JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS Vol. 53, No. 2, Apr. 2018, pp. 937–963

doi:10.1017/S0022109018000030

Staying on Top of the Curve: A Cascade Model of Term Structure Dynamics

Laurent E. Calvet, Adlai J. Fisher, and Liuren Wu*

Abstract

This paper specifies term structure dynamics by a recursive cascade of heterogeneously persistent factors. The cascade naturally orders economic shocks by their adjustment speeds, and generates smooth interest-rate curves in closed form. For a class of specifications, the number of parameters is invariant to the size of the state space, and the term structure converges to a stochastic limit as the state dimension goes to infinity. High-dimensional specifications fit observed term structure almost perfectly, match the observed low correlation between movements in different maturities, and produce stable interest-rate forecasts that outperform lower-dimensional specifications.

I. Introduction

The development of arbitrage-free dynamic term structure models (DTSMs) ranks as one of the most important achievements of modern asset pricing. Theoretical characterization of broad classes of DTSMs (e.g., the affine specifications

*Calvet (corresponding author), laurent.calvet@edhec.edu, EDHEC Business School; Fisher, adlai.fisher@sauder.ubc.ca, University of British Columbia Sauder School of Business; and Wu, liuren.wu@baruch.cuny.edu, Baruch College Zicklin School of Business. The authors thank David Backus, Hendrik Bessembinder (the editor), Tomas Björk, Peter Carr, Anna Cieslak, Antonio Diez de los Rios, Bjorn Flesaker, Mike Gallmeyer, Robert Goldstein (a referee), Jeremy Graveline, Dmitry Kreslavskiy, Markus Leippold, Vadim Linetsky, Andrei Lyashenko, Fabio Mercurio, Per Mykland, Andrew Siegel, Harvey Stein, David Weinbaum, Yildiray Yildirim, Hao Zhou, an anonymous referee, and seminar participants at Bloomberg, Cheung Kong Graduate School of Business, Columbia University, FGV/EBAPE, HEC Paris, Northwestern University, the Stockholm School of Economics, Syracuse University, the University of Chicago, the University of Minnesota, the University of Zürich, Ziff Brothers Investments, the 2010 Econometric Society World Congress, the 2010 Financial Econometrics Workshop at the Fields Institute in Toronto, the 2010 International Conference on Computing in Economics and Finance, the 2010 International Symposium on Business and Industrial Statistics, the 2010 McGill University Risk Management Conference, the 2010 Northern Finance Association conference, the 2010 Triple Crown Conference at Rutgers University, the 2011 Annual European Finance Association conference, and the 2011 Society of Financial Econometrics conference. Earlier versions of this paper circulated under the title “Dimension-Invariant Dynamic Term Structures.”

of Duffie and Kan (1996)) paved the way for an explosion of empirical work (e.g., Dai and Singleton (2000), Duffee (2002)).[1]

The vast majority of empirical studies consider term structure models with a small number of factors. Low-dimensional DTSMs, however, have important limitations. Estimated models imperfectly fit the yield curve at a given point in time, with economically significant pricing errors (Bali, Heidari, and Wu (2009)). DTSMs are consequently rarely used to impute forward rates, a key starting point for pricing derivatives (Heath, Jarrow, and Morton (1992)). Low-dimensional DTSMs generate substantially higher cross-correlations between changes in yields of different maturities than are observed in the data, suggesting that these specifications do not capture important sources of cross-sectional variation (Dai and Singleton (2003)). Furthermore, low-dimensional DTSMs tend to predict interest rates less accurately than a simple random walk (Duffee (2002), Ang and Piazzesi (2003)).

A natural solution is to consider DTSMs with higher state dimensions, which requires critical improvements in identification methodology. A generic 3-factor affine construction already calls for more than 20 parameters, many of which are not well identified empirically (Duffee and Stanton (2008), Duffee (2011)). When additional factors are considered, the number of parameters grows quadratically with the number of states, which is an instance of the curse of dimensionality. Recent literature has devoted significant effort to enhancing identification and simplifying estimation by i) deriving closed-form approximations of the likelihood function (Ait-Sahalia and Kimmel (2010)); ii) recasting term structure models in linear regression and other convenient forms (Adrian, Crump, and Moench (2013), Hamilton and Wu (2012)); and iii) developing normalizations with observable yields as states (Collin-Dufresne, Goldstein, and Jones (2008), Joslin, Singleton, and Zhu (2011)). Converting the term structure model into a vector autoregression (VAR) simplifies estimation methodology, but in a high-dimensional unconstrained VAR, the strength of identification for a large number of parameters remains an issue. Thus, in all canonical DTSMs, increasing the state dimension results in a rapid increase in the dimension of the parameter space and weak identification, which are major impediments to empirical progress.

aim

In this paper, we develop parsimonious classes of arbitrage-free DTSMs in which the number of parameters is invariant to the number of factors. These dimension-invariant models eliminate the trade-off between the richness of the state space and the strength of identification, thereby defeating the curse of dimensionality inherent to general DTSMs. Our approach is based on a cascade construction for factor dynamics. We posit a lowest-frequency first component that provides a central tendency to which the second factor mean reverts. The remaining factors, which may be arbitrary in number, follow a recursion whereby

[1]Important early contributions include Balduzzi, Das, Foresi, and Sundaram (1996), Brennan and Schwartz (1979), Constantinides (1992), Cox, Ingersoll, and Ross (1985), Longstaff and Schwartz (1992), Pearson and Sun (1994), and Vasicek (1977). Duffie, Pan, and Singleton (2000) and Duffie, Filipović, and Schachermayer (2003) progressively generalize the class of affine models and Leippold and Wu (2002) introduce the quadratic class of arbitrage-free DTSMs. Further extensions are provided by Backus, Foresi, Mozumdar, and Wu (2001), Dai and Singleton (2002), Piazzesi (2005), and Rudebusch and Wu (2008).

each component mean reverts around the preceding factor. We take the highest-frequency factor to be the instantaneous interest rate. The cascade structure naturally ranks the factors by their adjustment speeds.

While a generic factor structure can be rotated in multiple ways to sharpen economic interpretation, our cascade structure highlights the interplay between economic shocks of different durations. For example, Dalio (2012) contrasts long-term debt cycles, which can span decades or even a century, to moderate-term productivity/business cycles, which last about 5 years on average in the United States.[2] Furthermore, within each business cycle, monetary policies experience adjustments and revisions, generating shorter-term fluctuations. At even higher frequencies, shocks to the supply and demand for bonds can also generate price fluctuations over daily to quarterly horizons. These different macroeconomic, monetary, and microstructure shocks all impact bond prices, as documented, for example, in Ang and Piazzesi (2003), Diebold, Rudebusch, and Aruba (2006), Gallmeyer, Hollifield, and Zin (2005), and Piazzesi (2005).[3] By requiring that higher-frequency factors mean revert to lower-frequency factors, our cascade structure captures the nesting behavior of the different economic cycles and highlights their impacts on the interest-rate term structure.

In fixed-income markets, the most frequently quoted interest-rate maturities are spaced about evenly on a logarithmic scale. Our baseline specification correspondingly assumes that the adjustment speeds follow a geometric progression. We tightly specify volatilities and risk premia and obtain a model in which term structure dynamics are fully specified by as few as 5 parameters, irrespective of the number of factors. As a result, low- and high-dimensional specifications of the cascade DTSM are equally well identified.

The cascade DTSM accommodates a wide range of examples. For instance, within the Gaussian-affine class, low-dimensional versions of the cascade include the 1-factor model of Vasicek (1977) and the 2-factor model of Balduzzi, Das, and Foresi (1998). While the volatility of each factor is constant in our baseline specification, tractable extensions permit time variation in factor volatility with only a few additional, well-identified parameters. We show how to incorporate time-varying risk premia of the form consistent with recent empirical evidence (Cochrane and Piazzesi (2005), (2008)). Furthermore, the cascade implies that bond yields can be written as linear combinations of exponential functions of bond maturities. It is therefore consistent with the use of exponential basis functions for stripping spot and forward curves (Nelson and Siegel (1987), Svensson (1995)). The cascade also complements recent work modifying the Nelson–Siegel model (Christensen, Diebold, and Rudebusch (2011)) or proposing a set of exponential-based forward-rate curves (Chua, Foster, Ramaswamy, and Stine (2008)) that can

[2]The National Bureau of Economic Research has identified 33 business cycles over the 156-year period between 1854 and 2009.

[3]See also Ang, Dong, and Piazzesi (2007), Ang, Piazzesi, and Wei (2004), Balduzzi, Bertola, and Foresi (1997), Bekaert, Cho, and Moreno (2005), Gallmeyer, Hollifield, Palomino, and Zin (2005), Heidari and Wu (2010), Hördahl, Tristanoi, and Vestin (2006), Lu and Wu (2009), Rudebusch (2002), and Rudebusch, Swanson, and Wu (2006).

be reconciled with arbitrage-free dynamics.[4] We advance this literature by developing a parsimonious class of base functions generated by a cascade DTSM.

When the number of high-frequency factors goes to infinity, the sequence of cascade DTSMs converges to a well-defined, infinite-dimensional DTSM. We prove convergence under an explicit set of sufficient conditions on the sequence of factor adjustment speeds, volatilities, and risk premia. We also provide a condition under which the sequence of term structures and its limit have bounded maximal Sharpe ratios, addressing Duffee's (2010) concern that high-dimensional DTSMs may have unreasonably large Sharpe ratios. These results imply that finite-dimensional versions of the cascade can approximate arbitrarily closely the rich dynamics of infinite-state frameworks. The cascade thereby builds a bridge between standard, low-dimensional DTSMs and the infinite-dimensional term structures developed using the mathematics of random fields (Kennedy (1994), (1997), Goldstein (2000)).[5]

We demonstrate empirically the value added by our high-dimensional cascade. We estimate specifications with 2–10 factors using a 20-year panel of U.S. dollar London Interbank Offered Rates (LIBOR) and swap rates with maturities ranging from 6 months to 30 years. The dimension-invariant feature allows us to identify the parameters of the models with high statistical significance regardless of the number of factors. Estimates of the maximal Sharpe ratio are U-shaped in the number of factors and are all within a sensible interval, ranging from 0.22 for the 3-factor specification to 0.89 for the 10-factor model.

Comparing the model performance across different number of factors, we find that the 10-factor model significantly outperforms lower-dimensional models by several metrics. First, the root-mean-squared pricing errors of the 10-factor model average less than 1 basis point (bp), an order of magnitude smaller than its 3-factor counterpart and very close to the range of common bid–ask spreads. The near-perfect fit makes the model an ideal candidate for stripping smooth and dynamically consistent forward-rate curves for derivative pricing. Second, high-dimensional DTSMs accurately reflect empirical cross-correlations of interest-rate changes at different maturities. Third, the 10-factor model significantly outperforms autoregressive specifications and either matches or outperforms the random walk in predicting interest-rate movements out of sample.

The cascade's parsimony, large state space, and invariance of the parameter space to the size of the state space are reminiscent of earlier work by Calvet and Fisher (2001), (2004), (2007), (2008), which specifies the volatility term structure

[4]In contrast to the DTSM literature, which derives the term structure behavior from state dynamics and risk premia assumptions, a separate line of research directly proposes functional forms that can be used as the basis for stripping spot and forward interest-rate curves, such as exponential basis functions, polynomials (Chambers, Carleton, and Waldman (1984)), cubic splines (McCulloch (1975), Litzenberger and Rolfo (1984)), step functions (Ronn (1987)), and piecewise linear specifications (Fama and Bliss (1987)). Filipović (1999) shows that popular functional forms, such as the exponential basis functions proposed by Nelson and Siegel (1987), cannot be made consistent with any interest-rate dynamics. Since then, the literature has placed increased emphasis on excluding dynamic arbitrage on the assumed functional forms.

[5]Related work and applications include Collin-Dufresne and Goldstein (2003), Han (2007), Longstaff, Santa-Clara, and Schwartz (2001), Longstaff and Schwartz (2001), and Santa-Clara and Sornette (2001).

using a multiplicative cascade.[6] The additive cascade proposed in the present paper provides a new approach to modeling the term structure of interest rates with a high degree of analytical tractability. We anticipate that variants and extensions of the additive cascade can successfully be applied to many other areas of study, including variance dynamics and general time-series analysis.

