Multifrequency Cascade Interest Rate Dynamics and Dimension-Invariant Term Structures

Laurent E. Calvet
HEC Paris and NBER

Adlai J. Fisher
Sauder School of Business, University of British Columbia

Liuren Wu
Zicklin School of Business, Baruch College

ABSTRACT

We develop a class of dynamic term structure models that accommodates arbitrarily many interest-rate factors with very few parameters. The model builds on a cascade interest-rate dynamics that naturally ranks the factors by their rates of mean reversion, with each revolving around the next lower-frequency factor. The model further achieves dimension invariance by parameterizing the distributions of coefficients of the different frequency components. The net result is a class of term structure models with merely five parameters regardless of the number of factors. Using a panel of 15 LIBOR and swap rates, we estimate 15 models with one to 15 factors. The extensive estimation exercise shows that the 15-factor model significantly outperforms the other lower-dimensional specifications. The high-dimensional specification generates root mean squared pricing errors less than one basis point, thus making it an ideal candidate as a basis for forward rate curve stripping. The model also overcomes several known limitations of low-dimensional specifications by matching the observed low cross-correlations between changes in different interest rate series and by generating markedly better out-of-sample forecasting performance on interest rate movements.

JEL Classification: E43, E47, G10, G12, C51.

Keywords: Interest rate dynamics, term structure, cascade dynamics, dimension invariance, power law scaling.

We thank Anna Cieslak, David Backus, Peter Carr, Francis Diebold, Bjorn Flesaker, Michael Gallmeyer, Robert Goldstein, Jeremy Graveline, Burton Hollifield, Kris Jacobs, Dmitry Kreslavskiy, Markus Leippold, Vadim Linetsky, Andrei Lyashenko, Fabio Mercurio, Harvey Stein, David Weinbaum, Yildiray Yildirim, Hao Zhou, and seminar participants at Bloomberg, Cheung Kong Graduate School of Business, Columbia University, FGV/EBAPE, Northwestern University, Syracuse University, University of Chicago, University of Zurich, Ziff Brothers Investments, the Third Risk Management Conference at Mont Tremblant, Québec, the 2010 McGill University Risk Management Conference, the 3rd Triple Crown Conference at Rutgers University, and the Workshop on Financial Econometrics at the Fields Institute, Toronto, for discussions and comments. We welcome comments, including references we have inadvertently overlooked.
1. Introduction

The interest rate term structure experiences shocks of all frequencies; yet researchers often focus on a subset of the frequencies based on a particular purpose. For example, macro economists often focus on the effects of long-term macroeconomic shocks on the term structure. These studies often find that in the long run, positive shocks to inflation raise the interest rate level across all maturities, whereas positive shocks to real output growth raise short-term rates more than long-term rates. At intermediate horizons, central banks have historically implemented monetary policy through their influence on a short-term rate. Monetary surprises thereby directly impact the short-end of the yield curve and spread across the whole term structure through their influence on market expectations of future short rate movements (e.g., Balduzzi, Bertola, and Foresi (1997), Woodford (2003), Piazzesi (2005), and Heidari and Wu (2009)). The dynamic term structure model literature also starts with the specification of a short rate dynamics and derives its implication on the whole term structure through no-arbitrage arguments. See, for example, Duffie and Kan (1996) and Duffie, Pan, and Singleton (2000) for general affine specifications and Leippold and Wu (2002) for general quadratic specifications. On the other hand, market microstructure researchers, market makers, and broker dealers often direct their attentions to movements at even higher frequencies, and recognize that supply-demand shocks originate not just from short maturities, but rather from all tradeable securities. Large transactions of a particular fixed-income instrument can significantly move rates at the associated maturities, followed by quick dissipation along the yield curve through hedging practices.

To understand how macroeconomic shocks, monetary policy surprises, and supply-demand imbalances interact with each other to affect the whole interest rate term structure, it is imperative to develop models that include both low and high frequencies. In principle, this task can be handled by the existing dynamic term structure modeling frameworks as they can theoretically accommodate arbitrarily high dimensions. In practice, however, empirical works have mostly been performed on low-dimensional specifications, most commonly three-factor models. The implementation of higher-dimensional models is typically regarded as

1Recent works that link macroeconomic shocks to the interest rate term structure include, among others, Rudebusch (2002), Ang and Piazzesi (2003), Ang, Piazzesi, and Wei (2004), Bekarc, Cho, and Moreno (2005), Gallmeyer, Hollifield, and Zin (2005), Diebold, Rudebusch, and Aruba (2006), Hordahl, Tristaino, and Vestin (2006), Rudebusch, Swanson, and Wu (2006), Gallmeyer, Hollifield, Palomino, and Zin (2005), and Lu and Wu (2009).

2Examples include Balduzzi, Das, Foresi, and Sundaram (1996), Dai and Singleton (2000, 2002), Baekus, Foresi, Mozumdar,
impractical due to the curse of dimensionality. A general three-factor affine model has 20-30 free parameters, and a general three-factor quadratic model has even more parameters. Many of these parameters cannot be estimated with statistical significance. Further increasing the dimension leads to a rapid increase in the number of parameters, making their identification practically infeasible. Another common issue with these hidden-factor specifications is factor rotation, which induces additional identification issues and makes the estimated factors hard to interpret.

In this paper, we overcome these practical limitations by developing a class of extremely parsimonious dynamic term structure models that completely eliminate both the factor rotation and the curse of dimensionality. Our model builds on a cascade interest rate dynamic structure, where a high-frequency component mean-reverts to a lower-frequency component, which mean-reverts to another component of even lower frequency, and so on. The cascade structure naturally ranks the different factors according to their frequency measured by their mean reversion speeds. The ranking completely eliminates factor rotation and makes explicit the economic meanings of the different factors. Building upon the cascade structure, we further achieve dimension invariance in the number of parameters by parameterizing the coefficient distribution of the different frequency components. Specifically, by assuming power law scaling on the distribution of the mean-reversion speeds, constant and identical instantaneous variance rates, and constant and identical market prices of risk from all frequency components, we obtain a class of dynamic term structure models with merely five parameters, regardless of the number of factors. This dimension-invariance feature allows us to estimate low and high-dimensional models with equal ease and accuracy.

To gauge the empirical performance of the models, we collect 13 years of data on six U.S. dollar LIBOR series with maturities from one to 12 months and nine swap rates with maturities from two to 30 years. We estimate 15 models with the number of factors going from one to 15, respectively, and compare their performances in pricing the 15 interest rate series. The 15-factor model significantly outperforms lower-dimensional models. Its root mean squared pricing errors are less than one basis point, an order of magnitude smaller than its three-factor counterpart. The 15-factor model also captures the low observed cross-correlations between interest rate changes of different maturities, and significantly outperforms both

the random walk and an autoregressive specification in predicting future LIBOR movements.

The model’s near-perfect fitting makes it an ideal candidate as a basis for stripping forward rate curves from swap rates and/or coupon bonds. The basis functions from our model are similar to the exponential functional forms in Nelson and Siegel (1987), but our model overcomes inconsistencies of the Nelson-Siegel model (Filipović (1999)). Furthermore, the high-dimensional structure of our model accommodates considerable flexibility in fitting different term structure shapes than the essentially three-factor structure of the Nelson-Siegel model, or even the six-factor structure of Svensson (1994). The model-generated forward rate curves also overcome the discontinuity experienced in the piece-wise constant forward assumption commonly implemented in the industry.

Dai and Singleton (2002) summarize several limitations of low-dimensional term structure models. Our high-dimensional dimension-invariant cascade model makes progresses in overcoming all these limitations. First, low-dimensional term structure models often imply high cross-correlations between interest rate changes of different maturities, but the actual cross-correlation estimates are often much lower. By generating a near exact fitting to the observed rates, our high-dimensional specification overcomes this limitation in matching the observed low cross-correlations between interest rate changes. Second, interest-rate forecasts generated from low-dimensional term structure models often perform worse than a simple random walk assumption (Duffee (2002)). This poor performance comes from three major sources: (1) model out-of-sample stability due to the identification issues on the many model parameters, (2) imperfect fitting as the fitting errors carry over to become forecasting errors especially at short forecasting horizons, and (3) misspecified factor dynamics as low-dimensional models only capture the low-frequency movements with small mean-reversion speeds whereas the predictability of interest-rate changes mainly comes from the high frequency components with high mean reversion speeds. Our dimension-invariant high-dimensional specification resolves all three issues as the dimension-invariance eliminates the identification and out-of-sample model stability issue, and the high-dimensional structure generates near perfect-fitting and captures both high and low-frequency interest-rate movements. Finally, Dai and Singleton (2002), as well as Li and Zhao (2006) and Heidari and Wu (2009), observe the poor option pricing performance of low-dimensional term structure models. These models write option payoffs on model values whereas in reality options are paid off
against observed interest rates. Therefore, it is nearly impossible for a model to price interest rate options well if it cannot even match the observed underlying interest rates perfectly. The current practice in the industry is to take the observed interest rates as given and price options based exclusively on the volatility specification, such as the forward rate models of Ho and Lee (1986), Hull and White (1993), and Heath, Jarrow, and Morton (1992). By providing a near perfect match to the observed LIBOR and swap rates, our high-dimensional dynamic term structure model can be used as a basis for stripping the forward rate curve, which is used as inputs in these option pricing models. Furthermore, we show that for future research our dimension-invariant cascade structure can also be applied to the volatility dynamics for the purpose of option pricing.

In other related literatures, our cascade interest rate structure has its origin in Balduzzi, Das, and Foresi (1998), who develop a two-factor stochastic central tendency model, where the short rate is allowed to mean revert to a stochastic central tendency. Our cascade structure adds to this idea by allowing as many layers of stochastic tendency as the data call for, providing a way to model an arbitrary number of interest rate factors with a natural ranking according to mean reversion speeds. Multiplicative cascades and power-law scaling have also been proposed by Calvet and Fisher (2001, 2004, 2007) in capturing volatility dynamics.

The remainder of the paper is organized as follows. Section 2 develops the theory on cascade interest rate dynamics and dimension-invariant term structure models. Section 3 describes the data, estimation methodology, and the estimation results. Section 4 compares the performance difference between high- and low-dimensional models. Section 5 considers a variety of specification tests of the model assumptions. Section 6 concludes. All proofs are in the Appendix.

