \ldots, \phi_N)$ *are the* N *solutions to the characteristic equation*

$$f(\phi) = 1 - \sum_{i=1}^{N}\frac{z_i^2}{\lambda_i - \phi} = 0,$$

and the n*-th eigenvector in* $\Gamma = [\gamma_1 \cdots \gamma_N]$ *is*

$$\gamma_n = \frac{\Psi(\Lambda - \phi_n I_N)^{-1}z}{\|\Psi(\Lambda - \phi_n I_N)^{-1}z\|}.$$

Proof. See Appendix A.

Lemma 2 characterizes the eigenstructure of V_D with respect to the approximating matrix $\tilde{V}_D$. When ρ is in the vicinity of one, the loadings on the principal components for difference returns are well approximated by perturbations of sine functions.[9] Again, this is due to the strong local correlation which now arises from a *weighted* partial sum of processes across maturities, with weights depending on ρ. The approximations of the eigenvalues also show that when ρ is near unity, the first eigenvalue dominates as $1+\rho^2-2\rho\cos(\pi/(2N+1))$ is close to zero, and the "level" factor will account for the vast majority of the cross-sectional variation in $dr_t^{(n)}$. However, the data are generated by N factors and the first few principal components are not sufficient to accurately span the true factor space and consequently describe the driving forces of the term structure.

Remark 1. For fixed N, Lemma 2 provides an alternative approximate (in distribution) orthogonal expansion for a highly (positively) correlated autoregressive process; i.e., for fixed t

$$dr_{t+1}^{(n+1)} - \kappa^{(n)} \approx_d \sum_{j=1}^{N}\sqrt{\lambda_j}\cdot\psi_{n,j}\,v_j, \tag{14}$$

where the analytical expressions for $\{\psi_{n,j}\}$ and $\{\lambda_j\}$ are given in Lemmas 1 and 2, respectively, and v_j are *i.i.d.* standard Gaussian random variables. Equation (14) can be compared to the KL

[9] Some earlier papers in mathematical finance (Forzani and Tolmasky, 2003; Lord and Pelsser, 2007; Salinelli and Sgarra, 2007) have alerted the finance literature of the sources of the level, slope and curvature effect and their approximations as perturbations to sine functions. Lemma 2 provides more general and complete approximations to the eigendecomposition of interest.

expansion of an Ornstein-Uhlenbeck process (see, for example, Hassler and Hosseinkouchack, 2019).

□

Figure 3. Principal Component Output: Difference Returns. This figure shows the output from principal component analysis applied to the variance-covariance matrix of difference returns shown in equation (12) for $\rho = 0.99$ and maximum maturity $N = 20$ based on Lemma 2. The left chart shows the principal component loadings for the first three principal components of V_D and of $\tilde{V}_D$ (denoted as "PC Approx."); the right chart shows the scree plot for the first ten principal components.

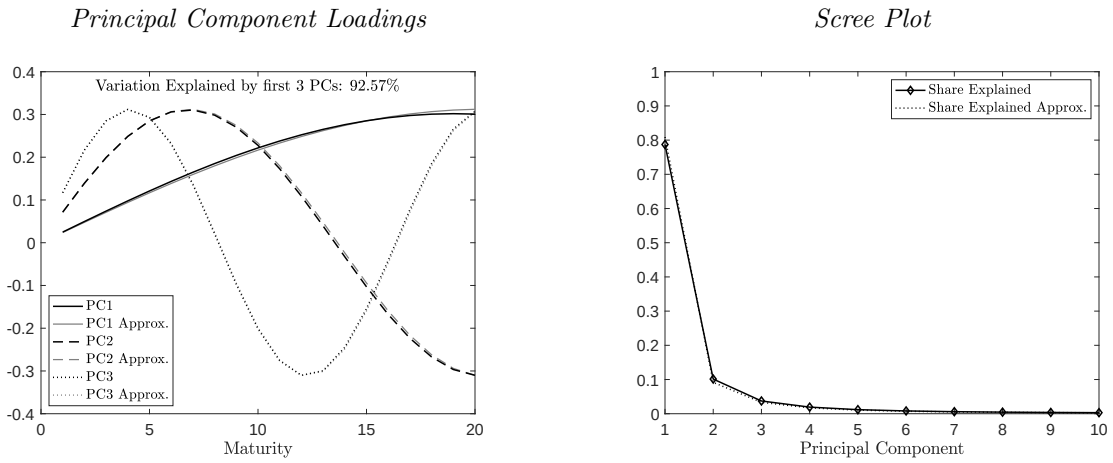


Figure 3 presents the results of Lemma 2 in graphical form for the special case of $\rho = 0.99$ and $N = 20$. We again observe the familiar polynomial shape of the loadings (left graph). In the right graph, we show the corresponding scree plot associated with the 10 largest eigenvalues. The first component "explains" a little less than 80% of the variation, the second component accounts for a little more than 10%, and the third component much less, again suggesting only a few factors; however, the data are generated by N different processes (see equation (11)).