The remainder of the paper is organized as follows: Section II develops the general cascade model. Section III provides assumptions under which the number of parameters of the model is invariant to the number of factors and gives sufficient conditions for convergence of the term structure to a stochastic limit as the dimensionality of the state space grows. Section IV describes the data and reports estimation results. Section V considers applications to yield-curve stripping, interest-rate cross-correlations, and interest-rate and principal-component forecasting. An Internet Appendix (available at www.jfqa.org) provides proofs, robustness checks, and extensions.

II. A Cascade Model of the Interest-Rate Term Structure

We consider an infinite-horizon, continuous-time economy. Time is indexed by $t \in [0,\infty)$. Uncertainty is represented by a filtered probability space $\{\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\geq 0}\}$, where Ω is the sample space, $\mathcal{F}$ is a σ-algebra on Ω, $\mathbb{P}$ is the physical measure, and $(\mathcal{F}_t)_{t\geq 0}$ is a filtration on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We assume that the usual conditions of right continuity and completeness with respect to the null sets of $\mathbb{P}$ are satisfied.

A. The Multifrequency Cascade

We model the dynamics of the instantaneous interest rate under the physical measure, $\mathbb{P}$, by a cascade of n diffusions. Consider a constant $\theta_r \in \mathbb{R}$ and an n-dimensional Wiener process $\mathbf{W}_t = (W_{1,t}, \ldots, W_{n,t})^\top$ with independent components. The dynamics of the n-dimensional state vector $\mathbf{X}_t = (x_{1,t}, \ldots, x_{n,t})^\top \in \mathbb{R}^n$ are defined as follows:

Assumption 1. Factor Structure. We let $x_{0,t} = \theta_r$ and specify the factors by the cascade of diffusions:

$$dx_{j,t} = \kappa_j (x_{j-1,t} - x_{j,t})\, dt + \sigma_j\, dW_{j,t}, \tag{1}$$

for every $j \in \{1, \ldots, n\}$, where $\kappa_1, \ldots, \kappa_n$ and $\sigma_1, \ldots, \sigma_n$ are strictly positive constants. Furthermore, the adjustment speeds satisfy $\kappa_1 < \kappa_2 < \cdots < \kappa_n$.

The first factor, $x_{1,t}$, is an Ornstein–Uhlenbeck process mean reverting to θ_r. For every $j > 1$, the jth factor, $x_{j,t}$, mean reverts toward the $(j-1)$th factor, $x_{j-1,t}$. The factors mean revert more quickly as j increases.

Assumption 2. Short Rate. The instantaneous interest rate coincides with the last factor: $r_t = x_{n,t}$ for all $t \geq 0$.

The cascade can be conveniently rewritten in matrix notation:

$$d\mathbf{X}_t = \boldsymbol{\kappa}\, (\boldsymbol{\theta} - \mathbf{X}_t)\, dt + \boldsymbol{\Sigma}^{1/2} d\mathbf{W}_t, \tag{2}$$

[6] See also Calvet and Fisher (2013) for a recent survey of multiplicative cascades in finance and Calvet, Fearnley, Fisher, and Leippold (2015) for an application of these methods to equity options.

where $\boldsymbol{\theta}=(\theta_r,\ldots,\theta_r)^\top$, $\boldsymbol{\Sigma}$ is a diagonal matrix with diagonal elements $\sigma_1^2,\ldots,\sigma_n^2$, and $\boldsymbol{\kappa}$ is a matrix with diagonal elements $\kappa_{j,j}=\kappa_j$, off-diagonal elements $\kappa_{j,j-1}=-\kappa_j$, and all other elements equal to 0.

In a general version of equation (2), factors can interact through both the covariance matrix, $\boldsymbol{\Sigma}$, and off-diagonal elements of the adjustment matrix, $\boldsymbol{\kappa}$. At the same time, as Dai and Singleton (2000) emphasize, the general system (2) is not well identified, so that restrictions on the matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{\kappa}$ must be imposed in estimation. Our specification imposes a diagonal structure on the covariance matrix $\boldsymbol{\Sigma}$ and captures factor interactions through the sparse lower triangular $\boldsymbol{\kappa}$ matrix, which delivers a well-identified multifrequency cascade.

B. Factor Representation of the Short Rate and Response Functions

The contribution of the different factors is illustrated by the following representation of the short rate:

Proposition 1. Factor Representation of the Short Rate. The instantaneous interest rate, r_t, is an affine function of the initial state vector $\mathbf{X}_0=(x_{1,0},\ldots,x_{n,0})^\top$ and integrals of the innovations driving each factor between 0 and t:

$$(3)\qquad r_t = \theta_r+\sum_{j=1}^{n}a_j(t)(x_{j,0}-\theta_r)+\sum_{j=1}^{n}\sigma_j\int_0^t a_j(t-s)\,dW_{j,s}.$$

Each response function $a_j(\tau)$ is proportional to the convolution product of exponential probability density functions:

$$(4)\qquad a_j(\tau) = (K_j*\ldots*K_n)(\tau)/\kappa_j,$$

where $*$ denotes the convolution operator, and for every $i\in\{1,\ldots,n\}$, $K_i(\tau)=\kappa_i e^{-\kappa_i\tau}$ if $\tau\geq 0$ and $K_i(\tau)=0$ if $\tau<0$.

Fisher and Gilles (1996) show that in any affine model, the vector of factors, $\mathbf{X}_t$, can be written as a linear combination of the initial factors, $\mathbf{X}_0$, the long-run levels of the factors, $\boldsymbol{\theta}$, and stochastic integrals of the Wiener process, $\mathbf{W}_t$, between dates 0 and t. Proposition 1 shows how the general expression simplifies for our cascade. Furthermore, since the factors have constant volatilities, the cascade belongs to the solvable class defined by Grasselli and Tebaldi (2008). The cascade allows us to leverage the benefits of these earlier results in a tightly specified setting.

For a given j, the response function, $a_j(\tau)$, quantifies how a unit shock to factor j at time $t-\tau$ impacts the instantaneous interest rate at time t. By equation (4), the response function, $a_j(\tau)$, depends on both j and n, but for expositional simplicity, we keep the dependence with respect to n implicit throughout the main text. The response functions, $a_j(\tau)$, are rescaled convolution products of exponential densities and therefore take positive values for every $\tau\in(0,\infty)$. They also satisfy $\int_0^{+\infty}a_j(\tau)\,d\tau=1/\kappa_j$ for all j.

The nth response function, $a_n(\tau)=e^{-\kappa_n\tau}$, is equal to unity for $\tau=0$, a consequence of the assumption that the short rate coincides with the nth factor. The function $a_n(\tau)$ then decays with the time horizon at a rate controlled by κ_n.

By equation (4), the $(n-1)$th response function is

$$(5) \qquad a_{n-1}(\tau) \quad = \quad \frac{\kappa_n}{\kappa_n - \kappa_{n-1}} \left(e^{-\kappa_{n-1}\tau} - e^{-\kappa_n \tau} \right).$$

It is equal to 0 at the origin, $a_{n-1}(0)=0$, indicating that shocks to the $(n-1)$th factor are not immediately incorporated into the short rate. The function then increases, peaks at $\overline{\tau}_{n-1}=(\ln\kappa_n - \ln\kappa_{n-1})/(\kappa_n-\kappa_{n-1})$, and decreases toward 0 as τ goes to infinity.

We show more generally:

Proposition 2. Response Functions. For every $j \in \{1,\dots,n\}$, the response function is given by

$$(6) \quad a_j(\tau) \quad = \quad \sum_{i=j}^{n} \alpha_{i,j}\kappa_i e^{-\kappa_i \tau}, \quad \text{where } \alpha_{i,j} \quad = \quad \frac{\prod_{k=j}^{n} \kappa_k}{\kappa_i \kappa_j \prod_{k=j,k\neq i}^{n} (\kappa_k - \kappa_i)},$$

which is hump-shaped if $j<n$. The maximum response time, $\bar{\tau}_j = \arg\max_\tau a_j(\tau)$, monotonically decreases with j. Furthermore, $0 \leq \sum_{j=1}^{n} a_j(\tau) \leq 1$ for all $\tau \geq 0$.

The humped shape of the response functions follows from the cascade. Instantaneously, only the highest-frequency shock enters the short rate. A lower-frequency shock impacts the corresponding factor, which then impacts the adjacent higher-frequency factor, and so on, until the short rate itself is affected. Each factor therefore drives the short rate at different horizons. The response functions decline to 0 as τ goes to infinity. By equation (3), the long-run expectation of the short rate, $\lim_{\tau\to\infty} \mathbb{E}_t^{\mathbb{P}}(r_{t+\tau})$, is equal to θ_r.

C. Risk Premia and Bond Prices

We assume that there are no arbitrage opportunities, or equivalently that there exists a risk-neutral measure $\mathbb{Q}$.

Assumption 3. Affine Risk Premia. The risk-neutral measure $\mathbb{Q}$ is defined by the following Radon–Nikodým derivative with respect to the physical measure:

$$(7) \qquad \left.\frac{d\mathbb{Q}}{d\mathbb{P}}\right|_t \quad \equiv \quad \prod_{j=1}^{n} \exp\left(-\int_0^t g_{j,s}\sigma_j \, dW_{j,s} - \frac{1}{2}\int_0^t g_{j,s}^2 \sigma_j^2 \, ds \right).$$

The market prices of risk $g_{j,t}$ are affine in $\mathbf{X}_t$:

$$(8) \qquad g_{j,t} \quad = \quad \gamma_j + \boldsymbol{\lambda}_j^\top \mathbf{X}_t,$$

where $\gamma_j \in \mathbb{R}$ and $\boldsymbol{\lambda}_j$ is an $n\times 1$ real column vector.

The assumption allows each factor to have its own potentially time-varying price of risk, $g_{j,t}$. By Girsanov's theorem, the factors are also affine processes under $\mathbb{Q}$:

$$(9) \qquad dx_{j,t} \quad = \quad \left[-g_{j,t}\sigma_j^2 dt + \kappa_j (x_{j-1,t} - x_{j,t}) \right] dt + \sigma_j dW_{j,t}^{\mathbb{Q}},$$

where $\mathbf{W}_t^{\mathbb{Q}} = (W_{1,t}^{\mathbb{Q}},\dots,W_{n,t}^{\mathbb{Q}})^\top$ is a standard Brownian motion under $\mathbb{Q}$.

The factor dynamics under $\mathbb{Q}$ can be rewritten in matrix notation as

$$(10) \qquad d\mathbf{X}_t = \boldsymbol{\kappa}^{\mathbb{Q}}(\boldsymbol{\theta}^{\mathbb{Q}} - \mathbf{X}_t)\,dt + \boldsymbol{\Sigma}^{1/2} d\mathbf{W}_t^{\mathbb{Q}},$$

where $\boldsymbol{\Lambda} = (\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_n)^\top$ is an $n \times n$ matrix, $\boldsymbol{\kappa}^{\mathbb{Q}} = \boldsymbol{\kappa} + \boldsymbol{\Sigma}\boldsymbol{\Lambda}$ is the risk-neutral adjustment matrix, and

$$(11) \qquad \boldsymbol{\theta}^{\mathbb{Q}} = (\boldsymbol{\kappa}^{\mathbb{Q}})^{-1}(\kappa_1\theta_r - \gamma_1\sigma_1^2, -\gamma_2\sigma_2^2, \ldots, -\gamma_n\sigma_n^2)^\top$$

is the long-run mean of the state vector.

Let $P(\mathbf{X}_t, \tau) = \mathbb{E}_t^{\mathbb{Q}}[\exp(-\int_t^{t+\tau} r_s\,ds)]$ denote the time-t value of a zero-coupon bond with 1 dollar par value and expiry date $t+\tau$, where τ denotes the time to maturity in years. Since the DTSM satisfies the conditions of Duffie and Kan (1996), zero-coupon bond values are exponential affine in the state vector,

$$(12) \qquad P(\mathbf{X}_t, \tau) = \exp[-\mathbf{b}(\tau)^\top \mathbf{X}_t - c(\tau)],$$

where the coefficients $\mathbf{b}(\tau)$ and $c(\tau)$ solve the usual system of ordinary differential equations.[7]

D. Closed-Form Bond Prices under Constant Risk Premia

When $\boldsymbol{\lambda}_j = 0$ for all j, risk premia are constant and bond prices are available analytically.