2. Cascade Interest Rate Dynamics and Dimension-Invariant Term Structure

We fix a filtered complete probability space \( \{\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{0 \leq t}\} \) that satisfies the usual technical conditions. Let \( P(t, T) \) denote the time-\( t \) value of a zero-coupon bond with one dollar par value and expiry date \( T \) and

\(^3\)The two-factor structure has also been adopted by Hull and White (2006) in pricing interest rate options. Carr and Wu (2007) apply this two-factor structure to the variance rate specification for pricing currency options. More recently, Carr, Gabaix, and Wu (2009) assume a three-factor cascade structure in pricing interest-rate options.
let \( r_t \) denote the instantaneous interest rate defined by continuity:

\[
     r_t \equiv \lim_{T \downarrow t} \frac{-\ln P(t, T)}{T - t}.
\]  

(1)

2.1. Multifrequency cascade interest rate dynamics

We model the dynamics of the instantaneous interest rate \( r_t \) via a multi-frequency cascade structure. Let \( W_t = (W_{1,t}, \ldots, W_{n,t})^\top \) denote a standard Wiener process with independent components, and let \( X_t = (x_{1,t}, \ldots, x_{n,t})^\top \) denote the \( n \) frequency components underlying the interest rate movements. The dynamics of the different frequency components are linked through the following cascade structure,

\[
     dx_{j,t} = \kappa_j (x_{j-1,t} - x_{j,t}) dt + \sigma_{j,t} dW_{j,t}, \quad j = 1, 2, \ldots, n,
\]

(2)

where each frequency component \( x_{j,t} \) mean reverts to the level of a lower, neighboring frequency component \( x_{j-1,t} \), and as such the mean-reverting speeds of the different frequency components are naturally ranked by \( \kappa_1 < \kappa_2 < \cdots < \kappa_n \). We let the lowest frequency be a constant \( x_{0,t} = \theta_r \), representing the long-run mean of all frequency components.

Intuitively, one can think of \( x_{j-1,t} \) as an moving average of \( x_{j,t} \). Starting at the highest frequency \( x_{n,t} \) and performing moving averages recursively, one generates lower and lower frequency components. Since the reciprocal of the mean reversion speeds \( \kappa_j \) has the unit of time (in years), the economic meaning of each frequent component becomes quite explicit.

The choice of the number and distribution of the frequency components can be determined by data and need. From a single time series, the highest identifiable frequency is governed by the observation frequency. For example, if one wants to model daily interest rate series, the highest identifiable frequency becomes daily, with \( \kappa_n \sim 252 \) (assuming 252 business days in a year). One can add increasingly higher frequency components as the data observations become more frequent. As we will show in a later section, the limit of \( n \to \infty \) is well-defined.
To the other end of the frequency spectrum, the lowest identifiable frequency is determined by the sample length. The lowest frequency cannot be over ten years ($\kappa_1 < 0.1$) if we only have ten years of data. One can add increasingly lower frequency components as the length of the time series increases. One often says that over the very long run, regime switches and one can no longer use the same model to model the whole time-series behavior. This is not true, as one can always use a low-frequency component to model the variation of regimes. In the limit, it is reasonable to think that the mean reversion speed of the lowest frequency converges to zero as it is difficult to predict regime changes in the very long run.

We can write the statistical dynamics of the state vector $X_t$ in matrix notation as

$$dX_t = \kappa(\theta - X_t)\,dt + \sqrt{\Sigma}dW_t,$$

(3)

where the mean-reversion speed matrix $\kappa$ has a block-diagonal form, with the diagonal elements $\kappa_{j,j} = \kappa_j$, the off-diagonal elements $\kappa_{j,j-1} = -\kappa_j$, and all other elements being zero. The long-run mean vector $\theta$ has identical elements equal to $\theta_\tau$, and the instantaneous covariance matrix $\Sigma$ is a diagonal matrix with the $j$-th diagonal element given by $\sigma^2_{jj}$.

In general hidden-state models such as affine models of Duffie and Kan (1996) and quadratic models of Leippold and Wu (2002), the mean-reversion matrix and the covariance matrix are left unconstrained, except some technical conditions. Under these general specifications, factors can rotate and many of the parameters are not identifiable. Factor rotation makes the the economic meaning of the factors elusive, and one must perform careful specification analysis as in Dai and Singleton (2000) to resolve the identification issue. The cascade structure in (2), or the block-diagonal form of $\kappa$ in the matrix notation of (3), completely eliminates factor rotation by ranking the factors based on their relative frequencies.

To understand how the different frequency components interact one another, it is useful to consider an alternative representation as a sum of stochastic integrals.

**Proposition 1** Under the cascade dynamics in (2), the $n$-th frequency component $x_{n,t}$ can be written as the
The coefficient $a_j(\tau)$ can be solved as convolution products of exponential probability density functions,

$$a_j(\tau) = (K_j \ast \ldots \ast K_n)(\tau)/\kappa_j, \quad K_j(\tau) = \kappa_j e^{-\kappa_j \tau},$$

(5)

with $\ast$ denoting the convolution operation.

The function $a_j(\tau)$ quantifies the response of $x_{n,t}$ to a unit shock from the $j$-th frequency component at time $t - \tau$. The response to unit shocks from its own (the highest) frequency component $W_{n,s}$ starts at one when $\tau = 0$ (instantaneously) and decays exponentially with the time horizon, $a_n(\tau) = e^{-\kappa_n \tau}$, where the decay speed is governed by the mean-reversion speed of this frequency component $\kappa_n$. The response to shocks from the lower frequency $(n - 1)$ is determined by the convolution of two exponential density functions originated from the two highest frequencies:

$$a_{n-1}(\tau) = (K_{n-1} \ast K_n)(\tau)/\kappa_{n-1} = \frac{\kappa_n}{\kappa_n - \kappa_{n-1}} \left( e^{-\kappa_{n-1} \tau} - e^{-\kappa_n \tau} \right).$$

(6)

In this case, the response at $\tau = 0$ is zero, becomes strictly positive at intermediate horizons, and converges to zero again as $\tau$ approaches to infinity. The hump-shaped response function reaches its maximum at horizon $\bar{\tau}_{n-1} = (\ln \kappa_n - \ln \kappa_{n-1})/(\kappa_n - \kappa_{n-1})$.

In general, the responses to shocks of lower frequencies are obtained through convolutions of exponential densities. These convolutions can be solved in analytical form.

**Proposition 2** For all $j < n$, the response functions $a_j(\tau)$ are hump-shaped and their maximum response horizons are monotonically decreasing with $j$. The functions satisfy the closed-form expressions:

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

(7)
Furthermore,

\[ 0 \leq \sum_{j=1}^{n} a_j(\tau) \leq 1 \quad \text{(8)} \]

for all \( \tau > 0 \)

The hump-shaped response functions are a direct result of the cascade structure. Instantaneously, only the highest frequency shock \( W_{n,\tau} \) enters \( x_{n,\tau} \). Nonetheless, all lower-frequency shocks impact the stochastic trends, and they ultimately affect the level of \( x_{n,\tau} \) at longer horizons. The cascade structure dictates that lower-frequency shocks enter into the higher-frequency component at increasingly longer horizons.

2.2. Dimension-invariant term structures

Building upon the cascade instantaneous interest rate dynamics in (2), we construct a simple, parsimonious, analytically tractable, and dimension-invariant dynamic term structure model, where the number of parameters in the model is invariant to the dimensions of the state vector.

First, to link the frequency components to the interest rate term structure, we set the instantaneous interest rate equal to the highest frequency component,

\[ r_t = x_{n,\tau}. \quad \text{(9)} \]

As a result, the response function \( a_j(\tau) \) in equation (7) captures the responses of the instantaneous interest rate to shocks of all frequencies. Equation (9) represents the most direct and simple way of linking the different frequency components to the interest rate term structure. We consider an alternative quadratic specification \( (r_t = x_{n,\tau}^2) \) in a later section on specification analysis. The alternative quadratic specification is equally parsimonious and dimension-invariant, but is not as analytically tractable as the direct linkage in equation (9).

Second, since the cascade structure ranks the factors according to their frequencies, we achieve dimension invariance by parameterizing the distribution of the coefficients governing the dynamics of the different
frequency components. We start with the distribution of the mean-reversion speeds \( \{ \kappa_j \}^n_{j=1} \), which directly determines the frequency of each component. We parameterize the distribution of the mean reversion speeds via a power law scaling,

\[
\kappa_j = \kappa_r b^{(j-1)}, \quad b > 1, \quad j = 1, 2, \cdots, n,
\]

where we use \( \kappa_r \) to denote the mean reversion speed of the lowest frequency component \( x_{1,t} \) and use the power scaling coefficient \( b > 1 \) to control the spacing between different frequency components. With this assumption, two parameters control the mean reversion speeds of all frequency components, regardless of how many frequency components we incorporate in the model.

Ur empirically verified assumption of power-law scaling in the mean reversion rates of interest rate factors relates to a broad prior literature on power laws in finance. and

Power-law scaling is a common phenomenon observed in many areas of natural science. Such scaling has been advocated by Mandelbrot (1963, 1967) for the tails and autocorrelations of many financial data sets, and appears in the more recent works of Calvet and Fisher (2001, 2004, 2007) as well as Gabaix (2009) for different types of applications. In the econometric literature, the geometric lag model (Koyck (1954)) is often used as a parsimonious way of including many (sometimes infinite) lags of the same variable, where the coefficients across different lags are scaled according to a power law.

In our application, the power-law scaling is also a natural choice, as the spacing in time becomes increasing wider at lower frequencies. As an example, consider a six-factor structure with the highest frequency being one year \( (\kappa_6 = 1) \), the lowest frequency being 32 years \( (\kappa_1 = 1/32) \), and the power scaling coefficient \( b = 2 \). The mean reversion speed sequence becomes \( (1/32, 1/16, \cdots, 1/2, 1) \). The spacing between the lowest two frequencies is 16 years whereas the spacing between the highest two frequencies is one year. This relative frequency spacing is in line with the actual spacing of the interest rate instruments issued in the market. For example, the most commonly quoted maturities for interest-rate swaps are at two, three, five, ten, 15, and 30 years, with increasingly wider time spacing at longer maturities.

Next, we make parametric assumptions on the distribution and dynamics of the instantaneous volatilities

\[4\text{See Calvet and Fisher (2008) for further discussion.}\]
of the different frequency components $\{\sigma_{j,t}\}_{j=1}^{\nu}$. Here, we note that our purpose is to build a dynamic term structure model describing the behavior of interest rates across maturities, and that it is well-documented that the interest-rate volatilities are not spanned by the interest rate term structure. Thus, it is difficult to fully identify the volatility dynamics from the interest rate term structure alone. Accordingly, we make a simplifying assumption on the volatility dynamics for our term structure modeling, assuming that the instantaneous volatilities of all frequency components are constant and identical,

$$\sigma_{j,t} = \sigma_r.$$  \hfill (11)

Lastly, to price bonds of all maturities, we also need to specify the market prices of risk from all frequency components. Past empirical studies have shown that from the interest rate term structure, one can readily identify the risk-neutral dynamics of the interest rates, but often finds it difficult to pin down the market prices of risk (Bali, Heidari, and Wu (2009)). In line with the simplifying assumption on the volatility dynamics, we also make a simplifying assumption on the market prices of risk, assuming that the market prices of risk from all frequency components are constant and identical,

$$\gamma_{j,t} = \gamma_r.$$  \hfill (12)

With the dimension-invariant assumptions in (10), (11), and (12), we can derive the interest rate dynamics under the risk-neutral measure $\mathbb{Q}$ as,

$$dx_{j,t} = -\gamma_r \sigma_r^2 dt + \kappa_r b^{(j-1)}(x_{j-1,t} - x_{j,t}) dt + \sigma_r dW_{j,t}^Q,$$  \hfill (13)

where the risk-neutral mean-reversion speed remains the same as under the statistical measure $\mathbb{P}$. The risk-neutral mean of the lowest frequency component $x_{1,t}$ becomes $\theta_{r,t}^Q = \theta_r - \gamma_r \sigma_r^2 / \kappa_r$. A negative market price will make the risk-neutral mean higher than the statistical mean. Furthermore, under the risk-neutral measure, each frequency component $x_{j,t}$ varies around an adjusted level of its lower-frequency counterpart.