Both panels of Figure 3 also display the approximating eigenvectors and eigenvalues from Lemma 2. We observe that the approximations are very close to the exact values. Furthermore, in Appendix B in the Supplementary Material we provide tight bounds on the eigenvalues of V_D by making a simple modification to the bounds proposed in Ipsen and Nadler (2009) for a rank-one update of a symmetric matrix. Appendix B in the Supplementary Material also presents expressions for bounds on the eigenvectors of V_D (see also Figure B.1 in the Supplementary Material).

We conclude the section with several remarks on the implications of Lemma 2. First, observe that in the setup of Lemma 1, PCA of difference returns, rather than excess returns, will recover the true factor space. In contrast, Lemma 2 demonstrates that in the presence of strong local correlation – a robust feature of the data – we encounter difficulties in characterizing the minimal dimension of the factor space even for difference returns. Specifically, PCA of difference returns will indicate a low-dimensional factor space while the omitted factors, albeit small in terms of explained variation, can be shown to contain important economic information (see examples in Section 3.3).

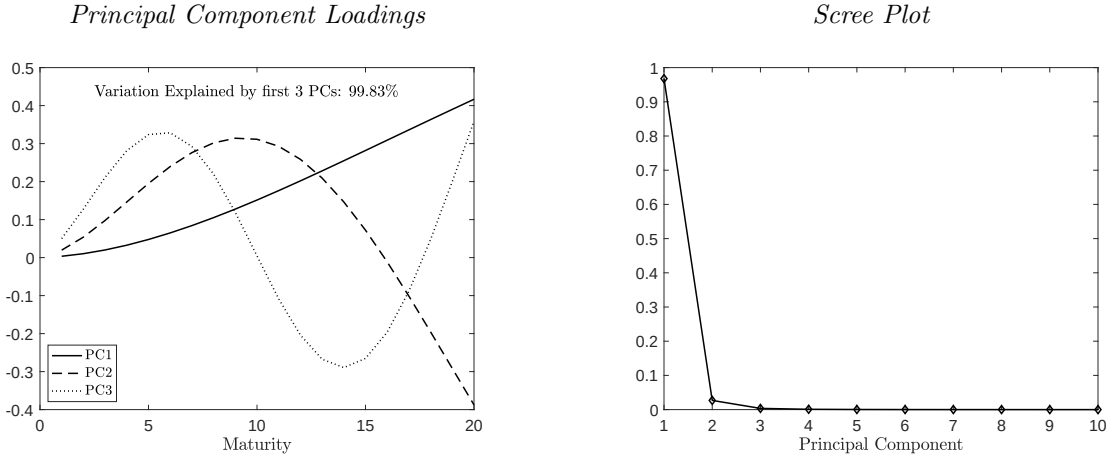
Second, consider the general form of the covariance matrix of R

$$V_R = C_1 V_D C_1', \tag{15}$$

where V_D has the special form in equation (12). Here, V_R^{-1} is no longer tridiagonal but instead pentadiagonal. This makes the analytical characterization of the eigenvalues and eigenvectors more onerous. However, it can be shown (Crump and Gospodinov, 2019) that the eigenvectors are solutions to a (fifth-order) difference equation, so they can be represented as polynomial transformations of the persistence parameter ρ and the maturity N (as in Lemma 2).

Figure 4 shows the principal component output based on the matrix in equation (15) for $\rho = 0.99$ and $N = 20$. The first thing to note is that the shape of the loadings is largely unchanged as compared to Figure 2. This demonstrates that the first three principal components exhibit near observationally equivalent behavior despite the widely different values of ρ in the two cases. Moreover, now both the loadings and the scree plot in Figure 4 are essentially indistinguishable from their empirical counterparts in the left column in Figure 1 for excess returns. It is important to emphasize again that the true data generating process has N factors. Finally, note that as ρ approaches one, the cross-sectional dependence of $rx_t^{(n)}$ can be approximated by a near-integrated process of order two and further intuition about its properties can be gained from the literature on eigenvalues of integrated Brownian motion (see Theorem 7 of Freedman, 1999).