Proposition 3. Pricing under Constant Premia. For every $j \in \{1, \ldots, n\}$, the price loading, $b_j(\tau)$, satisfies

$$(13) \qquad b_j(\tau) = \int_0^\tau a_j(\tau')\,d\tau' = \sum_{i=j}^{n} \alpha_{i,j}\left(1 - e^{-\kappa_i \tau}\right),$$

where the coefficients $\alpha_{i,j}$ are defined in equation (6). The intercept, $c(\tau)$, is given by

$$(14) \qquad \begin{aligned} c(\tau) = {} & \theta_r\kappa_1 \sum_{i=1}^{n} \alpha_{i,1}\left(\tau - \frac{1-e^{-\kappa_i\tau}}{\kappa_i}\right) \\ & - \sum_{j=1}^{n} \gamma_j\sigma_j^2 \sum_{i=j}^{n} \alpha_{i,j}\left(\tau - \frac{1-e^{-\kappa_i\tau}}{\kappa_i}\right) \\ & - \sum_{j=1}^{n} \frac{\sigma_j^2}{2} \sum_{i=j}^{n}\sum_{k=j}^{n} \alpha_{i,j}\alpha_{k,j} \\ & \times \left[\tau - \frac{1-e^{-\kappa_i\tau}}{\kappa_i} - \frac{1-e^{-\kappa_k\tau}}{\kappa_k} + \frac{1-e^{-(\kappa_i+\kappa_k)\tau}}{\kappa_i+\kappa_k}\right]. \end{aligned}$$

Furthermore, the long-run level of the state vector under $\mathbb{Q}$ is

$$\boldsymbol{\theta}^{\mathbb{Q}} = \left(\theta_r - \frac{\gamma_1\sigma_1^2}{\kappa_1}, \theta_r - \frac{\gamma_1\sigma_1^2}{\kappa_1} - \frac{\gamma_2\sigma_2^2}{\kappa_2}, \ldots, \theta_r - \sum_{i=1}^{n}\frac{\gamma_i\sigma_i^2}{\kappa_i}\right)^\top.$$

[7]See Duffie and Kan (1996) for the derivation and solution details.

Each price loading, $b_j(\tau)$, is the integral of the corresponding response function, $a_j(\tau)$. The intercept, $c(\tau)$, is the sum of i) a term proportional to the long-run interest rate under the physical measure, θ_r, ii) a term driven by the risk premia, γ_j, and iii) a convexity adjustment due to interest-rate volatility.

The long-run yield, $y_\infty = \lim_{\tau\to+\infty}[\mathbf{b}(\tau)^\top \mathbf{X}_t + c(\tau)]/\tau$, is given by

$$y_\infty = \theta_r - \sum_{j=1}^{n} \frac{\sigma_j^2}{\kappa_j^2}\left(\gamma_j \kappa_j + \frac{1}{2}\right), \tag{15}$$

as the Internet Appendix shows. In the absence of risk premia ($\gamma_j = 0$), the long-run yield, y_∞, is lower than the long-run level of the short rate, θ_r, and the average yield curve counterfactually slopes downward. However, if the risk premium coefficients, γ_j, are sufficiently negative, the inequality is reversed: $\theta_r < y_\infty$, and the average term structure is upward-sloping.

For every vector of constant premia $(\gamma_1, \ldots, \gamma_n)$, the Radon–Nikodým derivative of the risk-neutral measure $\mathbb{Q}$ simplifies to

$$\left.\frac{d\mathbb{Q}}{d\mathbb{P}}\right|_t \equiv \exp\left(-\sum_{j=1}^{n} \gamma_j \sigma_j W_{j,t} - \frac{t}{2}\sum_{j=1}^{n} \gamma_j^2 \sigma_j^2\right). \tag{16}$$

The maximal Sharpe ratio of any asset is therefore

$$\mathrm{SR}_{\max} = \left(\sum_{j=1}^{n} \gamma_j^2 \sigma_j^2\right)^{1/2}. \tag{17}$$

A portfolio achieves $\mathrm{SR}_{\max}$ if its instantaneous return has a perfect negative correlation with instantaneous variation in the Radon–Nikodým derivative.

E. Forward Curve

Under Assumptions 1–3, the instantaneous forward rate is affine in the state vector:

$$f(\mathbf{X}_t, \tau) \equiv -\frac{\partial \ln P}{\partial \tau}(\mathbf{X}_t, \tau) = \mathbf{b}'(\tau)^\top \mathbf{X}_t + c'(\tau), \tag{18}$$

as equation (12) implies. If risk premia are constant ($\boldsymbol{\Lambda} = 0$), the forward-rate loadings and the short-rate response functions are identical: $\mathbf{b}'(\tau) = \mathbf{a}(\tau)$ (by Proposition 3), and the biased expectations hypothesis holds.[8] If instead, risk premia are time-varying ($\boldsymbol{\Lambda} \neq 0$), the coefficient $\mathbf{b}'(\tau)$ generally differs from $\mathbf{a}(\tau)$. The specification of the risk-neutral measure in Assumption 3 allows considerable flexibility in linking forward rates to expected future spot rates, as we further explore in the Internet Appendix.

III. Dimension-Invariant Term Structures

A. Baseline Specification

This section develops parsimonious cascades, for which the dimension of the parameter space is invariant to the number of factors. In fixed-income markets,

[8] See Piazzesi (2010) for a general discussion.

the progression of the most quoted maturities is approximately geometric. For example, the most commonly quoted maturities for interest-rate swaps are at 2, 3, 5, 10, 15, and 30 years, with each maturity being a multiple of between 1.5 and 2 times the adjacent maturity. These quoting conventions presumably reflect the market's view of how information is distributed along the yield curve. To reflect this empirical observation, we propose

Assumption 4. Geometrically Distributed Adjustment Speeds. The sequence of adjustment speeds follows a geometric progression: $\kappa_j = \kappa_1 b^{j-1}$ for all j, where $b > 1$ and $\kappa_1 > 0$ are constant real numbers.

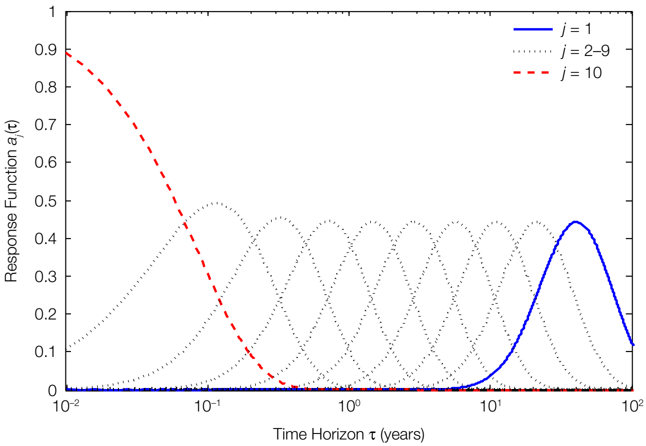
The first factor's adjustment speed, κ_1, and the spacing parameter, b, control the adjustment speeds of all factors.

Figure 1 plots the response functions, $a_j(\tau)$, for a specification with $n = 10$ geometrically distributed frequency components. The durations of the factors range from 1 month ($\kappa_{10} = 12$) to 30 years ($\kappa_1 = 1/30$), implying that the spacing parameter is $b = 1.92$. The response function of the highest-frequency component, $a_n(\tau)$, decays from unity to 0 as τ progresses from 0 to $+\infty$. All other response functions are hump-shaped, as Proposition 2 explains. Moreover, their maxima are spaced almost evenly on a logarithmic time scale, as we further discuss in Section III.B.

FIGURE 1

Short Rate Response to Unit Shocks from Different Frequency Components

Figure 1 illustrates the response of the instantaneous interest rate to unit shocks from each of the 10 frequency components across different time horizons. The solid line denotes response to shocks from the lowest frequency, and the dashed line denotes response to shocks from the highest frequency. The dotted lines represent responses to intermediate frequency components. The responses are computed with the parameters $\kappa_1 = 1/30$ and $\kappa_{10} = 12$, with power scaling in between.



In order to obtain a tightly specified cascade, we make additional assumptions that are consistent with the intuition that each factor carries roughly equal amounts of information and risk:

Assumption 5. Volatilities. Factor volatilities are identical: $\sigma_j = \sigma_1$ for all j.

Assumption 6. Risk Premia. The market prices of risk on all risk factors are time-invariant and identical: $\boldsymbol{\lambda}_j = 0$ and $\gamma_j = \gamma_1$ for all j.

Assumptions 1–6 define the baseline version of the term structure model. We observe that it is fully specified by the 5 parameters: κ_1, b, σ_1, θ_r, and γ_1. By Proposition 3, the long-run level of the first factor under the risk-neutral measure is given by $\theta_r^{\mathbb{Q}} = \theta_r - \gamma_1 \sigma_1^2 / \kappa_1$. Thus, there is a one-to-one mapping between $\theta_r^{\mathbb{Q}}$ and the market price of risk coefficient γ_1. Since $\theta_r^{\mathbb{Q}}$ offers the advantage of being directly comparable to θ_r, we define the parameter vector of the baseline model as

$$\left(\kappa_1, b, \sigma_1, \theta_r, \theta_r^{\mathbb{Q}}\right) \quad \in \quad (0,\infty) \times (1,\infty) \times (0,\infty) \times \mathbb{R}^2.$$

The maximal Sharpe ratio (equation (17)) is then $\mathrm{SR}_{\max} = |\gamma_1| \sigma_1 \sqrt{n}$, or equivalently

$$\text{(19)} \qquad \mathrm{SR}_{\max} \quad = \quad \kappa_1 |\theta_r^{\mathbb{Q}} - \theta_r| \sqrt{n} / \sigma_1.$$

Sections IV and V investigate the empirical performance of the baseline model. The Internet Appendix considers extensions that weaken Assumptions 1–6.

B. Limiting Behavior

We now investigate the behavior of the cascade as the number of high-frequency factors increases. Figure 1 suggests that the response functions are spaced almost evenly on a logarithmic time scale. In order to better understand this property, let $L^2(\mathbb{R}_+)$ denote the space of square-integrable measurable functions $y : \mathbb{R}_+ \longrightarrow \mathbb{R}$, and let $\|y\|_2 = [\int_0^{+\infty} y(\tau)^2 d\tau]^{1/2}$.

Proposition 4. Scaling of the Response Functions. Under Assumptions 1 and 4, the response function of every factor j converges, both uniformly on $\mathbb{R}_+$ and in $L^2(\mathbb{R}_+)$, to a continuous function $\bar{a}_j$ as the number of factors, n, goes to infinity. Furthermore, the limiting response functions satisfy the invariance condition $\bar{a}_{j+1}(\tau) = \bar{a}_j(b\tau)$ for all $j \geq 1$.

The proposition implies that the graphs of the limiting response functions are translated versions of each other on a log scale, as is evident in Figure 1.

The asymptotic scaling of the response functions suggests that, under some appropriate regularity conditions, the short-rate process should converge as well. Consider a fixed initial state $\mathbf{X}_0 = \{x_{j,0}\}_{j=1}^{\infty}$ and fixed parameters $(\kappa_1, b, \{\sigma_j\}_{j=1}^{\infty}, \theta_r)$ satisfying Assumptions 1–4. By Proposition 2, the response functions of an n-dimensional cascade take values between 0 and 1. We infer from Proposition 1 that the variance of the short rate $r_{n,t} = x_{n,t}$ conditional on the initial state satisfies

$$\text{(20)} \qquad \mathrm{var}(r_{n,t}) \;=\; \sum_{j=1}^{n} \sigma_j^2 \int_0^t a_j^2(s)\, ds \;\leq\; \sum_{j=1}^{n} \sigma_j^2 \int_0^{\infty} a_j(s)\, ds \;=\; \sum_{j=1}^{n} \frac{\sigma_j^2}{\kappa_j}.$$

This inequality suggests that the stochastic component of the short rate converges under

Condition 1. Size of the Shocks. $\sum_{j=1}^{\infty} b^{-j} \sigma_j^2 < \infty$.