\footnote{See evidence in Collin-Dufresne and Goldstein (2002), Fan, Gupta, and Ritchken (2003), Heidari and Wu (2003), and Li and Zhao (2006).}
$x_{j-1,t} - \gamma_r \sigma_r^2 / \kappa_j$. The level adjustment declines as the frequency increases. In matrix form, the risk-neutral state vector dynamics can be written as,

$$dX_t = \left( \kappa \theta^Q - \kappa X_t \right) dt + \sigma_r dW_t^Q,$$

where the vector $\kappa \theta^Q$ has the first element being $\kappa_r \theta_r - \gamma_r \sigma_r^2$ and the remaining elements being $-\gamma_r \sigma_r^2$.

Under the linear instantaneous interest rate function in (9) and the affine risk-neutral dynamics in (14), the zero-coupon bond values are exponential affine in the state vector (Duffie and Kan (1996)),

$$P(X_t, \tau) = \exp \left( - b(\tau)^T X_t - c(\tau) \right),$$

where the coefficients solve the ordinary differential equations,

$$b'(\tau) = e_n - \kappa b(\tau),$$
$$c'(\tau) = b(\tau)^T \left( \kappa \theta^Q \right) - \frac{1}{2} b(\tau)^T b(\tau) \sigma_r^2,$$

starting at $b(\tau) = 0$ and $c(\tau) = 0$, with $e_n$ denoting a vector with the value one in the $n$-th position and zeros elsewhere. We can solve the ordinary differential equations in analytical forms.

**Proposition 3** Under the cascade factor dynamics in (2), instantaneous interest-rate function in (9), and the dimension-invariant assumptions (10), (11), and (12), merely five parameters $(\kappa_r, b_r, \sigma_r, \theta_r, \gamma_r)$ control the interest-rate dynamics under both the statistical and the risk-neutral measures, regardless of the number of frequency components. The values of zero-coupon bonds are also determined by these five parameters and the values of the state vector,

$$P(X_t, \tau) = \exp \left( - \sum_{j=1}^n b_j(\tau) x_{j,t} - c(\tau) \right),$$
where the coefficients $b_j(\tau)$ and $c(\tau)$ are given by

$$\begin{align*}
b_j(\tau) &= \sum_{i=j}^{n} \alpha_{i,j} \left( 1 - e^{-\kappa_i \tau} \right), \\
c(\tau) &= \kappa_r \theta_r \sum_{i=1}^{n} \alpha_{i,1} \left( 1 - \frac{e^{-\kappa_i \tau}}{\kappa_i} \right) - \gamma_r \sigma_r^2 \sum_{j=1}^{n} \sum_{i=j}^{n} \alpha_{i,j} \left( 1 - \frac{e^{-\kappa_i \tau}}{\kappa_i} \right), \\
&\quad - \frac{\sigma_r^2}{2} \sum_{j=1}^{n} \sum_{i=j}^{n} \sum_{k=j}^{n} \alpha_{i,j} \alpha_{k,j} \left( 1 - \frac{e^{-\kappa_k \tau}}{\kappa_k} - \frac{e^{-\kappa_i \tau}}{\kappa_i} + \frac{1 - e^{-(\kappa_i + \kappa_k) \tau}}{\kappa_i + \kappa_k} \right),
\end{align*}$$

where the values of $\alpha_{i,j}$ are given in equation (7) in Proposition 2.

Using this result, we can price bonds analytically given values of the state variables. We can also analyze the behavior of the instantaneous forward rate, which is affine in the state vector,

$$f(X_t, \tau) \equiv -\frac{\partial \ln P(X_t, \tau)}{\partial \tau} = \sum_{j=1}^{n} d_j(\tau) x_{jt} + e(\tau),$$

with $d(\tau) = b'(\tau)$ and $e(\tau) = c'(\tau)$. Specifically,

$$\begin{align*}
d_j(\tau) &= \sum_{i=j}^{n} \alpha_{i,j} \kappa_i e^{-\kappa_i \tau} = a_j(\tau), \\
e(\tau) &= \kappa_r \theta_r \sum_{i=1}^{n} \alpha_{i,1} \left( 1 - e^{-\kappa_i \tau} \right) - \gamma_r \sigma_r^2 \sum_{j=1}^{n} \sum_{i=j}^{n} \alpha_{i,j} \left( 1 - e^{-\kappa_i \tau} \right) \\
&\quad - \frac{\sigma_r^2}{2} \sum_{j=1}^{n} \sum_{i=j}^{n} \sum_{k=j}^{n} \alpha_{i,j} \alpha_{k,j} \left( 1 - e^{-\kappa_k \tau} - e^{-\kappa_i \tau} + e^{-(\kappa_i + \kappa_k) \tau} \right). \tag{23}
\end{align*}$$

From (21), we can regard $d_j(\tau)$ as the loading coefficient of the $j$-th factor on the forward rate at different maturities, or the contemporaneous response function of the forward rate to unit shocks from each of the $j$-th frequency component. Equation (22) shows that these factor loading coefficients at different maturities are identical to the response function of the instantaneous short rate at different time horizons. As such, the time-series statistical dynamics of the instantaneous interest rate and the cross-sectional term structure behavior of the forward rate curve are tightly linked.

The intercept $e(\tau)$ of the forward rate curve contains three components. The first term is proportional to the long-run mean $\theta_r$ of the instantaneous interest rate. The second term captures the contribution of the
market prices of risk. The last term captures the convexity adjustments. In the limit of infinite maturity \( \tau \to \infty \), the long-term forward rate converges to a constant,

\[
f_\infty = \theta_r - \frac{\sigma_r^2}{\kappa_r} \left( \gamma_r \kappa_r \frac{1 - b^{-n}}{1 - b^{-1}} + \frac{1 - b^{-2n}}{2} \right).
\] (24)

Comparing the statistical mean of the short rate \( \theta_r \) with the long-run forward rate \( f_\infty \), we anticipate a downward sloping mean forward curve in the absence of a risk premium \( (\gamma_r = 0) \). The downward slope is generated by the convexity term \( \sigma_r^2 / \kappa_r^2 \). The mean forward curve becomes upward sloping when the market price is sufficiently negative to overcome the convexity effect, \(-\gamma_r \kappa_r \frac{1 - b^{-n}}{1 - b^{-1}} > \frac{1 - b^{-2n}}{2}, \) or \(-\gamma_r > \frac{1 + b^{-n}}{2 \kappa_r} \).

We also observe that for large \( n \) the impact of increasing \( n \) becomes small, suggesting convergence as we now investigate.

### 2.3. Limiting behavior

We give conditions under which both the short rate and term structure weakly converge as the number of factors \( n \) grows. We first consider the dynamics of the short rate \( r_t = x_{n,t} \). Intuitively, using the property of the response functions in (8), we know that \( a_j(\tau) \leq 1 \) and thus \( a_j^n(\tau) \leq a_j(\tau) \). We can use these results to derive an upper bound for the unconditional variance of the instantaneous interest rate:

\[
Var(r_t) = \sigma_r^2 \sum_{j=1}^{n} \int_{0}^{\infty} a_j^2(s)ds \leq \sigma_r^2 \sum_{j=1}^{n} \int_{0}^{\infty} a_j(s)ds = \sigma_r^2 \sum_{j=1}^{n} \frac{1}{\kappa_j},
\] (25)

which is finite as long as \( \kappa_j > 0 \) for all \( j \). Using the power law scaling assumption, we can rewrite the upper bound as

\[
Var(r_t) \leq \sigma_r^2 \sum_{j=1}^{n} \frac{1}{\kappa_j} = \frac{\sigma_r^2 1 - b^{-n}}{\kappa_r 1 - b^{-1}}.
\] (26)

Therefore, as long as \( \kappa_r > 0 \) and \( b > 1 \), the unconditional variance of the instantaneous interest rate is finite regardless of how many factors are included in the cascade. This leads naturally to the following result:

**Proposition 4** For fixed \((\kappa_r, \sigma_r, \theta_r, b)\), let \( r_{n,t} \) denote the instantaneous interest rate \( r_t \) when the number of
factors in the cascade is given by \( n \). As \( n \to \infty \), the sequence of processes \( r_{n,t} \) weakly converges to a limit point with continuous sample paths and finite variance.

It is natural to denote by \( r_t \) the limit point of the sequence \( r_{n,t} \). In empirical work, as we use a larger number of components \( n \), we gain an increasingly accurate approximation of this limit point.

When in addition the parameter \( \gamma_r \) is also fixed, the entire term structure of interest rates weakly converges as \( n \) grows. In particular, the long-run forward rate converges to,

\[
\lim_{n \to \infty} r_{n,t} = \theta_r - \frac{\sigma_r^2}{\kappa_r^2} \left( \frac{\gamma_r \kappa_r}{1 - b^{-1}} + \frac{1}{2} + \frac{1}{b^{-2}} \right) \tag{27}
\]

as \( n \to \infty \), confirming that the risk premium and convexity terms determine the average term structure slope.

### 2.4. Future applications and extensions

While our dimension-invariant cascade dynamic term structure model focuses on the behavior of the interest rate term structure, one can readily extend the specification to analyze the behavior of interest rate options and bond excess returns for future research.

#### 2.4.1. Option pricing and stochastic volatility

Since it is difficult to identify the volatility dynamics from the interest rate term structure, our simplifying assumption of constant and identical volatility in (11) is appropriate for term structure modeling. For interest rate option pricing, the current practice, e.g., Heath, Jarrow, and Morton (1992), is to take the observed interest rates as given and focus on modeling the volatility dynamics. In this case, one can specify an analogous dimension-invariant cascade dynamics for interest rate volatilities.