Figure 4. Principal Component Output: Excess Returns. This figure shows the output from principal component analysis applied to the variance-covariance matrix of excess returns in equation (15) for $\rho = 0.99$ and maximum maturity $N = 20$. The left chart shows the principal component loadings for the first three principal components; the right chart shows the scree plot for the first ten principal components.



3.3 Examples

In this section, we explore the economic significance of our theoretical results. In the two examples, we show that variation beyond the assumed factor structure can have an outsize distortion in common applications such as portfolio allocation and hedging. Let $R_t = \left(rx_t^{(2)}, \ldots, rx_t^{(N+1)}\right)'$ be the vector of excess returns with corresponding variance matrix, V_R, given by equation (15), with λ_i and φ_i being its ith ordered eigenvalue and eigenvector. For reference, when $\rho = 0.99$ and $N = 20$, the first three principal components account for 99.83% of the variation (see Figure 4). Since both examples are static problems, the variance matrix of the risky assets coincides with V_R, and we omit the subscript t for simplicity of notation. It is important to emphasize that these two

examples, and our main results more generally, involve population problems. As such, we abstract from the many challenges of estimating the relevant objects from the data.

Example 1 (Mean-Variance Optimization)**.** Consider the static portfolio optimization problem choosing portfolio weights ω with desired return $\bar{r}$

$$\omega^* = \min_{\omega} \frac{1}{2}\omega'\Gamma\omega \qquad \text{s.t.} \quad \omega'\mu = \bar{r} - r^{(1)},$$

where Γ is a symmetric, nonsingular weight matrix and $\mu = \mathbb{E}[R]$. Thus, the optimal portfolio weights satisfy $\omega^* \propto \Gamma^{-1}\mu$. Then, under the correct specification $\Gamma = V_R$, the optimal portfolio allocation satisfies $\omega^* \propto V_R^{-1}\mu$ with corresponding Sharpe ratio equal to $SR = \left(\mu' V_R^{-1}\mu\right)^{1/2}$.

However, suppose an investor incorrectly concludes that the number of factors is $m < N$ with additive idiosyncratic innovations featuring a scalar variance matrix. Then, in population, the investor utilizes $\Gamma = V_R^{(m)}$, where $V_R^{(m)} = \sum_{i=1}^{m} \lambda_i \varphi_i \varphi_i' + \frac{1}{N-m}\text{trace}\left(\sum_{i=m+1}^{N} \lambda_i \varphi_i \varphi_i'\right) \cdot I_N$, with corresponding Sharpe ratio

$$SR^{(m)} = \frac{\mu' \left(V_R^{(m)}\right)^{-1} \mu}{\sqrt{\mu' \left(V_R^{(m)}\right)^{-1} V_R \left(V_R^{(m)}\right)^{-1} \mu}}.$$

We can then compare the Sharpe ratios from the correctly specified variance matrix versus the incorrectly specified matrix based on a restricted factor space. The left chart of Figure 5 shows the reduction in the Sharpe ratio relative to the optimal portfolio choice, $SR^{(m)}/SR$, for different values of ρ and $N = 20$. The values of μ are calibrated using data on Eurodollar futures employed in Figure 1. The graph starts with $m = 3$. For $\rho = 0.99$, the Sharpe ratio with the misspecified variance matrix is far smaller – approximately 35% that of the optimal portfolio. As m increases, the relative Sharpe ratios slowly converge requiring more than ten principal components to achieve 90% of the optimal Sharpe ratio. Thus, although the approximation would appear, at first glance, to be sufficient, it leads to portfolio allocations that can be very far away from the optimal allocation. Interestingly, for smaller values of ρ, the effect is modestly muted. This arises because there is less commonality to be exploited for portfolio optimization and so the degree of misspecification has relatively less influence. However, values of ρ closer to one is the empirically relevant case. The intuition for this result is that the inverse of V_R is a pentadiagonal matrix and implies that each asset's portfolio weight only depends on excess returns from nearby maturities. In contrast, the inverse of $V_R^{(m)}$ is a full matrix and each asset's portfolio weight loads on excess returns for all maturities. This distorts the weights relative to the optimal choice and compromises the risk-return trade-off in the bond portfolio. □