We note that Condition 1 follows from Assumption 4 (scaling of adjustment speeds) and Assumption 5 (constant factor volatilities) and is therefore satisfied by the baseline specification defined in Section III.A. In this section, we consider weaker conditions guaranteeing convergence, which may guide further developments of the cascade.

We consider $T \in \mathbb{R}_+$ and denote by $L^2(\Omega \times [0,T])$ the Hilbert space of adapted square-integrable processes $y : \Omega \times [0,T] \to \mathbb{R}$. Let

$$\|y\|_{L^2(\Omega\times[0,T])} = \left[\mathbb{E}^{\mathbb{P}}\left(\int_0^T y_t^2\, dt\right)\right]^{1/2}.$$

The deterministic component of the short rate, $\theta_r + \sum_{j=1}^n a_j(t)(x_{j,0} - \theta_r)$, converges absolutely if $\sum_{j=1}^n \|a_j\|_2 |x_{j,0} - \theta_r| < \infty$. Since $\|a_j\|_2 \le 1/\sqrt{\kappa_j}$, it is sufficient to assume

Condition 2. Initial State. $\sum_{j=1}^\infty b^{-j/2} |x_{j,0} - \theta_r| < \infty$.

We show in the Internet Appendix:

Proposition 5. Convergence of the Short Rate. Consider an initial state vector $\{x_{j,0}\}_{j=1}^\infty$ and a vector of parameters $(\kappa_1, b, \{\sigma_j\}_{j=1}^\infty, \theta_r)$ such that Conditions 1 and 2 hold. When the number of factors n goes to infinity, the short rate $r_{n,t} = x_{n,t}$ converges in $L^2(\Omega \times [0,T])$ to the Gaussian process

$$(21) \qquad \bar{r}_t = \theta_r + \sum_{j=1}^\infty \bar{a}_j(t)(x_{j,0} - \theta_r) + \sum_{j=1}^\infty \sigma_j \int_0^t \bar{a}_j(t-s)\, dW_{j,s}.$$

Moreover, if there exist $\varepsilon \in (-\infty, 1/2)$ and $\eta \in (0, +\infty)$ such that the following conditions hold:

$$\sum_{j=1}^\infty b^{-\varepsilon j} \sigma_j^2 < \infty,$$
$$\sup_{1\le j<\infty} \left(b^{\eta j}\, |x_{j,0} - \theta_r|\right) < \infty,$$

then the limiting short-rate process has continuous sample paths.

The condition $\sum_{j=1}^\infty b^{-\varepsilon j} \sigma_j^2 < \infty$, which holds under the baseline specification, guarantees that the stochastic component of the short rate satisfies the Kolmogorov continuity condition. The assumption on the initial state, $\sup_{1\le j<\infty}(b^{\eta j}\, |x_{j,0} - \theta_r|) < \infty$, ensures that the deterministic component of the short rate is continuous; it requires that deviations from θ_r be less pronounced for components with higher adjustment speeds.

We now turn to the convergence of the term structure of zero-coupon yields. Let $\{\gamma_j\}_{j=1}^\infty$ denote a fixed sequence of risk premia that satisfies

Condition 3. Size of the Risk Premia. $\sum_{j=1}^\infty |\gamma_j| \sigma_j^2 / \kappa_j < \infty$.

The condition, which holds under the baseline specification, implies that the long-term yield (equation (15)) converges to the finite limit:

$$(22) \qquad \bar{y}_\infty = \theta_r - \sum_{j=1}^\infty \frac{\sigma_j^2}{\kappa_j^2}\left(\gamma_j \kappa_j + \frac{1}{2}\right),$$

as the number of components goes to infinity. This result suggests that the entire term structure converges. We indeed show:

Proposition 6. Convergence of the Term Structure. Consider an initial state vector $\mathbf{X}_0 = \{x_{j,0}\}_{j=1}^{\infty}$ and a vector of parameters $(\kappa_1, b, \{\sigma_j\}_{j=1}^{\infty}, \theta_r, \{\gamma_j\}_{j=1}^{\infty})$ such that Conditions 1–3 hold. When the number of factors n goes to infinity, the yield of a zero-coupon bond of maturity τ converges in $L^2(\Omega \times [0,T])$ to

$$\bar{y}(\mathbf{X}_t, \tau) \quad = \quad [\bar{\mathbf{b}}(\tau)^{\top}\mathbf{X}_t + \bar{c}(\tau)]/\tau \tag{23}$$

for every $\tau > 0$, where $\bar{\mathbf{b}}(\tau) = (\int_0^{\tau} \bar{a}_j(s)\,ds)_{j=1,\ldots,\infty}$ and the deterministic function $\bar{c}(\tau)$ is provided in the Internet Appendix.

This proposition applies to the baseline specification with an initial state satisfying Condition 2.

Investment opportunities in the limiting term structure are easily derived from equations (16) and (17). When $\sum_{j=1}^{\infty} \gamma_j^2 \sigma_j^2 = \infty$, the Radon–Nikodým derivative converges to 0 and no risk-neutral measure exists in the limiting economy. The maximal Sharpe ratio (equation (17)) correspondingly diverges to infinity as n gets large. When instead $\sum_{j=1}^{\infty} \gamma_j^2 \sigma_j^2 < \infty$, the Radon–Nikodým derivative has a positive limit and the corresponding probability measure $\mathbb{Q}$ can be used to price all bonds. The Sharpe ratio remains bounded, and the cascade with infinitely many factors can meet Duffee's (2010) requirement that a well-specified DTSM should generate reasonable investment opportunities.

The convergence of the cascade builds a bridge between standard, finite-factor DTSMs and the infinite-factor models described using the mathematics of random fields (Kennedy (1994), (1997), Goldstein (2000)). This bridge is useful because our model can be implemented using very simple empirical methods and yet approximates arbitrarily closely the rich dynamics of genuinely infinite-state term structures.

IV. Data and Estimation

We estimate the cascade on a panel of U.S. dollar 6-month LIBOR and swap rates with maturities of 2, 3, 4, 5, 7, 10, 15, 20, and 30 years. The data are obtained from Bloomberg and consist of weekly (Wednesday) closing mid-quotes spanning over 20 years from Jan. 4, 1995, to Oct. 28, 2015. There are 1,087 weekly observations for each of the 10 time series, for a total of 10,870 observations.

A. Summary Statistics

LIBOR and swap rates relate to zero-coupon bond prices by

$$\begin{aligned}
\text{LIBOR}(\mathbf{X}_t, \tau) &= \frac{100}{\tau}\left(\frac{1}{P(\mathbf{X}_t, \tau)} - 1\right), \\
\text{SWAP}(\mathbf{X}_t, \tau) &= 200\,\frac{1 - P(\mathbf{X}_t, \tau)}{\sum_{i=1}^{2\tau} P(\mathbf{X}_t, i/2)}.
\end{aligned}$$

The LIBOR follows actual/360 day-count convention, starting 2 business days forward. The swap contracts make semiannual payments and follow 30/360 day-count convention. Table 1 reports for each series the sample mean, standard deviation, skewness, excess kurtosis, and weekly autocorrelations of order 1, 5, 10, and 20. The average term structure is upward-sloping, while standard deviations decline with maturity. The skewness and excess kurtosis estimates are small. All interest-rate series are highly persistent, with first-order autocorrelations ranging from 0.993 to 0.998.

TABLE 1
Summary Statistics of LIBOR and Swap Rates

Table 1 reports summary statistics on weekly observations (Wednesday closing mid-quotes) of the 6-month LIBOR and swap rates with maturities of 2, 3, 4, 5, 7, 10, 15, 20, and 30 years. Each series contains 1,087 weekly observations from Jan. 4, 1995, to Oct. 28, 2015. Entries report the sample average (Mean), standard deviation (Std. Dev.), skewness (Skew.), excess kurtosis (Kurt.), and weekly autocorrelations of orders 1, 5, 10, and 20, respectively, for each series.

					Autocorrelation (weeks)			
Maturity	Mean	Std. Dev.	Skew.	Kurt.	1	5	10	20
6 months	3.128	2.312	0.116	−1.653	0.998	0.986	0.969	0.930
2 years	3.439	2.284	0.060	−1.490	0.996	0.981	0.961	0.922
3 years	3.699	2.179	−0.005	−1.398	0.996	0.978	0.957	0.917
4 years	3.925	2.074	−0.050	−1.317	0.995	0.976	0.954	0.912
5 years	4.118	1.976	−0.078	−1.251	0.995	0.974	0.950	0.907
7 years	4.417	1.823	−0.106	−1.158	0.994	0.971	0.945	0.898
10 years	4.705	1.689	−0.125	−1.083	0.994	0.969	0.940	0.888
15 years	4.980	1.584	−0.159	−1.029	0.993	0.967	0.937	0.882
20 years	5.094	1.546	−0.183	−1.016	0.993	0.967	0.937	0.880
30 years	5.154	1.508	−0.181	−0.985	0.993	0.967	0.934	0.874
Average	4.266	1.897	−0.071	−1.238	0.995	0.974	0.948	0.901

B. Estimation

We cast the dimension-invariant DTSM into a state-space form. Assume that bond prices are observed at time intervals of length Δt. The state-propagation equation is a discrete-time analog of the statistical dynamics in equation (2):

$$\text{(24)} \qquad \mathbf{X}_t = \mathbf{A} + \mathbf{\Phi}\mathbf{X}_{t-\Delta t} + \mathbf{\Sigma}_x^{1/2}\boldsymbol{\varepsilon}_t,$$

where $\Delta t = 1/52$, $\mathbf{\Phi} = \exp(-\Delta t\, \boldsymbol{\kappa})$, $\mathbf{I}_n$ is the n-dimensional identity matrix, $\mathbf{A} = (\mathbf{I}_n - \mathbf{\Phi})\boldsymbol{\theta}$, $\boldsymbol{\varepsilon}_t$ is independent and identically distributed (IID) $\mathcal{N}(0, \mathbf{I}_n)$, and $\mathbf{\Sigma}_x = \sigma_1^2\, \Delta t\, \mathbf{I}_n$.

The measurement equations are built from observations of LIBOR and swap rates:

$$\text{(25)} \qquad \mathbf{y}_t = \mathbf{h}(\mathbf{X}_t) + \mathbf{e}_t,$$

where $\mathbf{y}_t$ denotes the vector of observed rates, $\mathbf{h}(\mathbf{X}_t)$ is the vector of rates corresponding to the state $\mathbf{X}_t$ under the model, and $\mathbf{e}_t$ is a vector of measurement errors. We assume that the measurement errors are normally distributed IID random variables with 0 mean and variance σ_e^2.

When the state-space model is linear Gaussian, the Kalman (1960) filter provides the exact conditional mean and variance–covariance matrix of the latent state. In the case of the cascade, the state-propagation equation (24) is linear Gaussian but the measurement equation (25) is nonlinear in the state vector. For this reason, we use the unscented Kalman filter (Wan and van der Merwe (2001)) to handle the nonlinearity.

We consider specifications with 2–10 factors. Given the dimension-invariant nature of our specification, all the models are defined by 6 parameters, regardless of number of factors. The 5 parameters ($\kappa_1, b, \sigma_1, \theta_r$, and $\theta_r^{\mathbb{Q}}$) govern the term structure dynamics, while σ_e^2 quantifies the fitting error on the observed LIBOR and swap rates.

Table 2 reports the estimation results. As the number of factors increases, the frequencies of the components become more tightly spaced: The scaling coefficient b decreases from 11.19 with 2 factors to 1.87 with 10 factors. Other parameters tend to stabilize as the number of frequencies increases. For the 10-factor model, the adjustment speed of the lowest-frequency component is $\kappa_1 = 0.0388$, corresponding to a duration $(1/\kappa_1)$ of about 26 years. The scaling coefficient of $b = 1.869$ implies that the highest frequency has a duration of about a month. The statistical long-run mean, θ_r, is about 2.22%. The risk-neutral mean, $\theta_r^{\mathbb{Q}}$, is higher at 13.81%, accommodating the upward-sloping mean term structure. Due to model parsimony and data size, all parameters are estimated with small standard errors in all specifications.