The following candidate specification represents a natural extension of our term structure interest rate
dynamics specification for option pricing,

\[
\sigma_{j,t}^2 = v_{K,t} \tag{28}
\]

\[
dv_{k,t} = \kappa_k^e (v_{k-1,t} - v_{k,t}) dt + \omega_j \sqrt{v_{K,t}} dZ_{k,t}, \quad k = K, K-1, \ldots, 1, \tag{29}
\]

\[
v_{0,t} = \theta_v \tag{30}
\]

\[
\kappa_k^e = \beta^{k-1} \kappa_1^e, \quad \beta > 1, \tag{31}
\]

\[
\omega_j = \omega, \tag{32}
\]

\[
\rho = \mathbb{E}[dW_{j,t}dZ_{k,t}] / dt, \tag{33}
\]

where we replace the constant volatility specification by a system of \(K\)-dimensional cascade stochastic volatility dynamics. We set the instantaneous variance of all interest rate frequency components to the highest frequency variance component \(v_{K,t}\) in equation (28), which mean reverts to a lower-frequency stochastic variance component \(v_{K-1,t}\), and so on in equation (29), with \(\theta_v\) in equation (30) denoting the long-run mean of the instantaneous variance. We further achieve dimension-invariance by assuming power-law scaling on the mean-reversion speeds of the different variance frequency components in (31), constant and identical volatility of volatility coefficient \(\omega\) in (31), and identical correlation between the interest rate innovation and the variance innovation in (33).

We leave the topic of interest-rate option pricing for future research regarding further specification analysis and model estimation. One possible procedure for model estimation can be as follows. One first estimates our five-parameter term structure model with as many frequency components as needed to match the observed interest rates. Then, one can take the estimates of the five parameters as fixed, use the model to strip the forward rate curve, and proceed to estimate the parameters governing the variance dynamics in equations (28)-(33) using interest-rate options data. This estimation procedure extracts the interest-rate frequency components from the first step and extracts the interest-rate volatility factors from the second step.
2.4.2. Bond returns and time-varying risk premium

While it is difficult to extract the risk premium information from directly estimating term structure models, a much more direct way of understanding the interest rate risk premium behavior is to analyze the time series of bond returns over different investment horizons and on different bond maturities. For example, by analyzing bond returns over a one-year investment horizon, Cochrane and Piazzesi (2005) find that the one-year risk premiums affecting bonds of different maturities are all driven by a single factor, approximated by a tent-shaped function of forward rates.

For future research on risk premium analysis, one can use our cascade dynamic term structure model to extract the different frequency components from interest rates, \( \{x_{jt}\}_{j=1}^n \) and use the variance dynamics specified in (28)-(33) to extract the interest rate variance frequency components from interest-rate options, \( \{v_{kt}\}_{k=1}^K \). Then, one can analyze how bond excess returns over different investment returns depend on the level of these different frequency components,

\[
\text{Excess Bond Return}_{t+\Delta t} = a + \sum_{j=1}^n b_j x_{jt} + \sum_{k=1}^K c_k v_{kt} + \epsilon_{t+\Delta t}.
\]  

Through such regression analysis, one can study whether the coefficients \( \{b_j\} \) is tent-shaped at one-year investment horizon, whether the bond risk premium also depends on the interest rate variance at different frequencies, and how the dependence structure varies across different bond maturities and investment horizons.

3. Data and Estimation

We estimate the model using a panel data of U.S. dollar LIBOR and swap rates downloaded from Bloomberg. The LIBOR have maturities of one, two, three, six, nine, and 12 months, and the swap rates have maturities of two, three, four, five, seven, ten, 15, 20, and 30 years. The data are weekly (Wednesday) closing mid-quotes spanning 13 years from from January 4, 1995 to December 26, 2007, a total of 678 weekly observations for each series.
3.1. Summary statistics of LIBOR and swap rates

The LIBOR are simple-compounding interest rates that relate to the zero-coupon bond prices by

\[
LIBOR(X_t, \tau) = \frac{100}{\tau} \left( \frac{1}{P(X_t, \tau)} - 1 \right),
\]

where the maturity \( \tau \) follows actual/360 day-count convention, starting two business days forward. The swap rates relate to the zero-coupon bond prices by

\[
SWAP(X_t, \tau) = 100h \times \frac{1 - P(X_t, \tau)}{\sum_{i=1}^{\tau h} P(X_t, i/h)},
\]

where \( \tau \) denotes the maturity of the swap contract in years and \( h \) denotes the number of payments in each year. The swap contracts that we use make semi-annual payments \( (h = 2) \) and follow 30/360 day-count convention.

Table 1 reports summary statistics of the 15 interest-rate series. The table shows for each series the sample mean, standard deviation, skewness, excess kurtosis, and weekly autocorrelations of order one, five, ten, and 20. The average interest rates have an upward sloping term structure. The standard deviation shows a hump-shaped term structure that reaches its plateau at three-month maturity. All interest-rate series show small estimates for skewness and excess kurtosis. The interest rates are highly persistent, with first-order autocorrelations ranging from 0.9885 to 0.999.

Table [I about here.]

Figure 1 plots the time series of the LIBOR and swap data in the top panel. The short-term LIBOR started at 6% and varied between 5-6% in the 1990s, but dropped to about 1% during the recession around 2003. It reversed the trend and started to move up steadily from mid 2004 to 2006. The long-term swap rates did not move as much as the LIBOR did, thus generating variations in the term structure. The bottom panel plots the term structure of the LIBOR/swap rates at different dates, which show a wide variety of shapes including upward-sloping, downward-sloping, hump-shaped, and flat term structures.
3.2. Estimation methodology

We cast the model into a state space form, extract the distributions of the state variables at each date by a filtering technique, and estimate the model parameters using a quasi maximum likelihood procedure.

The state propagation equation is constructed based on a discretized version of the statistical dynamics of the interest rate factors $X_t$ in equation (3),

$$X_{t+1} = A + \Phi X_t + \sqrt{\Sigma} \varepsilon_{t+1},$$

where $\varepsilon \sim \mathcal{N}(0, I_n)$, $\Phi = \exp(-\kappa \Delta t)$, $A = (I - \Phi)\theta$, $\Sigma_x = \sigma_x^2 \Delta t I_n$, $\Delta t = 1/52$, and $I_n$ denoting an $n$-dimensional identity matrix.

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

$$y_t = h(X_t) + e_t, \quad h(X_t) = \begin{cases} \text{LIBOR}(X_t, i), & i = 1, 2, 3, 6, 9, 12 \text{ months} \\ \text{SWAP}(X_t, j), & j = 2, 3, 4, 5, 7, 10, 15, 20, 30 \text{ years} \end{cases},$$

where $y_t$ denotes the data observation, $h(X_t) = [\text{LIBOR}(X_t, i), \text{SWAP}(X_t, j)]$ denotes the model values of the corresponding LIBOR and swap rates as a function of the state vector $X_t$, and $e_t$ denotes a measurement error vector. We assume that the measurement errors are normally distributed iid random variables with zero mean and covariance matrix $\Sigma_y = \sigma_y^2 I_{15}$.

In systems where the state variables are Gaussian and the measurement equations are linear, the Kalman (1960) filter yields the efficient state updates in the least square sense. In our application, the state propagation equation (37) satisfies Gaussian linearity, but the measurement equations are nonlinear functions of the state variables. Traditional methods approximate the nonlinear function by using Taylor expansions to obtain extended forms of the Kalman Filter (EKF).\(^6\) Alternatively, Julier and Uhlmann (1997) propose the

\(^6\)Examples of EKF in term structure model estimation include Baadsgaard, Madsen, and Nielsen (2001), Chen and Scott (2003), Duan and Simonato (1999), and Duffee and Stanton (2008).
unscented Kalman Filter (UKF) to directly approximate the posterior density using a set of deterministically chosen sample points (sigma points). When these sigma points are propagated through the nonlinear measurement functions \( h(X_t) \), they capture the posterior mean and covariance accurately to the second order for any nonlinearity. We use the UKF approach to filter the mean and covariance of the states and measurement series. Specifically, we start with the linear Gaussian prediction on the state vector,

\[
\bar{X}_t = A + \Phi \hat{X}_{t-1}, \quad \bar{V}_{x,t} = \Phi \hat{V}_{x,t-1} \Phi^T + \Sigma_x,
\]

where \( \bar{X}_t \) and \( \bar{V}_{x,t} \) are the time-\((t-1)\) predicted value of the conditional mean and covariance matrix of the state vector. Based on these predictions, we draw a set of \( 2n+1 \) sigma vectors \( \chi_i \) on the state,

\[
\chi_{t,0} = \bar{X}_t, \quad \chi_{t,i} = \bar{X}_t \pm \sqrt{(k+\delta)(\bar{V}_{x,t})_j}
\]

where the weights are given by \( w_0 = \delta/(n+\delta) \) and \( w_i = 1/[2(n+\delta)] \) for \( i > 0 \), with \( \delta \) being a control parameter. We propagate the sigma points through the nonlinear measurement equation to obtain a set of sigma points on the measurements, \( \zeta_{t,i} = h(\chi_{t,i}) \), with which we compute the predicted mean \( \bar{y}_t \) and covariance matrix \( \bar{V}_{y,t} \) of the measurement series, as well as the covariance matrix between the state vector and the measurement \( \bar{V}_{xy,t} \):

\[
\bar{y}_t = \sum_{i=0}^{2k} w_i \zeta_{t,i}, \\
\bar{V}_{y,t} = \sum_{i=0}^{2k} w_i [\zeta_{t,i} - \bar{y}_t] [\zeta_{t,i} - \bar{y}_t]^T + \Sigma_y, \\
\bar{V}_{xy,t} = \sum_{i=0}^{2k} w_i [\chi_{t,i} - \bar{x}_t] [\zeta_{t,i} - \bar{y}_t]^T.
\]

Using these moment conditions, we apply the Kalman filter to obtain the filtered values of the mean \( \hat{X}_t \) and covariance \( \hat{V}_{y,t} \) of the state vector:

\[
\hat{X}_t = \bar{X}_t + K_t (y_t - \bar{y}_t), \quad \hat{V}_{y,t} = \bar{V}_{y,t} - K_t \bar{V}_{xy,t} K_t^T,
\]

where \( K_t = \bar{V}_{xy,t} (\bar{V}_{y,t})^{-1} \) denotes the Kalman gain.
Given the unscented Kalman filter forecasts on the conditional mean and covariance of the interest rate series at each date, we build the quasi log likelihood:

\[
l_t(\Theta) = -\frac{1}{2} \ln |\nabla_{y,t}| - \frac{1}{2} \left( (y_t - \bar{y}_t)^\top (\nabla_{y,t})^{-1} (y_t - \bar{y}_t) \right).
\]

(43)

We choose model parameters \(\Theta\) to maximize

\[
\max_{\Theta} \mathcal{L} = \sum_{t=1}^{N} l_t(\Theta),
\]

where \(N = 678\) denotes the number of weeks of observations. The procedure estimates the six parameters, \(\Theta \equiv [\kappa_r, \theta_r, \sigma_r, \theta_r^Q, b, \sigma_e^2]\), composed of the five model parameters and the pricing error variance \(\sigma_e^2\). We constrain \((\kappa_r, \theta_r, \sigma_r, \theta_r^Q, \sigma_e^2)\) to be positive and \(b\) to be greater than one in the estimation. The market price of risk is obtained from \(\gamma_r = \kappa_r (\theta_r^Q - \theta_r) / \sigma_e^2\).