Example 2 (Hedging Errors)**.** Suppose that we are interested in hedging a portfolio of bonds based on portfolio weights ω_p using m principal component portfolios

$$\min_{\omega \in \mathbb{R}^m} \mathbb{V}\left(\omega_p' R - \omega' \varphi^{(m)\prime} R\right), \qquad \varphi^{(m)} = (\varphi_1, \varphi_2, \ldots, \varphi_m). \tag{16}$$

Given a portfolio ω_p, we can then solve for the optimal choice of ω given by $\omega^* = \left(\varphi^{(m)\prime} V_R \varphi^{(m)}\right)^{-1} \varphi^{(m)\prime} V_R \omega_p$. The efficacy of the hedge can be assessed by calculating

$$\text{Corr}\left(\omega_p' R, \omega^{*\prime} \varphi^{(m)\prime} R\right) = \frac{\omega_p' V_R \varphi^{(m)} \omega^*}{\sqrt{\omega_p' V_R \omega_p \cdot \omega^{*\prime} \varphi^{(m)\prime} V_R \varphi^{(m)} \omega^*}}, \tag{17}$$

which measures the correlation in returns between the target portfolio and the mimicking portfolio based on only m principal components. Equivalently, the expression in equation (17) is the square root of the population R^2 of a regression of the target portfolio on the hedging portfolio.

The right chart of Figure 5 shows this population correlation for $m = 3$ and for 20 portfolios (single bonds of different maturities) across the same three values of ρ as in Example 1. The labels on the x-axis correspond to $\omega_p = e_i$, where e_i is an 20×1 vector with i-th element equal to one and all other elements equal to zero. The graph shows that this correlation is very far from one for single securities with short maturities for all three value of ρ. This is true despite the fact that the explained variation of the first three principal components is nearly 100% for $\rho = 0.99$ (see Figure 4) and 98.8% and 97.0% for $\rho = 0.75$ and $\rho = 0.50$, respectively. This reinforces the notion that factor analysis on excess returns is approximately invariant to the true factor space in difference returns and will always appear to have a low-dimensional representation. However, this masks that there are many more factors needed to span the space of bond returns, and so the hedging efficacy is compromised.

In fact, even larger hedging errors can coexist with high explained variation from a few factors. Intuitively, this can occur because there exist choices of ω_p that are elements of the subspace of the omitted "factors" when using only a small number of principal components. Clearly, when ω_p may be written as a linear combination of the columns of $\varphi^{(m)}$, then one can hedge this portfolio perfectly producing a correlation of one. In contrast, when ω_p may be written as a linear combination of basis vectors which span the space orthogonal to the column space of $\varphi^{(m)}$, then the hedging correlation is exactly zero. The reason is that the variation of the portfolio returns is exclusively driven by the omitted factors. More formally, consider the projection matrix (with eigenvalues zero or one) on the orthogonal complement of the space spanned by $V_R^{1/2} \varphi^{(m)}$

$$I_N - V_R^{1/2} \varphi^{(m)} \left(\varphi^{(m)\prime} V_R \varphi^{(m)}\right)^{-1} V_R^{1/2} \varphi^{(m)\prime}, \tag{18}$$

where $V_R^{1/2}$ is a symmetric square root of V_R. Let v_i, $i = 1, ..., N$, denote the ordered eigenvectors of the matrix in (18) so that the first m eigenvectors are those associated with an eigenvalue of 0 and the last $N - m$ eigenvectors are associated with an eigenvalue of 1. For any portfolio weights ω_p, we can write $\omega_p = V_R^{-1/2} (c_1 v_1 + \cdots + c_N v_N)$ for some $N \times 1$ vector $c = (c_1, \ldots, c_N)'$. Then, it can be shown that