TABLE 2

Parameter Estimates, Standard Errors, and Log Likelihoods

Table 2 reports the maximum likelihood estimates and standard errors (in parentheses) of the model parameters. Each row represents a set of parameter estimates of the cascade with $n=2$ to $n=10$ components. The column under $\mathcal{L}/T$ reports the maximized average weekly log likelihood value for each model. The last column reports the Vuong (1989) likelihood ratio test statistic $\mathcal{V}$ between the 10-factor model and each other model. Asymptotically, the statistic has a standard normal distribution. ** indicates significance at the 1% level.

	Parameters													
n	κ_1		θ_r		σ_1		$\theta_r^{\mathbb{Q}}$		b		σ_e^2		$\mathcal{L}/T$	$\mathcal{V}$
2	0.0303	(0.0004)	0.0000	(0.0000)	0.0150	(0.0000)	0.1076	(0.0012)	11.1919	(0.1929)	0.0200	(0.0001)	13.07	31.64**
3	0.0205	(0.0002)	0.0259	(0.0006)	0.0118	(0.0001)	0.0983	(0.0014)	14.5528	(0.1150)	0.0053	(0.0000)	18.34	25.38**
4	0.0268	(0.0005)	0.0000	(0.0000)	0.0129	(0.0000)	0.0668	(0.0008)	3.6545	(0.0306)	0.0011	(0.0000)	24.09	12.77**
5	0.0311	(0.0002)	0.0000	(0.0000)	0.0139	(0.0000)	0.0614	(0.0003)	3.1113	(0.0111)	0.0002	(0.0000)	28.89	7.36**
6	0.0306	(0.0002)	0.0000	(0.0000)	0.0136	(0.0000)	0.0603	(0.0003)	3.1483	(0.0121)	0.0002	(0.0000)	29.03	7.00**
7	0.0279	(0.0006)	0.0168	(0.0003)	0.0155	(0.0001)	0.1517	(0.0033)	2.0966	(0.0141)	0.0001	(0.0000)	30.95	4.51**
8	0.0387	(0.0009)	0.0186	(0.0004)	0.0161	(0.0001)	0.1337	(0.0030)	1.8444	(0.0109)	0.0001	(0.0000)	31.09	5.81**
9	0.0386	(0.0008)	0.0192	(0.0005)	0.0159	(0.0001)	0.1369	(0.0031)	1.8634	(0.0103)	0.0001	(0.0000)	31.22	5.96**
10	0.0388	(0.0008)	0.0222	(0.0005)	0.0160	(0.0001)	0.1381	(0.0029)	1.8690	(0.0101)	0.0001	(0.0000)	31.28	—

The maximal Sharpe ratio, $\mathrm{SR}_{\max}$, is easily imputed from the parameter estimates in Table 2. By equation (19), $\mathrm{SR}_{\max}$ is U-shaped in the number of factors, ranging from 0.22 for $n=3$ to 0.89 for $n=10$. The cascade therefore implies reasonable values of the maximal Sharpe ratio. In recent work, Duffee (2010) shows that general DTSMs with a large number of factors tend to produce extremely large Sharpe ratios. Our cascade avoids this difficulty and satisfies the criterion that a well-specified DTSM should generate reasonable investment opportunities.

C. In-Sample Fit

As Table 2 reports, in-sample fit improves as we add factors, even though the number of parameters is the same for all specifications. The log likelihood, $\mathcal{L}$, increases and the measurement-error variance, σ_e^2, decreases with the number of components, n. The improvement in fit is rapid for low values of n but levels off as we add more factors, consistent with convergence of the estimated term structure. To assess statistical significance, we report the Vuong (1989) statistic for the difference in likelihood between the 10-factor model and each of the lower-dimensional models.[9] In all cases, a hypothesis of equal likelihoods is rejected at the 1% level, implying that the 10-factor model significantly improves in-sample fit relative to lower-dimensional specifications.

To assess economic significance of the improvement in fit from using a high-dimensional specification, Table 3 reports root-mean-squared pricing errors on each interest-rate series from models with increasingly more frequency components, with the last row reporting the average root-mean-squared errors over the 10 interest-rate series for each model. The root-mean-squared pricing errors average over 12 bps for the 2-factor specification and over 6 bps for the 3-factor specification, which are economically very significant numbers (Bali et al. (2009)). The average root-mean-squared pricing error declines steadily with the number of frequencies and is lower than 1 bp for specifications with 7–10 frequencies.

TABLE 3

Root-Mean-Squared Pricing Errors

Table 3 reports the root-mean-squared pricing errors (in basis points) on the LIBOR and swap rates from the baseline version of the cascade with $n=2$ to $n=10$ components. The last row reports the grand average over the 10 interest-rate series.

	n								
Maturity	2	3	4	5	6	7	8	9	10
6 months	23.21	4.92	1.32	0.31	0.29	0.11	0.08	0.07	0.07
2 years	14.86	9.89	5.06	0.98	0.92	0.47	0.40	0.33	0.30
3 years	14.13	4.06	1.48	1.15	1.14	0.69	0.64	0.58	0.55
4 years	12.41	5.13	3.06	0.92	0.91	0.72	0.65	0.63	0.62
5 years	10.42	6.54	2.91	0.86	0.84	0.65	0.62	0.60	0.60
7 years	8.83	7.61	1.34	1.16	1.15	0.70	0.66	0.63	0.62
10 years	8.99	6.80	2.78	1.38	1.37	0.94	0.84	0.82	0.81
15 years	8.62	3.73	2.76	1.35	1.34	0.70	0.61	0.61	0.61
20 years	9.02	4.58	1.69	1.53	1.53	0.66	0.61	0.61	0.61
30 years	13.13	9.85	2.71	1.34	1.33	0.47	0.43	0.42	0.42
Average	12.36	6.31	2.51	1.10	1.08	0.61	0.55	0.53	0.52

V. Applications

A. Zero-Coupon Bond Prices and Forward-Rate Curves

The high-dimensional cascade provides a promising framework to generate stripped zero-coupon bond prices and forward-rate curves that match observed LIBOR and swap rates well, as we show in this section.

Forward-rate curves are fundamental starting points for derivative pricing approaches (e.g., Heath et al. (1992)). Typically, constructing a forward-rate curve from a discrete number of observations of coupon bonds or swap rates involves choosing a functional form (e.g., the exponential form of Nelson and Siegel (1987), Svensson (1995)) to link the forward rates across different maturities. Standard stripping methods, however, can be overly restrictive and produce economically significant fitting errors. These problems are pointedly discussed by Cochrane and Piazzesi (2008), who consider a data set fitted by Gurkaynak, Sack, and Wright (2006) to a 6-factor Svensson model. Cochrane and Piazzesi (2008)

[9]The Vuong (1989) test is a standard method to evaluate nonnested models and asymptotically has a standard normal distribution under the null of equal expected likelihoods.

comment on the impossibility of using the full cross section of these data: "Regressions using 15 maturities on the right hand side, generated from a six-factor model, are obviously hopeless" (p. 13). They also remark that a low-dimensional factor structure is often presumed to be sufficient for many applications, but "excess return forecasts imply multiple differences of the underlying price data ... so small amounts of smoothing have the potential to lose a lot of information in forecasting exercises" (p. 12).

To avoid the loss of information via dimension reduction, a common industry practice is to assume piecewise step functions for the forward curve. While this approach allows as many degrees of freedom as the number of observations, it produces unsmooth step functions that can give rise to arbitrage opportunities and induce distortions in forward-rate dynamics.

The dimension-invariant cascade DTSM solves these problems, permitting smooth, arbitrage-free yield-curve stripping that fits the observed data almost perfectly with a high number of factors. Since the forward rate is given by equation (18) under the model, the basis functions for the forward-rate curve are the response functions $a_j(\tau)$, $j=1\ldots n$. The components of the state vector, $x_{j,t}$, act as time-varying weights for the basis functions, and the rich state space provides the flexibility to closely match virtually all observed term structure shapes.

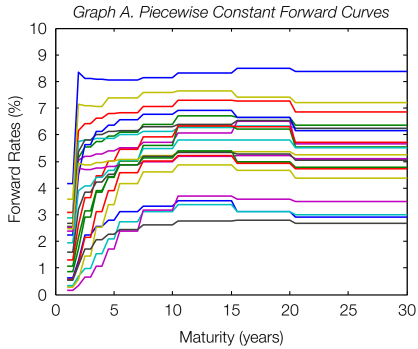
In Graph A of Figure 2, we illustrate the common industry approach, which assumes a piecewise constant step function for the forward rate and backs out the levels of the steps sequentially from low to high maturities. By design, the piecewise constant method matches observed rates exactly. However, the discontinuities in the forward curves indicate potentially large mispricing for any maturities not explicitly used in the construction of the forward curves. These discontinuities can induce instabilities when used as inputs to the forward-rate model of Heath et al. (1992).

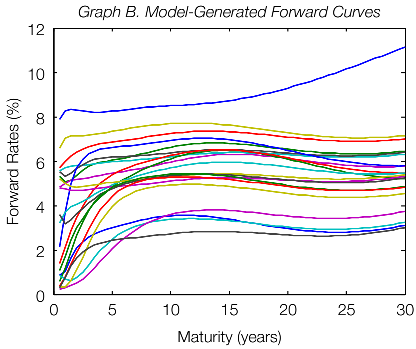
In Graph B of Figure 2, we plot the stripped forward rates $f(\hat{\mathbf{X}}_t,\tau)$ from the cascade DTSM, where $\hat{\mathbf{X}}_t$ denotes the filtered state vector at selected dates.

FIGURE 2

Term Structure of Forward Rates Stripped from LIBOR and Swap Rates

Figure 2 illustrates the term structure of forward rates at different dates, generated from the piecewise constant assumption (Graph A) and the estimated 10-factor cascade (Graph B).





The model provides a good starting point for interest-rate option pricing by generating smooth, dynamically consistent forward-rate curves that match to near perfection the cross section of observed rates.

B. Interest-Rate Cross-Correlations

Low-dimensional DTSMs typically imply counterfactually high cross-correlations between changes in nonoverlapping forward rates (Dai and Singleton (2003)). Intuitively, a low-dimensional model captures the systematic, common movements in the interest-rate term structure. Interest-rate fair values built purely from these common movements show high cross-correlation. A high-dimensional cascade has the potential to generate interest-rate fair values that match the cross-correlations observed in the data.

Measuring cross-correlations between nonoverlapping forward rates would require us to first strip the swap rates. The cross-correlation estimates would then depend on the selected stripping method and its basis functional forms. To avoid such sensitivities, we instead examine the cross-correlation between observed (LIBOR and swap) rates.

Panel A of Table 4 reports the correlation between weekly changes in the 6-month LIBOR and weekly changes in swap rates of different maturities. The first row reports historical correlations and other rows report the correlations produced

TABLE 4

Correlations of Weekly Changes in 6-Month LIBOR and Swap Rates

Table 4 reports correlation estimates between weekly changes in the 6-month LIBOR and weekly changes in swap rates of different maturities. The first row of each panel reports the correlation estimates from the historical data, and subsequent rows report correlation estimates of corresponding fair values produced from cascade models with $n=2$ to $n=10$ frequencies. Panel A of Table 4 reports the full-sample estimates from Jan. 4, 1995, to Oct. 28, 2015, and the model values are based on model parameters estimated using the full sample of data. Panel B reports the out-of-sample estimates from Jan. 7, 1998, to Oct. 28, 2015. Each year, the out-of-sample interest-rate fair values are generated from model parameters estimated using data up to the previous year.