Under normal circumstances, one would first specify the dimension of the state space before one can determine the parameter space. The number of parameters for a general term structure model increases rapidly with dimensionality. This curse of dimensionality has limited empirical works to low-dimensional specifications. By contrast, a distinct feature of our specification is that the number of parameters remains the same regardless of the number of factors.\(^7\) This dimension-invariant feature completely removes the curse-of-dimensionality problem and enables us to estimate high-dimensional models as easily as low-dimensional ones. As a result, dimensionality is no longer a concern but becomes a pure choice for us. Since the parameter space does not depend on the state space, we make the dimensionality choice in this last step. As we have 15 interest rate series, we proceed to estimate 15 different models, with the number of frequency components \(n = 1, 2, \ldots, 15\), respectively. Through this extensive estimation exercise, we analyze how many frequency components the data can identify and how high-dimensional models differ from low-dimensional models in capturing various aspects of the interest-rate behavior.

\(^7\)The one-factor model as one less parameter as we no longer need the power scaling coefficient for a one-factor structure.
3.3. Parameter estimates and likelihood tests

Table 2 reports parameter estimates, standard errors, and the maximized log likelihood for the 15 models. When \( n = 1 \), the model is equivalent to Vasicek (1977) and has no scaling parameter \( b \). All other models have the same number of parameters regardless of the number of frequencies. Given the extreme parsimony of the model specifications and the large amount of data (15 series, each with 678 observations), all six parameters are estimated with small standard errors.

The log likelihood \( (L) \) rises monotonically with the number of frequencies \( n \). The increase is rapid initially but levels off as we add more factors, consistent with the weak convergence demonstrated in Section 3. The estimates on the measurement error variance \( \sigma_e^2 \) decline with increasing model dimension. In the last column, we take the 15-factor model as the benchmark and compute the log likelihood difference between this benchmark model and the other 14 models. We represent this difference in terms of the Vuong (1989) statistic,

\[
\mathcal{V}_n = \sqrt{N} \left( m_{\rho^n} \right) / s_{\rho^n}, \quad \rho^n = l_{15}^n - l_n^n, \quad n = 1, 2, \ldots, 14, \tag{45}
\]

where \( \rho^n \) denotes the weekly log likelihood ratio difference between the 15-factor model and the model with \( n \) factors, \( m_{\rho^n} \) denotes the sample mean of the likelihood ratio, and \( s_{\rho^n} \) denotes the standard deviation of the likelihood ratio, which we compute by adjusting for the serial correlation of the likelihood ratio according to Newey and West (1987) with the lags chosen according to Andrews (1991). Asymptotically, the Vuong-statistic \( \mathcal{V}_n \) has a standard normal distribution. We say that the 15-factor benchmark model significantly outperforms the \( n \)-factor model at 95\% confidence level if the Vuong-statistic is greater than 1.64. The statistics suggest that the 15-factor model significantly outperform all lower-frequency models at the 95\% confidence level.

Comparing the parameter estimates under different number of frequency components reveals more insights about the model structure. The estimates for the mean-reversion speed of the lowest frequency component \( (\kappa_r) \) are similar regardless of how many factors we include. Thus, the estimation procedure identifies
the low frequency movements first. As we add more factors, higher frequencies are captured as well. The estimate for the scaling parameter \( (b) \) becomes smaller as \( n \) increases. As we allow more factors, the spacing of frequencies becomes finer.

Focusing on the best-performing 15-factor benchmark model, the lowest frequency is \( \kappa_r = 0.0572 \), corresponding to a time horizon \((1/\kappa_r)\) of about 17.5 years. The scaling coefficient \( b = 1.74 \) gives the highest frequency a mean reversion speed \( \kappa_{15} = 133.3 \), corresponding to a couple days. The estimate for the statistical long-run mean is close to zero, that for the risk-neutral mean is at 5.59%, and that for the instantaneous volatility is 1.56%. These estimates imply a negative market price of risk: \( \gamma_r = \kappa_r(\theta_r - \theta_r^Q)/\sigma_r^2 = -13.1. \)

The risk-neutral dynamics of the short rate determines the cross-sectional behavior of the interest rate across different maturities. In particular, from equation (18), we can write the instantaneous forward rate as an affine function of the state vector,

\[
   f(X_t, \tau) = d(\tau)^\top X_t + e(\tau).
\]

Thus, the mean forward rate term structure is given by

\[
   \mathbb{E}^P[f(X_t, \tau)] = d(\tau)^\top \theta_r + e(\tau),
\]

whereas the proportional coefficient \( d(\tau) \) measures the contemporaneous response of the forward rate curve to each unit shock in the state vector.

Figure 2 plots the mean forward rate curve in the top panel and the contemporaneous factor loading in the bottom panel, both as a function of the interest rate maturity. The mean forward rate curve is monotonically upward sloping, a direct result of the negative market price of risk. For the factor loading in the bottom panel, to differentiate the responses to different risk factors, we use a dashed line to denote the response to the highest frequency shock and a solid line to denote the response to the lowest frequency shock, while using dotted lines for all intermediate frequencies. Furthermore, given the power-law scaling on the fre-
quencies, we represent the time horizon $\tau$ in log scale in the bottom panel. As we discussed earlier, except for the highest frequency, which has an exponentially decaying response function, the response functions for all other frequencies possess hump-shaped functional forms, with the maximum responses occurring at increasingly longer horizons for lower frequency components.

[Figure 2 about here.]

4. Model performance comparison

The dimension-invariance feature of our specification allows us to estimate models of both low and high dimensions with equal ease and accuracy, thus enabling us to compare their relative performances. For ease of exposition, we choose the 15-factor model to represent the high-dimensional models and the three-factor model to present a low-dimensional benchmark. Since most term structure models estimated in the literature contain three or fewer factors, the three-factor model also serves as a benchmark for the literature.

4.1. In-sample fit and yield curve stripping

Table 3 reports the summary statistics of the pricing errors from the two representative models. The statistics include the sample mean, root mean squared error (Rmse), weekly autocorrelation (Auto), maximum absolute error (Max), and percentage explained variation (VR), defined as one minus the ratio of the mean squared pricing error to the variance of the original series. The summary statistics on the pricing errors from the three-factor model in panel A are similar to those reported in the literature for typical three-factor models. The root mean squared error averages over six basis points. Since the bid-ask spreads for swap rates average around half to one basis point, these pricing errors, albeit small, are still economically significant. The economic importance of the fitting errors becomes apparent when one forms interest-rate portfolios to hedge away the more persistent factors (Bali, Heidari, and Wu (2009)). By contrast, the 15-factor model fits observed interest rates to near perfection, with the root mean squared pricing error averaging less than one basis point and the explained variations being close to 100% for all series.
Combining near-perfect fit with extreme parsimony and tractability in an arbitrage-free framework, an obvious and practical application for the high-dimensional model is to serve as a basis function for stripping yield curves from swaps and coupon bonds. Typically, constructing a forward-rate curve from a discrete number of observations of coupon bonds or swap rates involves choosing a functional form to link the forward rates across different maturities and estimating parameters of the assumed functional form. Basis functions that have been proposed for yield curve stripping include polynomials (Chambers, Carleton, and Waldman (1984)), cubic splines (McCulloch (1975) and Litzenberger and Rolfo (1984)), step functions (Ronn (1987)), piece-wise linear specifications (Fama and Bliss (1987)), and exponentials (Nelson and Siegel (1987) and Svensson (1994)). From the forward rate expression in (21) and (22), we can think of \( \{a_j(\tau)\}_{j=1}^n \) as the basis functions derived from our model for the forward rate curve, with the frequency components \( \{x_{j,t}\}_{j=1}^n \) as weights that vary over time to capture the different term structure shapes. As shown in Proposition 1, these basis functions are convolutions of exponential density functions.

Of all the basis functions, the exponential function forms of Nelson and Siegel (1987) and Svensson (1994) have become popular choices in the academic literature. The Federal Reserve has also adopted a variation of this functional form for stripping yield curves from Treasury bonds. The original Nelson-Siegel basis function cannot be made consistent with any interest-rate dynamics (Filipović (1999)). As such, the stripped yield curve can potentially allow arbitrage opportunities. Christensen, Diebold, and Rudebusch (2008) propose modification of the original function to make it consistent with a three-factor dynamic term structure model, thus resolving the consistency issue. While the three-factor structure does capture major variations of the yield curve (Diebold and Li (2006)), the remaining fitting errors can still be economically significant (Bali, Heidari, and Wu (2009)) and the functional forms can become overly restrictive in some instances. Cochrane and Piazzesi (2008) further argue that one loses important information in fitting 15 interest rates series using even the six-factor Svensson model. Indeed, practitioners rarely use these models to strip the forward rate curve from LIBOR and swap rates. The common industry practice is to assume a piece-wise constant step function for the forward rate and estimate the levels of the steps sequentially from

\[\text{Table 3 about here.}\]
low to high maturities. The length of the steps are determined by the maturities of the LIBOR and swap rates used in the stripping procedure. The top panel of Figure 3 shows the piece-wise constant forward rate curves stripped from the LIBOR and swap rates based on this common industry practice.

Our cascade dimension-invariant term structure model also generates basis functions as convolutions of exponential functions. By using a high-dimensional structure, we can strip smooth, dynamically consistent forward rate curves while fitting the observed LIBOR and swap rates to near perfection, thus alleviating the information loss concern of Cochrane and Piazzesi (2008). The bottom panel of Figure 3 plots the forward rate curves generated from our estimated model with 15 factors. The overall shapes of the forward rate curves in the two panels are similar, and both types of forward curves can match the observed LIBOR and swap rates. Nevertheless, our model-generated curves in the bottom panel have several advantages over the piece-wise constant forward assumption in the top panel. First, our forward curves are dynamically consistent and thus exclude arbitrage opportunities. The piece-wise constant forward assumption matches the observed LIBOR and swap rates exactly, and thus does not introduce arbitrage opportunities on the observed interest rates. Nevertheless, the discontinuities can induce instabilities when used as inputs to price interest-rate options in the forward rate model of Heath, Jarrow, and Morton (1992). In this sense, our dynamic term structure model provides a good starting point for the interest-rate option pricing literature by generating a smooth, dynamically consistent forward rate curve that matches the observed LIBOR and swap rates.

4.2. Cross-correlations among different interest rate series

When measuring the cross-correlations between changes in non-overlapping forward rates, Dai and Singleton (2003) find that low-dimensional term structure models typically imply much higher correlations than those estimated from the data. Intuitively, a low-dimensional model captures the systematic, common movements in the interest rate term structure. By design, interest rate fair values built purely from these common
movements show high cross-correlation. With a high-dimensional structure, our model can accommodate both systematic movements across the whole term structure and idiosyncratic movements in a particular LIBOR/swap instrument. As such, the high-dimensional model has the promise of generating interest rate fair values that match the cross-correlation behavior observed in the data.

To measure cross-correlations between non-overlapping forward rates, one must strip the swap rates. The estimates would thus depend on the particular stripping method and the basis function. To avoid such contamination, we measure cross-correlations between the observed LIBOR and swap rates. Their overlapping nature dictates that their cross-correlations can be much higher than between non-overlapping forward rates. Nevertheless, our objective is to investigate whether high-dimensional model can match what is observed in the data.