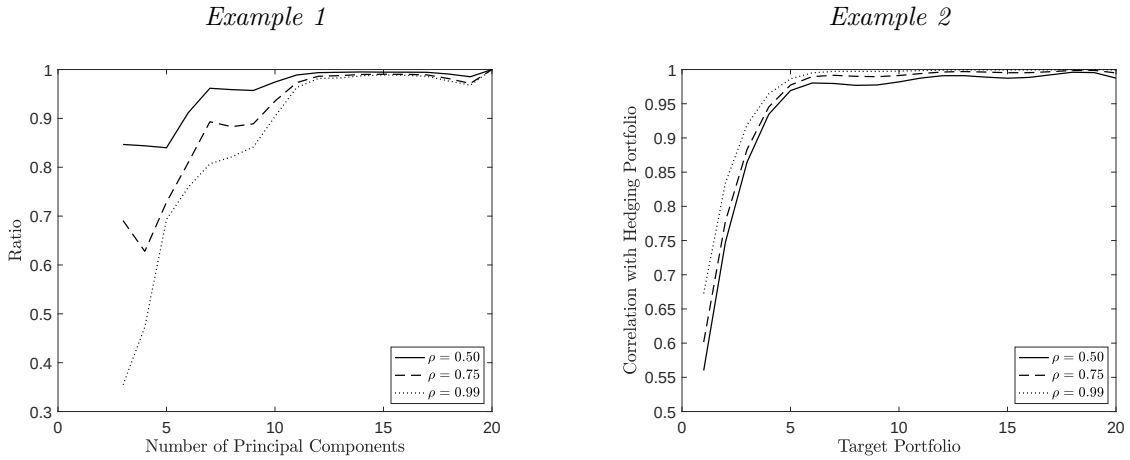
$$\text{Corr}\left(\omega_p' R, \omega^{*\prime} \varphi^{(m)\prime} R\right) = \frac{\|c_{1:m}\|}{\|c\|},$$

where $c_{1:m} = (c_1, \ldots, c_m)'$. It then follows immediately that setting $c_1 = ... = c_m = 0$ or $c_{m+1} =$

... $= c_N = 0$ renders the correlation 0 or 1, respectively, with all other values of c constituting intermediate cases.

□

Figure 5. Implications of Misspecified Factor Space. This figure shows the results from two numerical exercises introduced in Examples 1 and 2. The left chart compares the Sharpe ratios from the optimal portfolio allocation using the correct variance matrix versus that based on an approximate variance matrix obtained from PCA for $m = 3, \ldots, N$. The right chart reports the correlation between the hedging portfolio ($m = 3$) and a single-bond target portfolio.



Examples 1 and 2 suggest that relying on a low-dimensional factor structure – when the true space is characterized by many factors as in the setup in this paper – could result in economically significant hedging errors and bond misallocations. While these examples have an intentionally simple design, their implications are of practical relevance to institutions with large bond portfolios such as insurance companies or fund managers. Even in population, standard selection approaches will likely fail to determine the correct number of factors due to the monotonically decreasing eigenvalue structure.

4 Further Discussion and Conclusions

So how consequential is the extraction of a small number of principal components in practice? It depends on the purpose of the analysis. In certain applications, where the goal is statistical dimension reduction, using a small number of principal components may be adequate. More generally, any sufficiently smooth curve will be well approximated by a projection onto a few basis functions (such as those given in Lemma 1) that would be quite informative about its shape. On the other hand, if the interest lies in analyzing the term structure of interest rates, caution must be exercised in attaching economic interpretation to the factors and emphasizing goodness-of-fit metrics. Our theoretical results suggest that characterizing the factor space of bond returns with any degree of confidence faces formidable challenges. Examples 1 and 2 also show that there may be a significant cost to erroneously restricting the factor space even if a few components account for a high degree of explained variation.

It should be stressed that none of these points is specific to the term structure of interest rates and they would apply to term structures of other assets as well. Crump and Gospodinov (2019) provide extensive evidence of level-slope-curvature effects in various maturity-ordered asset markets such as oil futures, currency futures, inflation swaps, S&P 500 options, dividend prices, and international term structures of interest rates as well as evidence from nonfinancial applications such as climate data. The commonality across these examples is the natural cross-sectional ordering inherent in these variables along with strong local correlation. Thus, our results apply directly to any asset with finite maturity, suggesting that empirical results based on parsimonious, highly parameterized models may lack robustness.

In summary, our paper shows that determining the minimal dimension of the data generating process for the term structure – which is common practice in empirical work – is far more difficult than currently appreciated. At the heart of this challenge lies the extreme cross-sectional dependence across maturities that produces a polynomial pattern in the eigenvectors of the covariance matrices of these processes, regardless of the true dimension of their factor space. This is further exacerbated by factor analyses applied to bond returns or yields as the cross-sectional dependence is amplified by the overlap across maturities embedded in the definition of these objects. Moreover, the standard tools and metrics for factor models are ill suited to summarize the pattern of eigenvalues and eigenvectors induced by these local correlations as they always indicate a low-dimensional factor space. This echoes with the difficulties in identifying the true factor space discussed elsewhere (e.g., Brown, 1989) and argues against taking a strong stand on the exact factor structure of yield data when evaluating term structure models and simulating the yield curve (Crump and Gospodinov, 2019).