	Maturity (years)								
	2	3	4	5	7	10	15	20	30
Panel A. Full-Sample Estimates									
Data	0.59	0.53	0.49	0.46	0.43	0.39	0.34	0.31	0.29
Cascade									
$n=2$	0.95	0.91	0.86	0.82	0.77	0.72	0.68	0.66	0.65
$n=3$	0.85	0.80	0.76	0.73	0.67	0.61	0.56	0.53	0.51
$n=4$	0.70	0.61	0.57	0.53	0.48	0.44	0.39	0.37	0.34
$n=5$	0.63	0.55	0.51	0.48	0.44	0.40	0.36	0.34	0.31
$n=6$	0.63	0.55	0.51	0.48	0.44	0.40	0.36	0.34	0.31
$n=7$	0.61	0.54	0.50	0.47	0.44	0.40	0.35	0.33	0.30
$n=8$	0.60	0.53	0.50	0.47	0.44	0.40	0.35	0.33	0.30
$n=9$	0.60	0.53	0.50	0.47	0.44	0.40	0.35	0.32	0.30
$n=10$	0.60	0.53	0.50	0.47	0.44	0.40	0.35	0.32	0.30
Panel B. Out-of-Sample Estimates									
Data	0.56	0.50	0.46	0.43	0.40	0.36	0.31	0.28	0.26
Cascade									
$n=2$	0.90	0.82	0.76	0.71	0.66	0.61	0.58	0.57	0.56
$n=3$	0.75	0.69	0.64	0.60	0.54	0.48	0.43	0.40	0.38
$n=4$	0.63	0.56	0.51	0.47	0.42	0.38	0.33	0.31	0.27
$n=5$	0.62	0.54	0.49	0.45	0.41	0.37	0.32	0.30	0.27
$n=6$	0.60	0.52	0.48	0.45	0.41	0.37	0.32	0.29	0.26
$n=7$	0.60	0.51	0.47	0.45	0.41	0.36	0.32	0.29	0.26
$n=8$	0.59	0.51	0.47	0.45	0.41	0.36	0.32	0.29	0.26
$n=9$	0.59	0.51	0.47	0.44	0.41	0.36	0.32	0.29	0.26
$n=10$	0.59	0.51	0.47	0.44	0.41	0.36	0.32	0.29	0.26

by models with increasing number of factors. Consistent with Dai and Singleton (2003), the fitted values from low-dimensional models have much higher cross-correlation than are observed in historical data. As state variables are added to the model, the cross-correlations better match the historical data.

To gauge the stability of the model estimation approach, we conduct an out-of-sample exercise in which we reestimate model parameters once a year starting from 1998 by using data up to the year before. Then, at each date from Jan. 7, 1998, we generate out-of-sample interest-rate fair values based on model parameters estimated before that year. Panel B of Table 4 compares the correlation estimates from the historical data during this out-of-sample period from Jan. 7, 1998, to Oct. 28, 2015, with the corresponding estimates on the out-of-sample fair values produced by models with increasing number of factors. The results are very similar to the full-sample estimates reported in Panel A, highlighting the stability of our estimation approach and model structure. Overall, Table 4 shows that a high number of factors are needed to match the cross-correlations observed between weekly changes of interest rates.

C. Interest-Rate Forecasting

Several studies find that low-dimensional DTSMs fare worse than a simple random walk in forecasting interest-rate movements out of sample (e.g., Duffee (2002), Ang and Piazzesi (2003), and Bali et al. (2009)). We hypothesize that two limitations of typical DTSMs inhibit their performance. First, traditional 3-factor models do not fit observed interest rates closely in sample. Because the fitted yield curve is the starting point of model-based forecasts, fitting errors contaminate out-of-sample forecasts. Second, general 3-factor models involve over 20 parameters, many of which are poorly identified. Therefore, no matter how well a traditional 3-factor model fits the data in sample, its performance should deteriorate out of sample due to parameter instability.

Dimension-invariant DTSMs allow us to address both issues. High-dimensional versions of the cascade fit the current term structure to near perfection, as Section IV shows. The parsimony of the baseline cascade should imply that performance experiences little out-of-sample deterioration.

We compare the forecasting performance of a first-order autoregression (AR(1)) and cascades with increasing number of factors. To demonstrate the differential effects of in-sample fit and out-of-sample stability on forecasting performance, we calculate in-sample prediction errors for each method using full-sample parameter estimates and then report the out-of-sample performance of each method with rolling estimation.

We assess forecasting performance by a predictive variation (PV) measure. The PV of the jth interest-rate series over a horizon h is defined as

$$\mathrm{PV}_{j,h} = 1 - \frac{\sum_{t=1}^{N-h} \left(\hat{y}_{j,t+h} - y_{j,t+h}\right)^2}{\sum_{t=1}^{N-h} \left(y_{j,t} - y_{j,t+h}\right)^2},$$

where $y_{j,t+h}$ denotes the jth rate at time $t+h$ and $\hat{y}_{j,t+h}$ is the time-t forecast. The difference $(y_{j,t} - y_{j,t+h})$ represents the prediction error produced by the random-walk assumption, under which the time-t forecast is simply the time-t observation. Therefore, by design, the predictive variation $\mathrm{PV}_{j,h}$ is positive when the strategy outperforms the random-walk hypothesis and negative otherwise.

1. In-Sample Performance

Table 5 reports in-sample predictive variation at the 1-week (Panel A) and 4-week (Panel B) horizons. The first row of each panel reports the predictive variation from the AR(1) regression, whereas the remaining rows report predictive variation from cascades with 2–10 frequencies. Both panels show similar behaviors. In sample, the AR(1) generates the best performance in all cases. Because the AR(1) strategy nests the random walk, all predictive variations are positive by design. For the 6-month LIBOR series, the predictive variation is 14.25% at the 1-week horizon and 28.76% at the 4-week horizon. For the swap rate series, the predictive variation is smaller and close to 3% at the 1-week horizon and 10% at the 4-week horizon.

TABLE 5

In-Sample Predictive Variation from AR(1) and Cascade Models

Table 5 reports in-sample predictive variation (in percentage points) on each interest-rate series at the 1-week (Panel A) and 4-week (Panel B) forecasting horizons. Within each panel, the first row provides the performance from a first-order autoregressive regression (AR(1)) and the remaining rows report the performance from the cascade model with $n=2$ to $n=10$ frequencies. The predictive variation is defined as 1 minus the ratio of mean-squared predicting error to mean-squared interest-rate change. The forecasts are based on full-sample model parameter estimates from Jan. 1995 to Oct. 2015.

	Maturity (years)									
	0.5	2	3	4	5	7	10	15	20	30
Panel A. 1-Week Forecasting Horizon										
AR(1)	14.25	2.81	2.88	3.15	2.59	3.14	3.81	2.62	2.45	2.69
Cascade										
$n=2$	−707.65	−181.92	−125.10	−79.90	−47.46	−29.95	−37.77	−47.80	−59.45	−118.93
$n=3$	−84.43	−93.34	−11.90	−7.65	−12.95	−20.45	−18.95	−10.65	−19.71	−66.83
$n=4$	7.24	−30.02	−0.77	−1.45	−1.39	0.55	−4.65	−7.14	−4.00	0.04
$n=5$	8.90	−0.66	−1.40	−0.82	−0.56	0.83	1.00	−1.70	−3.36	1.07
$n=6$	10.44	−0.78	−1.76	−1.12	−0.76	0.75	0.93	−1.73	−3.42	1.02
$n=7$	9.02	−0.21	0.05	0.15	−0.44	−0.11	0.48	0.13	0.21	0.10
$n=8$	8.87	0.04	0.14	0.31	−0.31	−0.04	0.51	0.16	0.38	0.10
$n=9$	8.78	0.00	−0.01	0.29	−0.31	−0.07	0.47	0.15	0.38	0.07
$n=10$	8.87	−0.07	−0.10	0.25	−0.33	−0.10	0.45	0.14	0.38	0.06
Panel B. 4-Week Forecasting Horizon										
AR(1)	28.76	10.13	9.65	9.67	9.66	10.47	11.28	9.87	9.53	10.14
Cascade										
$n=2$	−64.55	−61.36	−33.18	−14.67	−3.94	0.03	−7.31	−18.00	−23.66	−34.44
$n=3$	−47.33	−41.13	−7.39	0.59	2.32	0.91	−3.04	−7.87	−12.77	−20.59
$n=4$	7.57	−18.15	−2.71	−0.15	0.30	−0.28	−3.59	−5.80	−4.17	2.73
$n=5$	20.52	−1.54	−2.53	−2.10	−1.26	0.38	0.48	−1.94	−2.90	1.65
$n=6$	21.79	−2.27	−3.30	−2.75	−1.76	0.07	0.29	−2.07	−3.05	1.48
$n=7$	19.95	−0.33	−0.67	−0.76	−0.79	−0.30	0.41	0.05	0.29	0.51
$n=8$	19.69	0.35	−0.23	−0.34	−0.45	−0.06	0.60	0.19	0.53	0.54
$n=9$	18.81	0.14	−0.51	−0.44	−0.50	−0.13	0.54	0.16	0.51	0.50
$n=10$	18.35	−0.09	−0.73	−0.57	−0.58	−0.19	0.48	0.13	0.49	0.46

Low-dimensional cascades (with 2–3 factors) perform much worse than the random walk, especially at short horizons. These results are expected. Low-dimensional models fit the current term structure imperfectly and therefore have the wrong starting point for interest-rate forecasts.

With 4 or more factors, the cascade generates positive predictive variation at both the 1-week and 4-week horizons for 6-month LIBOR. For the swap rates, the predictive variation estimates are close to 0, even for models with high number of factors. These findings suggest that the models, even with perfect fitting, cannot forecast swap rates much more accurately than the random walk.

There are two possible reasons for the lack of predictability on longer-term rates. First, longer-term rates are persistent by nature and difficult to predict, regardless of methods. This is evident from the poor in-sample performance of the AR(1) specification. Second, the cascade predicts future levels of the short rate from current levels of longer-maturity rates but does not consider the reverse channel. As a result, the cascade predicts short rates better than long rates.

2. Out-of-Sample Performance

To evaluate out-of-sample performance, we reestimate the autoregressive coefficients and the cascade model parameters once a year starting from 1998 by using data up to the end of the previous year, and then employ these parameters to generate out-of-sample forecasts over the current year. Table 6 reports the out-of-sample forecasting results in a format similar to Table 5. Although AR(1) is the strongest in-sample performer, its predictive power deteriorates dramatically out of sample. Specifically, the AR(1) models produce worse forecasts than the

TABLE 6

Out-of-Sample Predictive Variation from AR(1) and Cascade Models

Table 6 reports out-of-sample predictive variation (in percentage points) on each interest-rate series at the 1-week (Panel A) and 4-week (Panel B) forecasting horizons. Within each panel, the first row reports the performance from a first-order autoregressive regression (AR(1)) and the remaining rows report the performance from the cascade model with $n=2$ to $n=10$ frequencies. The predictive variation is defined as 1 minus the ratio of mean-squared predicting error to mean-squared interest-rate change. At each date starting from Jan. 1998, the forecasts are made using model parameters estimated with data up to the year before. The statistics are computed based on forecasts from Jan. 1998 to Oct. 2015.

	Maturity (years)									
	0.5	2	3	4	5	7	10	15	20	30
Panel A. 1-Week Forecasting Horizon										
AR(1)	−9.51	−4.39	−3.42	−3.05	−2.72	−2.26	−1.89	−1.91	−1.88	−1.75
Cascade										
$n=2$	−652.05	−254.06	−170.51	−107.06	−63.25	−42.44	−50.71	−71.51	−93.35	−254.92
$n=3$	−71.84	−123.32	−13.81	−8.66	−15.42	−23.17	−15.19	−9.72	−24.48	−94.69
$n=4$	1.49	−26.13	−0.89	−6.75	−5.88	−0.09	−4.29	−11.13	−2.82	−7.85
$n=5$	3.87	−1.70	−1.84	−1.72	−1.05	0.21	0.89	−1.11	−3.70	0.10
$n=6$	11.03	−1.89	−1.20	−1.16	−1.18	−0.86	−0.23	−0.88	−1.86	−0.17
$n=7$	10.18	−0.63	0.17	−0.20	−0.78	−0.21	0.28	−0.20	−0.19	0.15
$n=8$	10.59	−0.66	0.08	−0.22	−0.74	−0.06	0.34	−0.29	0.01	0.21
$n=9$	10.97	−0.79	−0.01	−0.28	−0.74	−0.20	0.14	−0.13	0.15	0.08
$n=10$	11.12	−0.80	−0.06	−0.26	−0.71	−0.23	0.13	−0.11	0.18	0.07
Panel B. 4-Week Forecasting Horizon										
AR(1)	−21.93	−18.65	−14.45	−12.54	−11.04	−9.36	−8.44	−8.43	−8.54	−7.87
Cascade										
$n=2$	−41.22	−91.52	−54.57	−28.53	−12.26	−3.65	−8.47	−20.93	−30.25	−71.52
$n=3$	−50.92	−55.09	−8.90	1.71	4.43	3.98	1.44	−5.85	−14.56	−33.71
$n=4$	0.76	−13.14	−2.20	−2.82	−2.17	−0.53	−2.39	−5.51	−3.15	−0.26
$n=5$	13.55	−2.21	−3.38	−3.43	−2.28	−0.41	−0.23	−2.08	−3.46	0.12
$n=6$	22.59	−2.55	−3.13	−3.39	−2.82	−1.78	−1.45	−1.95	−2.46	−0.38
$n=7$	22.35	−0.83	−1.21	−1.70	−1.64	−0.91	−0.21	−0.66	−0.73	−0.03
$n=8$	21.47	−1.04	−1.37	−1.73	−1.54	−0.64	0.04	−0.54	−0.34	0.24
$n=9$	21.69	−1.66	−1.81	−2.06	−1.76	−0.91	−0.28	−0.53	−0.31	0.01
$n=10$	20.96	−1.77	−1.93	−2.08	−1.75	−0.93	−0.28	−0.50	−0.27	0.02

random walk across all maturities and forecasting horizons. These findings indicate parameter instability and overfitting.