Our 15 interest-rate series generate a $(15 \times 15)$ correlation matrix. For ease of exposition, we take the six-month LIBOR as the basis instrument and measure its correlation with other LIBOR and swap rates. Figure 4 reports the correlation estimates between the weekly changes of the six-month LIBOR and weekly changes in other LIBOR and swap rates across different maturities. Circles denote the cross-correlation estimates from data. The solid line denotes estimates from model values generated from the 15-factor model. The dash line denotes estimates from model values generated from the three-factor model. The solid line matches the data very well, but the dashed line is well above the data. Therefore, by using a high-dimensional structure, we can readily overcome one of the major limitations of low-dimensional term structure models.

4.3. Interest rate forecasting

Several studies, e.g., Duffee (2002) and Bali, Heidari, and Wu (2009), find that low-dimensional dynamic term structure models perform miserably in forecasting interest-rate movements. The forecasts are often no better than assuming a simple random walk. We attribute this poor forecasting performance to two major
issues. First, these studies often estimate a general three-factor specification, which normally involves over 20 model parameters, many of which cannot be estimated with statistical significance. Thus, even the estimated dynamics generate good in-sample performance, the parameter identification issue dictates that the parameter estimates are unstable over time and as a result the out-of-sample performance can deteriorate significantly. Second, as we have observed earlier in Table 3, a three-factor model does not fit the observed interest rates perfectly. When one performs interest-rate forecasting based on the factor dynamics, the fitting error will carry over to become part of the forecasting error, and its relative contribution to the forecasting error increases with shrinking forecasting horizon. By contrast, the random walk hypothesis does not involve parameter estimation, and thus does not experience any out-of-sample deterioration. Furthermore, forecasts based on the random walk assumption start (and end) with the currently observed interest rate level. The forecasting error converges to zero as the forecasting horizon shrinks to zero.

While allowing much richer interest-rate dynamics than the random walk, our high-dimensional dimension-variant cascade model can also overcome the issues associated with traditionally estimated low-dimensional models. First, the model only has five parameters regardless of the dimension. This extreme parsimony is likely to bring stability to its out-of-sample performance. Second, the high-dimensional model fits the observed interest rates to near perfection, and as a result, the forecasting error no longer contains a fitting error component. To the extent that the interest-rate dynamics are well-specified, there is hope that our model can generate better forecasting performance than the random walk assumption.

To verify these ideas, we compare the forecasting performance of the three- and 15-factor dimension-invariant cascade specifications against two standard benchmarks. The first benchmark is the random walk hypothesis (RW) and the second is a first-order autoregression (AR). We first consider in-sample forecasting, under which we estimate the parameters for each approach using the entire sample period and compare the prediction errors from the same period. Under the AR strategy, we estimate an AR(1) regression on each interest-rate series $j$ for each forecasting horizon $h$,

$$y_{j,t+h} = a + by_{j,t} + e_{j,t+h}, \quad j = 1, \cdots, 15,$$

(48)

\[\text{As the forecasting horizon shrinks to zero, the forecasting error on the interest rate factors converges to zero, all errors left on the forecasts of the observed interest rate series are due to the model fitting error.}\]
where $y_{jt}$ denotes the time-$t$ observed value of the $j$-th interest rate (LIBOR or swap rate) series, and we consider forecasting horizons at $h = 1, 2, 3,$ and 4 weeks. Given the full-sample estimates on the autoregressive coefficients $(\hat{a}, \hat{b})$, we generate forecasted values $\bar{y}_{jt+h} = \hat{a} + \hat{b}y_{jt}$, and define the forecasting error as $e_{jt+h} = y_{jt+h} - \bar{y}_{jt+h}$. When forecasting using a term structure model, we use the filtered state values at each date $\hat{X}_t$ and the statistical factor dynamics to predict future values of the state over different horizons $h$,

\[
\bar{X}_{t+h} = A_h + \Phi_h \hat{X}_t,
\]

with $\Phi_h = \exp(-\kappa h \Delta t)$, $A_h = (I - \Phi_h) \theta$, and $\Delta t = 1/52$ denoting the length of each period. Based on the forecasted values of the state vector $\bar{X}_{t+h}$, we compute the forecasted values of the LIBOR and swap rates according to (35) and (36), respectively. Finally, the forecasting error from the random walk hypothesis, in sample or out of sample, is simply the interest rate changes over the forecasting horizon, $\Delta y_{jt+h} = y_{jt+h} - y_{jt}$.

To measure the forecasting performance, we compare the forecasting error $e_{jt+h}$ of each strategy to the forecasting error from the random walk hypothesis. We measure the performance difference using the predictive variation (PV), defined as one minus the ratio of mean squared predicting error to mean squared interest rate change:

\[
PV_j = 1 - \frac{\frac{1}{N-h} \sum_{t=1}^{N-h} (e_{jt+h})^2}{\frac{1}{N-h} \sum_{t=1}^{N-h} (\Delta y_{jt+h})^2}.
\]

The predictive variation is positive when the strategy outperforms the random walk hypothesis.

Table 4 reports the in-sample predictive variation estimates from different strategies and over different predicting horizons. By design, the in-sample forecasting error from the autoregressive regression is always smaller than the interest rate change. Hence, the in-sample predictive variation estimates from the autoregressive regression are all positive, as shown in panel A of Table 4. The estimates range from 17.53% to 68.12% for the LIBOR series, but mostly within 10% for the swap rate series, indicating that in sample, an AR(1) regression can explain a significant portion of changes in the LIBOR series, but a small portion for the swap rates. The estimates in panel B show that the three-factor model performs worse than the random walk hypothesis at the one-week horizon for all LIBOR and swap rate series. The under-performance is
especially strong for longer-term LIBOR series. At longer forecasting horizons, the predictive variations on one-, two-, and three-month LIBOR become positive, but the estimates on six-, nine-, and 12-month LIBOR remain negative. The predictive variation estimates on the swap rate series are negative at one week horizon, but become less negative at the four-week horizon. The negative estimates are in line with the results in Duffee (2002) and Bali, Heidari, and Wu (2009) for different specifications of three-factor affine models and verify that low-dimensional term structure models generate poor forecasts.

Panel C reports the predictive variation estimates from the 15-factor model. The estimates are positive for all LIBOR series across all four forecasting horizons. Thus, by using a higher dimension, the model generates markedly better forecasting performance and can outperform the random walk hypothesis on all LIBOR series. The predictive variation estimates on the swap rate series are all close to zero, suggesting that the model cannot do any better than the random walk hypothesis in predicting the swap rate changes. It is worth noting that even the in-sample predictive variation estimates from the autoregressive regression are very small on the swap rate series, suggesting that changes in swap rates are inherently difficult to predict.

Our cascade structure naturally separates different frequency components in the interest rate movements. The low-frequency components behave like random walks and it is therefore inherently difficult to predict changes in these components; however, high-frequency components can show strong mean reversion and it is easier to predict their changes via the mean-reversion behavior. Since low-dimensional models only identify the low-frequency components while discarding the predictable high-frequency components, it is not surprising that these models generate poor forecasting performance. Only by capturing the high-frequency components can a model fully capture the predictability inherent in the interest-rate series. The high-frequency components by design can only be captured by a high-dimensional model structure. Furthermore, since the strongly predictable high-frequency movements dominate the movements of short-term LIBOR rates, a high-dimensional model can significantly outperform the random walk hypothesis in predicting changes in the LIBOR movements.

To gauge the out-of-sample stability of each forecasting strategy, starting from January 7, 1998, we re-estimate the autoregressive coefficients and the model parameters at each date $t$ using the data up
to that date, and generate forecasts over the next four weeks based on the coefficient estimates on that
date. For this out-of-sample analysis, we focus on the performance of the AR(1) strategy and the 15-factor
model. Since the three-factor model cannot outperform the random walk hypothesis in sample, its out-
of-sample performance can only be worse. Table 5 reports the results from the out-of-sample forecasting
exercise. Panel A reports the predictive variation of the AR(1) regression. Although the autoregressive
regression generates the best in-sample forecasting performance, its out-of-sample performance deteriorates
dramatically. The performance is worse than the random walk hypothesis across all interest-rate series and
over all four forecasting horizons. The performance deterioration suggests that an autoregressive regression
on a highly persistent series is very likely to generate worse out-of-sample forecasting performance results.

Panel B reports the out-of-sample performance of the 15-factor model. The out-of-sample predictive
variation estimates are very close to the corresponding in-sample estimates. The 15-factor model out-
performs the random walk hypothesis on all LIBOR series and over all four forecasting horizons. The
out-of-sample stability of the 15-factor model derives largely from its extreme parsimony. Since only five
parameters control the dynamics and term structure behaviors of the 15 data series, these parameters can
be estimated with reasonable stability. By contrast, since the AR(1) regression is performed on each data
series, estimating the AR(1) relation on all 15 series generate 30 coefficients.

In performing time series analysis, a natural idea is to exploit the cross-correlations of the interest rates
and build a VAR(1) forecasting system. In theory, a VAR(1) system is more efficient in extracting the infor-
mation from the whole term structure. In practice, however, a general VAR(1) system would have too many
free parameters to be estimated with any accuracy. By contrast, our model essentially builds a constrained
VAR(1) system on the forward rates that exploits the information from the whole term structure, but with
merely five parameters. It is this combination of cross-sectional information with extreme parsimony that
generates the good forecasting performance both in sample and out of sample. The three-factor model is
also parsimonious and is thus also stable. Yet, it is not flexible enough to generate good performance. The
15-factor model possesses both the flexibility due to the high-dimensional structure and stability due to the
same degree of parsimony.
To gauge the statistical significance of the outperformance of the 15-factor model over the random walk hypothesis, panel B of Table 5 also reports the \( t \)-statistics of the performance difference between the the 15-factor model and the random walk hypothesis. The \( t \)-statistic is constructed on the differences in the squared forecasting errors between two forecasting methods. Asymptotically, it has a standard normal distribution. A positive \( t \)-static suggests that the model outperforms the random walk hypothesis. The statistics show that the model outperforms the random walk hypothesis for almost all LIBOR series and over most forecasting horizons. The outperformance is statistically significant in most cases. The \( t \)-statistics on the swap rate series are mostly insignificant.

5. Specification analysis

Under the cascade structure, we make three key parametric assumptions on the distributions of coefficients of different frequency components to make the model dimension invariant. In this section, we investigate the empirical reasonability of these assumptions and discuss the merits of alternative specifications.

5.1. Is power law scaling empirically accurate?

To verify the empirical validity of power-law scaling in the mean reversion speeds \( \kappa_j = \kappa_r b^{(j-1)} \), we estimate a 15-factor model with \( \{\kappa_j\}_{j=1}^n \) being free parameters without scaling constraints. The total number of parameters increases from the original six to 19.