Our analysis has direct implications for many unresolved questions about the term structure of interest rates. For example, our theoretical results call into question the ease with which we designate a factor as "spanned" or "unspanned" (e.g., Duffee, 2011; Joslin, Priebsch, and Singleton, 2014) as the distinction necessarily relies on an accurate characterization of the true factor structure of the data. Reconciling this question will have significant ramifications for understanding bond risk premiums. More generally, our paper brings to the fore a certain fragility in the existing data-driven methods for selecting factor models, as these methods are unreliable even in our simple stylized setup.

References

Ahn, S. C., and A. R. Horenstein (2013): "Eigenvalue Ratio Test for the Number of Factors," *Econometrica*, 81(3), 1203–1227.

Brown, S. J. (1989): "The Number of Factors in Security Returns," *Journal of Finance*, 44(5), 1247–1262.

Bunch, J. R., C. P. Nielsen, and D. C. Sorensen (1978): "Rank-One Modification of the Symmetric Eigenproblem," *Numerische Mathematik*, 31(1), 31–48.

Chen, H. (2002): "On a New Trigonometric Identity," *International Journal of Mathematical Education in Science and Technology*, 33(2), 306–309.

Crump, R. K., and N. Gospodinov (2019): "Deconstructing the Yield Curve," Staff Report 884, Federal Reserve Bank of New York.

Duffee, G. R. (2011): "Information in (and not in) the Term Structure," *Review of Financial Studies*, 24(9), 2895–2934.

——— (2013): "Forecasting Interest Rates," in *Handbook of Economic Forecasting*, ed. by G. Elliott, and A. Timmermann, vol. 2, chap. 7, pp. 385–426. Elsevier.

Forzani, L., and C. Tolmasky (2003): "A Family of Models Explaining the Level-Slope-Curvature Effect," *International Journal of Theoretical and Applied Finance*, 6(3), 239–255.

Freedman, D. (1999): "On the Bernstein-Von Mises Theorem with Infinite-Dimensional Parameters," *Annals of Statistics*, 27(4), 1119–1140.

Gáll, J. M., G. Pap, and M. C. van Zuijlen (2006): "Forward Interest Rate Curves in Discrete Time Settings Driven by Random Fields," *Computers and Mathematics with Applications*, 51(3-4), 387–396.

Garbade, K. (1996): *Fixed Income Analytics*. MIT Press.

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

Goldstein, R. S. (2000): "The Term Structure of Interest Rates as a Random Field," *Review of Financial Studies*, 13(2), 365–384.

Golub, G. H. (1973): "Some Modified Matrix Eigenvalue Problems," *SIAM Review*, 15(2), 318–334.

Gürkaynak, R., and J. H. Wright (2012): "Macroeconomics and the Term Structure," *Journal of Economic Literature*, 52(2), 331–367.

Hassler, U., and M. Hosseinkouchack (2019): "Ratio Tests Under Limiting Normality," *Econometric Reviews*, 38(7), 793–813.

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

Ipsen, I., and B. Nadler (2009): "Refined Perturbation Bounds for Eigenvalues of Hermitian And Non-Hermitian Matrices," *SIAM Journal on Matrix Analysis and Applications*, 31(1), 40–53.

Joslin, S., M. Priebsch, and K. J. Singleton (2014): "Risk Premiums in Dynamic Term Structure Models with Unspanned Macro Risks," *Journal of Finance*, 69(3), 1197–1233.

Kennedy, D. P. (1997): "Characterizing Gaussian Models of the Term Structure of Interest Rates," *Mathematical Finance*, 7(2), 107–118.

Lekkos, I. (2000): "A Critique of Factor Analysis of Interest Rates," *Journal of Derivatives*, 8(1), 72–83.

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

Lord, R., and A. Pelsser (2007): "Level-Slope-Curvature - Fact or Artefact?," *Applied Mathematical Finance*, 14(2), 105–130.

Müller, U. K., and M. W. Watson (2008): "Testing Models of Low-Frequency Variability," *Econometrica*, 76(5), 979–1016.

——— (2018): "Long-Run Covariability," *Econometrica*, 3(86), 775–804.

Onatski, A., and C. Wang (2021): "Spurious Factor Analysis," *Econometrica*, 89(2), 591–614.

Phillips, P. C. B. (1998): "New Tools for Understanding Spurious Regressions," *Econometrica*, 66(6), 1299–1325.

Piazzesi, M. (2010): "Affine Term Structure Models," in *Handbook of Financial Econometrics*, vol. 1, chap. 12, pp. 691–758. Elsevier.