By contrast, due to their parsimony, the cascade models perform similarly in and out of sample. For specifications with 6 or more factors, the out-of-sample predictive variation on the 6-month LIBOR is positive at about 11% at the 1-week horizon and is close to 22% at the 4-week horizon. The predictive variation on swap rates are as small out of sample as they are in sample.

D. Principal Component Forecasting

Principal component analysis has a long history in the term structure literature (Litterman and Scheinkman (1991), Knez, Litterman, and Scheinkman (1994), and Joslin et al. (2011)). We now analyze the ability of different methods to forecast the principal components of interest rates.[10] At each date, we estimate the covariance matrix of the weekly changes of the interest-rate series using data up to the previous year, and we form interest-rate portfolios mimicking the principal components using the eigenvectors of the estimated covariance matrix. The forecasts on each portfolio are then computed from the forecasts on each interest-rate series.

Table 7 reports the out-of-sample performance of each strategy for each mimicking portfolio. The first principal component corresponds to the largest eigenvalue and represents approximately an average of all interest rate series, whereas the 10th principal component corresponds to the smallest eigenvalue. The AR(1) strategy has predictive variation estimates that are generally more negative for the portfolios with larger eigenvalues. At the 1-week horizon, the estimates are negative for the first 6 principal components and slightly positive for the remaining 4. At the 4-week horizon, the estimates are negative for the first 9 principal components and positive only for the 10th component. These results are consistent with the property that lower principal components are more persistent and higher components are more transient. The AR(1) strategy cannot beat the random walk on persistent series but can beat it on highly transient components.

Low-dimensional cascades, which we expect to capture only the first few components of interest-rate movements, provide miserable predictions of higher-order principal components. Table 7 shows that 2- and 3-factor cascade models generate highly negative predictive variation estimates for all principal components and perform especially poorly for higher-order components. The analysis highlights a limitation of low-dimensional DTSMs: Even when they do match broad features of interest rate behaviors, they miss grossly on fitting higher-order principal components that relate to differences across interest-rate series.

Cascades with 4–6 factors generate slightly positive predictive variation for some low-frequency components, but the predictive variation of higher-order components remains highly negative. By contrast, the preferred cascade specifications with 9–10 factors generate positive predictive variation for the majority of principal components.

To provide a broader perspective on our forecasting results, consider the simple idea of exploiting information in the cross section of interest rates by using a

[10] We thank a referee for this suggestion.

TABLE 7

Out-of-Sample Predictive Variation of Principal Components

Table 7 reports out-of-sample predictive variation (in percentage points) of each principal component mimicking portfolio at the 1-week (Panel A) and 4-week (Panel B) forecasting horizons. The first principal component corresponds to the eigenvector with the largest eigenvalue, and the mimicking portfolios from 1 to 10 corresponds to eigenvalues of declining order. Within each panel, the first row reports the performance from a first-order autoregressive regression (AR(1)) and the remaining rows report the performance from the cascade model with $n=2$ to $n=10$ frequencies. The predictive variation is defined as 1 minus the ratio of mean-squared predicting error to mean-squared interest-rate change. At each date starting from Jan. 1998, the forecasts are made using model parameters estimated with data up to the year before. The statistics are computed based on forecasts from Jan. 1998 to Oct. 2015.

	Principal Component									
	1	2	3	4	5	6	7	8	9	10
Panel A. 1-Week Forecasting Horizon										
AR(1)	−2.46	−6.30	−6.64	−0.19	−2.80	−0.38	1.48	0.14	0.22	2.47
Cascade										
$n=2$	−3.22	−211.15	−2,468.13	−4,586.27	−3,350.18	−3,748.02	−532.96	−386.32	−400.53	−1,129.12
$n=3$	−0.53	−3.48	−161.50	−2773.90	−2142.20	−2551.08	−364.24	−388.81	−533.72	−775.63
$n=4$	−0.30	1.19	−9.21	−25.39	−2816.15	−438.84	−114.63	−209.64	−322.74	−375.67
$n=5$	−0.14	1.48	−4.28	−5.35	−18.40	−244.26	−203.79	−140.29	−90.79	−141.27
$n=6$	−0.42	3.25	4.62	0.45	−13.57	−199.89	−179.43	−121.93	−52.43	−98.51
$n=7$	−0.12	3.64	3.60	2.37	6.37	−16.83	−13.33	−15.98	3.52	−9.61
$n=8$	−0.13	3.64	4.62	2.41	7.81	−13.81	−4.45	−5.89	11.80	−6.55
$n=9$	−0.20	3.30	6.00	2.53	9.19	−8.11	1.04	2.69	9.33	5.30
$n=10$	−0.21	3.18	6.46	2.57	9.33	−7.44	4.25	5.23	10.79	6.85
Panel B. 4-Week Forecasting Horizon										
AR(1)	−10.47	−20.99	−21.31	−2.20	−18.68	−7.01	−1.18	−5.31	−14.94	2.46
Cascade										
$n=2$	−3.80	−64.68	−389.64	−1,232.84	−1,145.99	−1,475.52	−336.89	−191.92	−234.25	−806.72
$n=3$	−2.73	−4.84	−109.33	−711.27	−679.66	−944.93	−247.20	−202.65	−321.42	−544.81
$n=4$	−1.72	4.97	−19.09	−5.51	−1002.28	−165.36	−59.94	−106.78	−201.82	−246.08
$n=5$	−1.30	6.41	−3.71	−0.92	−3.34	−81.21	−123.19	−59.94	−47.54	−93.26
$n=6$	−1.97	9.71	13.26	0.21	−4.10	−78.72	−99.66	−46.51	−23.45	−62.76
$n=7$	−0.75	10.73	12.14	−0.70	2.72	−9.40	1.29	5.70	6.06	0.99
$n=8$	−0.68	10.30	12.54	−0.52	3.09	−9.33	6.86	8.46	10.78	3.89
$n=9$	−0.99	9.49	15.48	−0.27	4.22	−8.04	8.89	13.32	6.67	11.36
$n=10$	−0.99	8.75	15.50	−0.15	4.32	−8.71	10.73	13.70	6.95	12.41

general first-order vector autoregression (VAR(1)). While this approach would seem to have natural appeal, we can easily see flaws. Overfitting and parameter instability are problematic even when using simple AR(1) regressions, which restrict the VAR by setting the off-diagonal elements of the VAR transition matrix equal to 0. Obviously, out-of-sample instability would be an even larger problem in a general VAR(1).[11] The cascade model obtains the important advantages of a full VAR(1) system on the forward rates, but the entire VAR(1) system is specified by only 5 parameters, regardless of the number of factors, eliminating the difficulties associated with overfitting. The ability to parsimoniously incorporate information from the entire term structure is a necessity for generating good forecasts and drives our results.

The base version of the cascade DTSM accomplishes several objectives. First, it provides smooth, arbitrage-free yield-curve stripping while maintaining a close fit to the original data. The approach avoids the potential loss of information that occurs when high-dimensional data are compressed into a lower-dimensional factor structure. Second, higher-dimensional cascades match observed interest-rate cross-correlations more closely. Third, higher-dimensional cascades

[11]This intuition is consistent with Ang and Piazzesi (2003), who show that an unconstrained VAR(1) performs poorly in forecasting relative to the simple random-walk hypothesis.

offer improved short-term forecasts of LIBOR interest rates relative to low-dimensional specifications. The Internet Appendix verifies the robustness of our dimension-invariant scaling specifications and considers various extensions, including cascading stochastic-volatility specifications for interest-rate option pricing and time-varying market price of risk specifications to match the evidence on bond-return predictions (Cochrane and Piazzesi (2005), (2008)).

VI. Conclusion

We have developed a tractable class of dynamic term structure models that accommodate arbitrarily many factors of heterogeneous durations. The construction of the short rate is based on a cascade of diffusions, in which each state variable mean reverts toward the preceding one, capturing the idea that any economic fluctuation occurs within the context of lower-frequency cycles. The baseline version of the model makes parsimonious assumptions on the distribution of risk, adjustment speeds, and risk premia and generates a dimension-invariant structure specified by only 5 parameters, regardless of the number of factors. The term structure of zero-coupon bond yields converges to a well-defined stochastic limit as the number of high-frequency factors goes to infinity. The convergence result builds a bridge between standard low-dimensional term structure models and infinite-dimensional specifications that have been proposed in prior literature. We have also provided conditions under which arbitrage opportunities are ruled out and portfolio Sharpe ratios are bounded in the limiting economy, which offers guidance for future work.

The cascade DTSM offers several empirical benefits. A high-dimensional, 10-factor version fits a broad cross section of LIBOR and swap rates almost perfectly, with errors lower than 1 bp. The fitted model generates smooth forward curves, accurately matches the low cross-correlations observed between movements of interest rates at different maturities, and generates superior out-of-sample interest-rate forecasts relative to lower-dimensional DTSMs and autoregressive specifications. We have provided empirical tests of the primary scaling assumptions in the benchmark model and have shown how to incorporate time-varying risk premia in the cascade framework, producing results consistent with Cochrane and Piazzesi (2005). Finally, we have shown how to extend the base model to include a cascade volatility specification.

The paper opens a number of directions for future research. First, it would be interesting to use a high-dimensional cascade as the starting point for pricing interest-rate options, complemented with an additional volatility cascade of the type discussed in the Internet Appendix. The literature (e.g., Dai and Singleton (2000)) generally finds that low-dimensional DTSMs do a poor job of pricing interest-rate derivatives. An important difficulty is that the payoffs of interest-rate options are written on actually observed interest rates rather than on model values. A low-dimensional DTSM provides a relatively poor fit of observed interest rates and requires that the asset pricer accommodate deviations between observations and model valuation. Kennedy (1994) and Goldstein (2000), among others, propose using infinite-dimensional random fields in order to fully accommodate the idiosyncratic movements of each interest-rate series required to price interest-rate options. Our cascade dynamics build a natural bridge between finite-dimensional and infinite-dimensional models while matching observed interest rates, providing a promising foundation for pricing interest-rate options.

Second, the additive cascade that we develop as a building block of a DTSM seems attractive as a general time-series model for other applications in finance, economics, and statistics. The additive cascade accommodates an arbitrary number of components with a small number of parameters and permits straightforward filtering. It thus provides a convenient, tractable framework for decomposing series into fluctuations of multiple frequencies. We leave empirical and theoretical investigations of these extensions for future research.

References

Adrian, T.; R. Crump; and E. Moench. "Pricing the Term Structure with Linear Regressions." *Journal of Financial Economics*, 10 (2013), 110–138.

Ait-Sahalia, Y., and R. Kimmel. "Estimating Affine Multifactor Term Structure Models Using Closed-Form Likelihood Expansions." *Journal of Financial Economics*, 98 (2010), 113–144.