Given the large number of parameters, the optimization routine becomes much slower. In Figure 5, we plot the logarithm of the \( \kappa_j \) estimates in circles, and we also plot in solid line the linear relation implied by the power-law scaling assumption with \( \kappa_r = 0.0572 \) and \( b = 1.74 \). The estimates for the free parameters \( \kappa_j \) vary around the solid line, suggesting that the scaling assumption holds reasonably well.

The maximized log likelihood values of this model is 29,997, whereas that of the model with power-
law scaling constraint is 29,376. The likelihood ratio test with 13 degrees of freedom strongly rejects the constrained model. Nevertheless, the added freedom does not improve the forecasting performance on the interest rates. When we try to perform rolling estimation of the model to generate out-of-sample forecasting results, the estimation often experiences convergence issues, making the out-of-sample exercise difficult to proceed. When we use the whole-sample parameter estimates to perform the forecasting exercise, the grand average of the percentage predicted variation on the six LIBOR rates over the four different horizons is 21.13%. The analogous in-sample performance on the dimension-invariant constrained model is 21.10%. Thus, the model with the power-law scaling performs as well as the model with free scalings on the mean-reversion speeds. Meanwhile, the power-law scaling significantly improves the identification and stability of model parameters.

5.2. Do risks and risk premiums scale across frequencies?

Given well-documented difficulties in extracting the volatility dynamics and bond risk premiums from term structure model estimation, we make the simplifying assumption that the instantaneous volatilities and market prices of risk are constant and identical for all frequency components. To investigate whether the volatilities and market prices also scale across frequencies, we estimate a 15-factor model that allows different scalings on both the volatilities and the market prices. We let the instantaneous variance of each frequency component to scale as,

\[ \sigma_j^2 = \sigma_r^2 b^{(j-1)s} \zeta, \]

and we accommodate a more flexible risk premium specification at each frequency,

\[ \gamma_j \sigma_j^2 = \gamma_0 \sigma_j^2 - \gamma_1 \sigma_j^2 x_{j-1} + \gamma_2 \sigma_j^2 x_j, \]

where
where we allow different scalings for each of the three components,

\[
\begin{align*}
\gamma_0 \sigma_j^2 &= \gamma_0 \sigma_r b_t \delta^{(i-1)} s_0, \\
\gamma_1 \sigma_j^2 r_{j-1,t} &= \gamma_1 \sigma_r^2 b_t^{(j-1)} s_1 r_{t-1,t}, \\
\gamma_2 \sigma_j^2 r_{j,t} &= \gamma_2 \sigma_r^2 b_t^{(j-1)} s_2 r_{t,t}.
\end{align*}
\]

This relaxed specification adds two more market price coefficients (\(\gamma_1, \gamma_2\)) and four more scaling exponent parameters (\(s_\sigma, s_0, s_1, s_2\)). A zero estimate for a particular scaling exponent parameter, say \(s_\sigma\), indicates that the volatility \(\sigma_j\) does not scale across different frequencies. A positive estimate indicates that the component increases with higher frequency and a negative estimate indicates that the component declines with increasing frequency.

Table 6 reports the scaling parameter estimates on the risks and risk premiums. The three risk premium coefficients are all negative, suggesting that the risk premium increases with increasing deviation of \(x_{j,t}\) from its lower-frequency neighbor \(x_{j-1,t}\). The estimates for the scaling exponent on the instantaneous variance are negative, suggesting that the instantaneous variance become smaller as higher frequencies. However, the exponent estimates are much smaller than one in absolute magnitude. Thus, the instantaneous variance varies much slower with frequency than does the mean reversion speed. The scaling exponent on the constant risk premium component \(s_0\) is estimated to be positive, suggesting that the risk premium is larger in absolute magnitude for high frequency risks. Again, however, the estimate is much smaller than one and hence the variation is much smaller than does the mean reversion speed. The scaling exponent estimate on \(s_1\) is virtually zero. The scaling exponent estimate on \(s_2\) is negative at \(-1.76187\), suggesting that the risk premium can have some dependence on the risk level, but the dependence becomes smaller at higher frequencies.

Overall, the estimation suggests that the risks and risk premiums identified from the term structure do not vary nearly as much across frequencies as does the mean reversion speed. It is thus appropriate to assume stability across frequencies. Furthermore, when we allow the risk premiums to depend on the risk levels and its lower-frequency levels, the risk premiums do not show apparent scaling. If anything, it suggests that the time-varying component of the risk premium become smaller at higher frequency, while the constant component becomes larger. Hence, it is appropriate to maintain a constant market price for parsimony and
stability for term structure model estimation. On the other hand, parametrization across frequencies can be fruitful when one performs bond excess return regressions as in (34) to further investigate the behavior of bond risk premiums.

5.3. Alternative specifications to guarantee positive interest rates

The linear Gaussian structure of our cascade dimension-invariant term structure model allows extreme analytical tractability on how the risks dissipate over time and how the risk factors affect the term structure of interest rates. Yet, the Gaussian structure dictates that interest rates exhibit positive probabilities of becoming negative. To exclude positive probabilities of negative interest rates, we consider two alternative dimension-invariant specifications.

The first specifies the instantaneous volatility of each frequency to be proportional to the square root of the corresponding risk level, \( \sigma_{j,t} = \sigma_r \sqrt{x_{j,t}} \). The one-factor version of this alternative specification traces back to the classic model of Cox, Ingersoll, and Ross (1985). We further assume that the market price of risk is proportional to the square root of the risk level \( \gamma_r \sqrt{x_{j,t}} \) so that the risk premium is proportional to the risk level, and \( \kappa_{j,j}^{\Omega} = \kappa_{j,j} + \gamma_r \sigma_r^2 \). In this case, zero-coupon bond prices remain exponential affine functions of the frequency components, but the loading coefficients can no longer be solved analytically. Instead, they can be solved numerically from the following set of ordinary differential equations,

\[
\begin{align*}
  b' (\tau) & = e_n - (\kappa^{\Omega})^T b(\tau) - \frac{1}{2} \sigma_r^2 [b(\tau) \odot b(\tau)], \\
  c' (\tau) & = b_1 (\tau) \kappa_r \theta_r,
\end{align*}
\]  

(54)

starting at \( b(0) = 0 \) and \( c(0) = 0 \). When we estimate this alternative specification with 15 factors, we obtain the maximized log likelihood value at 29,377, almost identical to the maximized likelihood value of the linear Gaussian model. The in-sample forecasting on the six LIBOR rates generates a grand average of the predicted variation of 21.16\%, again similar to the value from the linear Gaussian specification (21.10\%).

The second alternative is to maintain the linear Gaussian structure of the frequency components but to
set the instantaneous interest rate to the square of the highest-frequency component,

\[ r_t = x_{n,t}^2. \] (55)

The quadratic short rate function dictates that zero-coupon bond values become exponential quadratic functions of the frequency components (Leippold and Wu (2002)),

\[ P(X_t, \tau) = \exp \left( -X_t^\top B(\tau) X_t - b(\tau)^\top X_t - c(\tau) \right), \] (56)

where the coefficients can be solved numerically from the following set of ordinary differential equations,

\[
\begin{align*}
B'(\tau) &= \mathbf{Z}_n - B(\tau) \kappa - \kappa^\top B(\tau) - 2B(\tau)^2 \sigma_r^2, \\
b'(\tau) &= 2B(\tau) \kappa \theta^Q - \kappa^\top b(\tau) - 2B(\tau) b(\tau) \sigma_r^2, \\
c'(\tau) &= b(\tau)^\top \kappa \theta^Q + \text{tr} B(\tau) \sigma_r^2 - \frac{1}{2} b(\tau)^\top b(\tau) \sigma_r^2,
\end{align*}
\]

starting at \( B(0) = 0, b(0) = 0, \) and \( c(0) = 0, \) where \( \mathbf{Z}_n \) is a matrix of zeros except for the \((n, n)\)th element being one. When we estimate this quadratic model with 15 factors, we obtain the maximized log likelihood value at 29,472, slightly higher than the linear Gaussian model. The in-sample forecasting on the six LIBOR rates over the four different horizons generates a grand average of the predicted variation of 21.05%, slightly lower than the value from the linear Gaussian specification (21.10%).

The two alternative specifications are the same as the linear Gaussian model in terms of dimension-invariance and parsimony, but are more numerically complex as the bond pricing solutions are no longer analytical. Since all three specifications generate similar performance, we recommend the linear Gaussian structure for its simplicity and analytical tractability when the probability of negative rates is not a serious concern. When it is, either of the two alternatives can be chosen according to the specific application.
6. Conclusion

We develop a class of dynamic term structure models that are free from factor rotation, extremely parsimonious, and dimension invariant. The model uses a cascade dynamic structure to rank factors according to their frequencies, and thus completely removes factor rotation. Furthermore, through parameterizing the distribution of the coefficients governing different frequency components, the model uses merely five parameters to govern the interest-rate statistical dynamics and the term structure behavior, regardless of the dimension of the state vector. The dimension-invariance completely removes the well-known curse-of-dimensionality problem in general specifications, and allows us to estimate models with arbitrarily high dimensions with equal ease and statistical significance.

Through an extensive estimation exercise, we show that the data ask for a much higher dimension that the three-factor structure traditionally focused on in the existing literature. We also show that a high-dimensional model can resolves several well-documented limitations of its low-dimensional counterparts by generating both near-perfect in-sample fitting and superior out-of-sample forecasting performances.

One of the drawbacks of traditional dynamic term structure models is that they perform poorly in pricing interest rate options. One potential reason for the poor performance is that the models do not fit the underlying interest rate curve perfectly and the options are written on observed interest rates, not on the model values. By accommodating a much higher-dimensional structure, our model can fit the interest rate term structure to near perfection. As a result, the forward rate curves derived from our model can be directly used as inputs in existing forward rate models such as Heath, Jarrow, and Morton (1992) for interest-rate option pricing. One can also specify the interest-rate variance dynamics in an analogous dimension-invariant cascade structure to obtain better pricing performance of interest-rate options across a wide range of option maturities. The extracted frequency components from interest rates and interest rate volatilities can then serve as the basis for bond risk premium analysis. We leave these areas for future research.
A. Appendix

A.1. Proof of Proposition 1

We prove the proposition by induction. Consider the one-factor case \((n = 1)\). We infer from Ito’s lemma that 
\[
d(e^{\kappa t}x_{1,t}) = \kappa_1 e^{\kappa t}x_{1,t} dt + e^{\kappa t}d(x_{1,t}).
\]
Since 
\[
d(x_{1,t}) = \kappa_1 (\theta_r - x_{1,t}) dt + \sigma_1 dW_{1,t},
\]
we have
\[
d(e^{\kappa t}x_{1,t}) = e^{\kappa t} (\kappa_1 \theta_r dt + \sigma_1 dW_{1,t}).
\]
Integrating both sides and then dividing by \(e^{\kappa t}\), we obtain the one-factor case of equation \((4)\).