Quoniam, J. M., and M. G. Greening (1968): "A Trigonometric Summation," *American Mathematical Monthly*, 75(4), 405–406.

Salinelli, E., and C. Sgarra (2007): "Shift, Slope and Curvature for a Class of Yields Correlation Matrices," *Linear Algebra and its Applications*, 426, 650–666.

Santa-Clara, P., and D. Sornette (2001): "The Dynamics of the Forward Interest Rate Curve with Stochastic String Shocks," *Review of Financial Studies*, 14(1), 149–185.

Shorack, G. R., and J. A. Wellner (1986): *Empirical Processes with Applications to Statistics*. Wiley.

Stock, J. H., and M. W. Watson (2016): "Dynamic Factor Models, Factor-Augmented Vector Autoregressions, and Structural Vector Autoregressions in Macroeconomics," in *Handbook of Macroeconomics*, ed. by J. B. Taylor, and H. Uhlig, vol. 2A, chap. 8, pp. 415–525. Elsevier.

Tanaka, K. (2017): *Time Series Analysis: Nonstationary and Noninvertible Distribution Theory*. Wiley, 2 edn.

Uhlig, H. (2009): "Comment on 'How Has the Euro Changed the Monetary Transmission Mechanism?'," in *NBER Macroeconomics Annual*, ed. by D. Acemoglu, K. Rogoff, and M. Woodford, vol. 24, pp. 141–152. MIT Press.

Willms, A. R. (2008): "Analytic Results for the Eigenvalues of Certain Tridiagonal Matrices," *SIAM Journal on Matrix Analysis and Applications*, 30(2), 639–656.

Yueh, W.-C. (2005): "Eigenvalues of Several Tridiagonal Matrices," *Applied Mathematics E-Notes*, 5, 66–74.

Appendix A Proofs of Main Results

The proofs of Lemmas 1 and 2 employ the relationship between the non-zero eigenvalues of a matrix and its inverse. The inverses of the covariance matrices of interest take the form of a tridiagonal matrix with certain restrictions on its elements. We repeatedly make use of the closed-form expressions for the eigenvalues and eigenvectors of a tridiagonal matrix of a specific form (Willms, 2008; Yueh, 2005).

Proof of Lemma 1: The inverse of $\sigma^2 C_1 C_1'$ takes the following tridiagonal form

$$(\sigma^2 C_1 C_1')^{-1} = \frac{1}{\sigma^2}\begin{bmatrix} 2 & -1 & 0 & 0 & .... & & & 0 \\ -1 & 2 & -1 & 0 & 0 & & & 0 \\ 0 & -1 & 2 & -1 & 0 & & & ... \\ ... & & & & ... & & & \\ & & & & & ... & & 0 \\ & & & & & -1 & 2 & -1 \\ 0 & & & & ... & 0 & -1 & 1 \end{bmatrix}.$$

The eigenvalues of this matrix are (Willms, 2008; Yueh, 2005)

$$\zeta_n = \frac{2 - 2\cos(\theta_n)}{\sigma^2},$$

where $\zeta_1 < \zeta_2 < \ldots < \zeta_N$, and the j-th element of the n-th eigenvector is

$$\varphi_{j,n} = \sin(j\theta_n),$$

where $0 < \theta_n < \pi$ satisfies the characteristic equation

$$\frac{\sin((N+1)\theta_n)}{\sin(\theta_n)} - \frac{\sin(N\theta_n)}{\sin(\theta_n)} = \frac{\cos((2N+1)\theta_n)}{\cos(\theta_n/2)} = 0.$$

This characteristic equation has N solutions of the form $\theta_n = \frac{(2n-1)\pi}{2N+1}$ for $n = 1, ..., N$. Thus, the eigenvalues of $\sigma^2 C_1 C_1'$ are

$$\lambda_n^0 = \zeta_n^{-1} = \frac{\sigma^2}{2[1 - \cos((2n-1)\pi/(2N+1))]}.$$

The eigenvectors have elements $\varphi_{j,n}$. Define $\psi_{j,n} = 2\varphi_{j,n}/\sqrt{2N+1}$. Then, by similar arguments as in Quoniam and Greening (1968), we have that $\|\psi_n\| = 1$. Finally, note that the diagonal elements of $C_1 C_1'$ are $1, \ldots, N$ so that trace $(C_1 C_1') = N(N+1)/2$. □