Ang, A.; S. Dong; and M. Piazzesi. "No-Arbitrage Taylor Rules." Working Paper, Columbia University and University of Chicago (2007).

Ang, A., and M. Piazzesi. "A No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables." *Journal of Monetary Economics*, 50 (2003), 745–787.

Ang, A.; M. Piazzesi; and M. Wei. "What Does the Yield Curve Tell Us about GDP Growth?" *Journal of Econometrics*, 131 (2004), 359–403.

Backus, D.; S. Foresi; A. Mozumdar; and L. Wu. "Predictable Changes in Yields and Forward Rates." *Journal of Financial Economics*, 59 (2001), 281–311.

Balduzzi, P.; G. Bertola; and S. Foresi. "A Model of Target Changes and the Term Structure of Interest Rates." *Journal of Monetary Economics*, 39 (1997), 223–249.

Balduzzi, P.; S. Das; and S. Foresi. "The Central Tendency: A Second Factor in Bond Yields." *Review of Economics and Statistics*, 80 (1998), 62–72.

Balduzzi, P.; S. Das; S. Foresi; and R. Sundaram. "A Simple Approach to Three-Factor Affine Term Structure Models." *Journal of Fixed Income*, 6 (1996), 43–53.

Bali, T.; M. Heidari; and L. Wu. "Predictability of Interest Rates and Interest-Rate Portfolios." *Journal of Business and Economic Statistics*, 27 (2009), 517–527.

Bekaert, G.; S. Cho; and A. Moreno. "New Keynesian Macroeconomics and the Term Structure." *Journal of Money, Credit, and Banking*, 42 (2005), 33–62.

Brennan, M., and E. S. Schwartz. "A Continuous Time Approach to the Pricing of Bonds." *Journal of Banking and Finance*, 3 (1979), 133–155.

Calvet, L. E.; M. Fearnley; A. J. Fisher; and M. Leippold. "What's Beneath the Surface? Option Pricing with Multifrequency Latent States." *Journal of Econometrics*, 187 (2015), 498–511.

Calvet, L. E., and A. J. Fisher. "Forecasting Multifractal Volatility." *Journal of Econometrics*, 105 (2001), 27–58.

Calvet, L. E., and A. J. Fisher. "How to Forecast Long-Run Volatility: Regime-Switching and the Estimation of Multifractal Processes." *Journal of Financial Econometrics*, 2 (2004), 49–83.

Calvet, L. E., and A. J. Fisher. "Multifrequency News and Stock Returns." *Journal of Financial Economics*, 86 (2007), 178–212.

Calvet, L. E., and A. J. Fisher. *Multifractal Volatility: Theory, Forecasting, and Pricing*. Burlington, MA: Academic Press (2008).

Calvet, L. E., and A. J. Fisher. "Extreme Risk and Fractal Regularity in Finance." *Contemporary Mathematics*, 601 (2013), 65–94.

Chambers, D. R.; W. T. Carleton; and D. W. Waldman. "A New Approach to Estimation of the Term Structure of Interest Rates." *Journal of Financial and Quantitative Analysis*, 19 (1984), 233–252.

Christensen, J. H. E.; F. X. Diebold; and G. D. Rudebusch. "The Affine Arbitrage-Free Class of Nelson–Siegel Term Structure Models." *Journal of Econometrics*, 164 (2011), 4–20.

Chua, C. T.; D. Foster; K. Ramaswamy; and R. Stine. "A Dynamic Model for the Forward Curve." *Review of Financial Studies*, 21 (2008), 265–310.

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

Cochrane, J. H., and M. Piazzesi. "Decomposing the Yield Curve." Working Paper, University of Chicago and Stanford University (2008).
Collin-Dufresne, P., and R. S. Goldstein. "Generalizing the Affine Framework to HJM and Random Field Models." Working Paper, Washington University in St. Louis (2003).
Collin-Dufresne, P.; R. S. Goldstein; and C. Jones. "Identification of Maximal Affine Term Structure Models." *Journal of Finance*, 63 (2008), 743–795.
Constantinides, G. M. "A Theory of the Nominal Term Structure of Interest Rates." *Review of Financial Studies*, 5 (1992), 531–552.
Cox, J. C.; J. E. Ingersoll; and S. R. Ross. "A Theory of the Term Structure of Interest Rates." *Econometrica*, 53 (1985), 385–408.
Dai, Q., and K. Singleton. "Specification Analysis of Affine Term Structure Models." *Journal of Finance*, 55 (2000), 1943–1978.
Dai, Q., and K. Singleton. "Expectation Puzzles, Time-Varying Risk Premia, and Affine Models of the Term Structure." *Journal of Financial Economics*, 63 (2002), 415–441.
Dai, Q., and K. Singleton. "Term Structure Dynamics in Theory and Reality." *Review of Financial Studies*, 16 (2003), 631–678.
Dalio, R. "Economic Principles." Working Paper, Bridgewater Associates (2012).
Diebold, F. X.; G. D. Rudebusch; and S. B. Aruba. "The Macroeconomy and the Yield Curve: A Dynamic Latent Factor Approach." *Journal of Econometrics*, 131 (2006), 309–338.
Duffee, G. R. "Term Premia and Interest Rate Forecasts in Affine Models." *Journal of Finance*, 57 (2002), 405–443.
Duffee, G. R. "Sharpe Ratios in Term Structure Models." Working Paper, Johns Hopkins University (2010).
Duffee, G. R. "Forecasting with the Term Structure: The Role of No-Arbitrage Restrictions." Working Paper, Johns Hopkins University (2011).
Duffee, G. R., and R. H. Stanton. "Evidence on Simulation Inference for Near Unit-Root Processes with Implications for Term Structure Estimation." *Journal of Financial Econometrics*, 6 (2008), 108–142.
Duffie, D.; D. Filipović; and W. Schachermayer. "Affine Processes and Applications in Finance." *Annals of Applied Probability*, 13 (2003), 984–1053.
Duffie, D., and R. Kan. "A Yield-Factor Model of Interest Rates." *Mathematical Finance*, 6 (1996), 379–406.
Duffie, D.; J. Pan; and K. Singleton. "Transform Analysis and Asset Pricing for Affine Jump Diffusions." *Econometrica*, 68 (2000), 1343–1376.
Fama, E. F., and R. R. Bliss. "The Information in Long-Maturity Forward Rates." *American Economic Review*, 77 (1987), 680–692.
Filipović, D. "A Note on the Nelson–Siegel Family." *Mathematical Finance*, 9 (1999), 349–359.
Fisher, M., and C. Gilles. "Estimating Exponential-Affine Models of the Term Structure." Working Paper, Federal Reserve Bank of Atlanta (1996).
Gallmeyer, M.; B. Hollifield; F. Palomino; and S. Zin. "Arbitrage-Free Bond Pricing with Dynamic Macroeconomic Models." *Journal of Monetary Economics*, 52 (2005), 921–950.
Gallmeyer, M.; B. Hollifield; and S. Zin. "Taylor Rules, McCallum Rules, and the Term Structure of Interest Rates." *Journal of Monetary Economics*, 52 (2005), 921–950.
Goldstein, R. "The Term Structure of Interest Rates as a Random Field." *Review of Financial Studies*, 13 (2000), 365–384.
Grasselli, M., and C. Tebaldi. "Solvable Affine Term Structure Models." *Mathematical Finance*, 18 (2008), 135–153.
Gurkaynak, R. S.; B. Sack; and J. H. Wright. "The U.S. Treasury Yield Curve: 1961 to the Present." Finance and Economics Discussion Series 2006-28, Federal Reserve Board (2006).
Hamilton, J., and J. C. Wu. "Identification and Estimation of Gaussian Affine Term Structure Models." *Journal of Econometrics*, 168 (2012), 315–331.
Han, B. "Stochastic Volatilities and Correlations of Bond Yields." *Journal of Finance*, 62 (2007), 1491–1524.
Heath, D.; R. Jarrow; and A. Morton. "Bond Pricing and the Term Structure of Interest Rates: A New Technology for Contingent Claims Valuation." *Econometrica*, 60 (1992), 77–105.
Heidari, M., and L. Wu. "Market Anticipation of Fed Policy Changes and the Term Structure of Interest Rates." *Review of Finance*, 14 (2010), 313–342.
Hördahl, P.; O. Tristanoi; and D. Vestin. "A Joint Econometric Model of Macroeconomic and Term Structure Dynamics." *Journal of Econometrics*, 131 (2006), 405–444.
Joslin, S.; K. J. Singleton; and H. Zhu. "A New Perspective on Gaussian Dynamic Term Structure Models." *Review of Financial Studies*, 24 (2011), 926–970.

Kalman, R. E. "A New Approach to Linear Filtering and Prediction Problems." *Transactions of the ASME–Journal of Basic Engineering*, 82 (1960), 35–45.

Kennedy, D. "The Term Structure of Interest Rates as a Gaussian Random Field." *Mathematical Finance*, 4 (1994), 247–258.

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

Knez, P. J.; R. Litterman; and J. Scheinkman. "Explorations into Factors Explaining Money Market Returns." *Journal of Finance*, 49 (1994), 1861–1882.

Leippold, M., and L. Wu. "Asset Pricing under the Quadratic Class." *Journal of Financial and Quantitative Analysis*, 37 (2002), 271–295.

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

Litzenberger, R. H., and J. Rolfo. "An International Study of Tax Effects on Government Bonds." *Journal of Finance*, 39 (1984), 1–22.

Longstaff, F. A.; P. Santa-Clara; and E. S. Schwartz. "The Relative Valuation of Caps and Swaptions: Theory and Empirical Evidence." *Journal of Finance*, 56 (2001), 2067–2109.

Longstaff, F. A., and E. S. Schwartz. "Interest Rate Volatility and the Term Structure: A Two-Factor General Equilibrium Model." *Journal of Finance*, 47 (1992), 1259–1282.

Longstaff, F. A., and E. S. Schwartz. "Valuing American Options by Simulation: A Simple Least-Squares Approach." *Review of Financial Studies*, 14 (2001), 113–147.

Lu, B., and L. Wu. "Macroeconomic Releases and the Interest Rate Term Structure." *Journal of Monetary Economics*, 56 (2009), 872–884.

McCulloch, J. H. "The Tax-Adjusted Yield Curve." *Journal of Finance*, 30 (1975), 811–830.

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

Pearson, N. D., and T.-S. Sun. "Exploiting the Conditional Density in Estimating the Term Structure: An Application to the Cox, Ingersoll, and Ross Model." *Journal of Finance*, 49 (1994), 1279–1304.

Piazzesi, M. "Bond Yields and the Federal Reserve." *Journal of Political Economy*, 113 (2005), 311–344.

Piazzesi, M. "Affine Term Structure Models." In *Handbook of Financial Econometrics, Vol. 1*, Y. Ait-Sahalia and L. P. Hansen, eds. Amsterdam: North Holland Elsevier (2010), 691–766.

Ronn, E. I. "A New Linear Programming Approach to Bond Portfolio Management." *Journal of Financial and Quantitative Analysis*, 22 (1987), 439–466.

Rudebusch, G. D. "Term Structure Evidence on Interest Rate Smoothing and Monetary Policy Inertia." *Journal of Monetary Economics*, 49 (2002), 1161–1187.

Rudebusch, G. D.; E. T. Swanson; and T. Wu. "The Bond Yield 'Conundrum' from a Macro-Finance Perspective." *Monetary and Economic Studies*, 24 (2006), 83–128.

Rudebusch, G. D., and T. Wu. "A Macro-Finance Model of the Term Structure, Monetary Policy and the Economy." *Economic Journal*, 118 (2008), 906–926.

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

Svensson, L. E. O. "Estimating Forward Interest Rates with the Extended Nelson and Siegel Method." *Quarterly Review, Sveriges Riksbank*, 3 (1995), 13–26.

Vasicek, O. A. "An Equilibrium Characterization of the Term Structure." *Journal of Financial Economics*, 5 (1977), 177–188.

Vuong, Q. "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses." *Econometrica*, 57 (1989), 307–333.

Wan, E. A., and R. van der Merwe. "The Unscented Kalman Filter." In *Kalman Filtering and Neural Networks*, S. Haykin, ed. New York: Wiley & Sons Publishing (2001).