Proof of Lemma 2: The inverse of V_D takes the following tridiagonal form

$$V_D^{-1} = \frac{1}{\sigma^2}\begin{bmatrix} 1+\rho^2 & -\rho & 0 & 0 & .... & & & 0 \\ -\rho & 1+\rho^2 & -\rho & 0 & 0 & & & 0 \\ 0 & -\rho & 1+\rho^2 & -\rho & 0 & & & ... \\ ... & & & & ... & & & \\ & & & & & ... & -\rho & 0 \\ & & & & & -\rho & 1+\rho^2 & -\rho \\ 0 & & & & ... & 0 & -\rho & 1 \end{bmatrix}.$$

Then, $V_D^{-1} = \tilde{V}_D^{-1} + xx'$, where x is an $N \times 1$ vector with last element equal to $\sqrt{\frac{\rho(1-\rho)}{\sigma^2}}$ and all other elements equal to zero. Thus, $\tilde{V}_D^{-1}$ has (N, N) element equal to $\frac{1-\rho+\rho^2}{\sigma^2}$ and all other elements identical to those of V_D^{-1}. The eigenvalues of this matrix are (Willms, 2008; Yueh, 2005)

$$\zeta_n = \frac{1 + \rho^2 - 2\rho\cos(\theta_n)}{\sigma^2},$$

where $\zeta_1 < \zeta_2 < \ldots < \zeta_N$. Furthermore, the j-th element of the n-th eigenvector is

$$\varphi_{j,n} = \sin(j\theta_n),$$

where $0 < \theta_n < \pi$ again satisfies the characteristic equation

$$\frac{\sin((N+1)\theta_n)}{\sin(\theta_n)} - \frac{\sin(N\theta_n)}{\sin(\theta_n)} = \frac{\cos((2N+1)\theta_n)}{\cos(\theta_n/2)} = 0.$$

This characteristic equation has N solutions of the form $\theta_n = \frac{(2n-1)\pi}{2N+1}$ for $n = 1, ..., N$. Thus, the eigenvalues for $\tilde{V}_D$ are

$$\lambda_n = \zeta_n^{-1} = \frac{\sigma^2}{1 + \rho^2 - 2\rho\cos((2n-1)\pi/(2N+1))}$$

and, as in Lemma 1, the eigenvectors are ψ_n. By the results in Chen (2002, equation (7)), setting $n = 2N+1$ and $\alpha = \frac{\pi}{2N+1}$, we obtain that

$$\sum_{n=1}^{N} \frac{\sigma^2}{1 + \rho^2 - 2\rho\cos((2n-1)\pi/(2N+1))} = \frac{\sigma^2}{2}\frac{(2N+1)(1-\rho^{2N+1})(1+\rho) - (1+\rho^{2N+1})(1-\rho)}{(1+\rho^{2N+1})(1-\rho^2)(1+\rho)},$$

where we have used that $\cos(\cdot)$ is an even function. By the Sherman-Morrison formula, we obtain that $V_D = \tilde{V}_D - \varkappa\varkappa'$.

Next, using that $\Psi\Psi' = I_N$, observe that

$$\Psi\Lambda\Psi' + \varkappa\varkappa' = \Psi(\Lambda + zz')\Psi',$$

where $z = \Psi'\varkappa$. Thus, the problem reduces to the eigenvalue decomposition of $\Lambda + zz'$ with eigenvalues that satisfy the equation $\det(\Lambda + zz' - \phi I_N) = 0$. Then, we have (Golub, 1973),

$$\begin{aligned}
\det(\Lambda + zz' - \phi I_N) &= \det((\Lambda - \phi I_N) + zz') \\
&= (1 + z'(\Lambda - \phi I_N)^{-1} z)\det(\Lambda - \phi I_N) \\
&= \left(1 + \sum_{i=1}^{N} \frac{z_i^2}{\lambda_i - \phi}\right)\left(\prod_{i=1}^{N}(\lambda_i - \phi)\right).
\end{aligned}$$

The second multiplicative term is nonzero as long as $0 < \rho < 1$, and the eigenvalues $(\phi_1, ..., \phi_N)$ can be obtained as the N roots from the characteristic equation

$$f(\phi) = 1 + \sum_{i=1}^{N} \frac{z_i^2}{\lambda_i - \phi} = 0.$$

The eigenvector corresponding to the n-th eigenvalue is then given by (Bunch, Nielsen, and Sorensen, 1978)

$$\gamma_n = \frac{\Psi(\Lambda - \phi_n I_N)^{-1} z}{\|\Psi(\Lambda - \phi_n I_N)^{-1} z\|}.$$

□