We now assume that the proposition \((4)\) holds for an \((n - 1)\)-factor structure,
\[
x_{n-1,t} = \theta_r + \sum_{j=1}^{n-1} (x_{j,0} - \theta_r) a_j^{n-1}(t) + \sum_{j=1}^{n-1} \sigma_j \int_0^t a_j^{n-1}(t-s) dW_{j,s},
\]
(57)
where \(a_j^{n-1} = (K_j \cdots K_{n-1})/K_j\). Ito’s lemma implies 
\[
d(e^{\kappa t}x_{n,t}) = \kappa_n e^{\kappa t}x_{n-1,t} dt + \sigma_n e^{\kappa t}dW_{n,t}.
\]
Integrating both sides and then dividing both sides by \(e^{\kappa t}\), we have
\[
x_{n,t} = e^{-\kappa t}x_{n,0} + \int_0^t \kappa_n e^{-\kappa(t-s)} x_{n-1,s} ds + \sigma_n \int_0^t e^{-\kappa(t-s)} dW_{n,s}.
\]
Substitute out \(x_{n-1,s}\) according to equation \((57)\),
\[
\int_0^t \kappa_n e^{-\kappa(t-s)} x_{n-1,s} ds = \theta_r (1 - e^{-\kappa t}) + \sum_{j=1}^{n-1} (x_{j,0} - \theta_r) \int_0^t \kappa_n e^{-\kappa(t-s)} a_j^{n-1}(s) ds
\]
\[
+ \sum_{j=1}^{n-1} \sigma_j \int_0^t \kappa_n e^{-\kappa(t-s)} \left[ \int_0^t a_j^{n-1}(s-u) dW_{j,u} \right] ds.
\]
Let \(a_j^n(t) = e^{-\kappa t}a_j(t)\), and \(a_j^n(t) = \int_0^t \kappa_n e^{-\kappa(t-s)} a_j^{n-1}(s) ds\) for \(j \leq n-1\), we have,
\[
x_{n,t} = \theta_r + \sum_{j=1}^{n} (x_{j,0} - \theta_r) a_j^n(t) + \sigma_n \int_0^t e^{-\kappa(t-s)} dW_{n,s} + \sum_{j=1}^{n-1} \sigma_j \int_0^t \left[ \int_u^t \kappa_n e^{-\kappa(t-s)} a_j^{n-1}(s-u) ds \right] dW_{j,u}.
\]
We observe that
\[
\int_u^t \kappa_n e^{-\kappa_n(t-s)} a_{j}^{n-1}(s-u)ds = \int_0^{t-u} \kappa_n e^{-\kappa_n(t-u-s')} a_{j}^{n-1}(s')ds' = a_{j}^n(t-u).
\]

Thus, the proposition holds for the \( n \)-factor structure. We conclude that the proposition holds for all \( n \).

The coefficients are defined recursively by
\[
a_{j}^n(t) = \int_0^t \kappa_n e^{-\kappa_n(t-s)} a_{j}^{n-1}(s)ds \text{ starting with } a_{j}^1(t) = e^{-\kappa_j t}.
\]

Let \( K_i(t) = \kappa_i e^{-\kappa_i t} \) for \( t > 0 \) denote an exponential probability density, we can write the recursive valuation in terms of convolution products of exponential probability density functions,
\[
a_{j}^n = K_n * a_{j}^{n-1} = K_n * K_{n-1} * a_{j}^{n-2} = \cdots = K_n * K_{n-1} * \cdots K_j / \kappa_j.
\]

This proves the convolution result in equation (5).

\section*{A.2. Proof of Proposition 2}

Since \( a_j = (K_j * \cdots * K_n) / \kappa_j \), we infer that
\[
a_j(\tau) = \frac{\kappa_{j+1}}{\kappa_j} \int_0^\tau K_j(\tau-s)a_{j+1}(s)ds.
\]

We differentiate this relation with respect to \( \tau \)
\[
a_j'(\tau) = \frac{\kappa_{j+1}}{\kappa_j} \left[ \int_0^\tau -\kappa_j K_j(\tau-s)a_{j+1}(s)ds + K_j(0)a_{j+1}(\tau) \right].
\]

This implies the relations: \( a_j'(\tau) = \kappa_{j+1}a_{j+1}(\tau) - \kappa_ja_j(\tau) \) and
\[
a_j''(\tau) = \kappa_{j+1}a_{j+1}'(\tau) - \kappa_ja_j'(\tau).
\]

An interior local optimum of \( a_j \) therefore satisfies
\[
a_j''(\tau) = \kappa_{j+1}a_{j+1}'(\tau)
\]

38
since $a_j'(\tau) = 0$.

We now show by backward induction that for all $j = n - 1, \cdots, 1$, the function $a_j(\tau)$ is a single peaked function reaching a maximum at $\bar{\tau}_j$. Furthermore, $\bar{\tau}_1 \geq \cdots \geq \bar{\tau}_n$.

The property holds for $j = n - 1$. The function $a_{n-1}(\tau) = \frac{\kappa_n}{\kappa_n - \kappa_{n-1}} (e^{-\kappa_{n-1}\tau} - e^{-\kappa_n\tau})$ is a hump-shaped function, which reaches a maximum when $\tau = \bar{\tau}_{n-1} = \ln(\kappa_n/\kappa_{n-1})/(\kappa_n - \kappa_{n-1})$.

Assume that the property holds for $j + 1$. Let $\bar{\tau}_j$ denote the smallest local maximum of $a_j$. We know that $a_j''(\bar{\tau}_j) \leq 0$, and that $a_j''(\bar{\tau}_j) = \kappa_{j+1}a''(\bar{\tau}_j)$. Hence $a_{j+1}''(\bar{\tau}_j) \leq 0$, which implies that $\bar{\tau}_j \geq \bar{\tau}_{j+1}$. If the function is nonmonotonic, there exists a local minimum $\tau > \bar{\tau}_j$. Since $\tau > \bar{\tau}_j \geq \bar{\tau}_{j+1}$, we know that $a_j''(\tau) = \kappa_{j+1}a''(\bar{\tau}_j) < 0$, which is a contradiction. We conclude that $a_j$ is single peaked and reaches a maximum at $\bar{\tau}_j \geq \bar{\tau}_{j+1}$.

The analytical solutions and proofs for the convolutions of exponential density functions are given, among other places, in Akkouchi (2008).

Inequality (8) can be proved by a forward recursion. Starting at $n = 1$, the condition holds since $a_1(t) = e^{-\kappa_1 t} \leq 1$ for $t \geq 0$. We now assume that the inequality holds for an $(n - 1)$-factor structure. We infer that

$$\sum_{j=1}^{n} a_j^n(t) = e^{-\kappa_n t} + \sum_{j=1}^{n-1} \int_0^t \kappa_n e^{-\kappa_n(t-s)} a_j^{n-1}(s) ds$$

$$= e^{-\kappa_n t} + \int_0^t \kappa_n e^{-\kappa_n(t-s)} \sum_{j=1}^{n-1} a_j^{n-1}(s) ds.$$

Since $\sum_{j=1}^{n-1} a_j^{n-1}(s) < 1$ for all $s > 0$, we have

$$\sum_{j=1}^{n} a_j^n(t) \leq e^{-\kappa_n t} + \int_0^t \kappa_n e^{-\kappa_n(t-s)} ds = 1.$$

We conclude that the inequality holds for all $n$. 39
A.3. Proof of Proposition 3

Duffie and Kan (1996) has shown that the zero-coupon bond values are exponential-affine in the state vector under affine dynamics. Our model satisfies the affine conditions. The analytical solutions for the affine coefficients in (19) and (20) can be directly solved from the ordinary differential equations in (16) and (17). They can also be derived from the integral form of the instantaneous interest rate dynamics.

We infer from (4) that

\[
\int_0^\tau r_t \, dt = \theta_\tau + \sum_{j=1}^n (x_{j,0} - \theta_r) \int_0^\tau a_j(t) \, dt + \sum_{j=1}^n \sigma_j \left[ \int_0^\tau a_j(t-s) \, dt \right] dW_{j,s}.
\]

Let

\[
b_j(\tau) = \int_0^\tau a_j(t) \, dt = \sum_{i=j}^n \alpha_{i,j} \left( 1 - e^{-\kappa_i \tau} \right).
\]  

Equation (5) implies that \( \int_0^\infty a_j(\tau) d\tau = 1/\kappa_j \). By (59), we infer that \( \sum_{i=j}^n \alpha_{i,j} = 1/\kappa_j \). Hence,

\[
b_j(\tau) = \frac{1}{\kappa_j} - \sum_{i=j}^n \alpha_{i,j} e^{-\kappa_i \tau}.
\]  

Equation (5) implies that \( \int_0^\infty a_j(\tau) d\tau = 1/\kappa_j \). By (59), we infer that \( \sum_{i=j}^n \alpha_{i,j} = 1/\kappa_j \). Hence,

\[
b_j(\tau) = \frac{1}{\kappa_j} - \sum_{i=j}^n \alpha_{i,j} e^{-\kappa_i \tau}.
\]  

We have,

\[
\int_0^\tau r_t \, dt = b(\tau)^\top X_0 + \theta_r \left[ \tau - \sum_{j=1}^n b_j(\tau) \right] + \sum_{j=1}^n \sigma_j \int_0^\tau b_j(\tau-s) \, dW_{j,s}.
\]

The price of a zero-coupon bond with maturity \( \tau \) is therefore:

\[
P(X_0, \tau) = \mathbb{E}^Q \exp \left( - \int_0^\tau r_t \, dt \right) = \exp \left[ -c(\tau) - b(\tau)^\top X_0 \right]
\]
where
\[
c(\tau) = \theta_r \left[ \tau - \sum_{j=1}^{n} b_j(\tau) \right] - \sum_{j=1}^{n} \gamma_j \sigma_j^2 \int_{0}^{\tau} b_j(s) ds - \sum_{j=1}^{n} \frac{\sigma_j^2}{2} \int_{0}^{\tau} b_j^2(s) ds.
\]

We now successively compute each of the three terms on the right-hand side.

First, we verify that
\[
\sum_{j=1}^{i} \alpha_{i,j} = \frac{\kappa_i}{\kappa_1} \alpha_{i,1}.
\]

Hence
\[
\sum_{j=1}^{n} b_j(\tau) = \sum_{i=1}^{n} \sum_{j=i}^{n} \alpha_{i,j} (1 - e^{-\kappa_i \tau}) = \sum_{i=1}^{n} \sum_{j=1}^{i} \alpha_{i,j} (1 - e^{-\kappa_i \tau}) = \kappa_1 \sum_{i=1}^{n} \frac{1 - e^{-\kappa_1 \tau}}{\kappa_i}.
\]

Since \(\sum_{i=1}^{n} \alpha_{i,1} = 1/\kappa_1\), we infer that
\[
\theta_r \left[ \tau - \sum_{j=1}^{n} b_j(\tau) \right] = \kappa_1 \theta_r \sum_{i=1}^{n} \frac{\alpha_{i,1} (1 - e^{-\kappa_i \tau})}{\kappa_i}.
\]

Second, we observe that
\[
\int_{0}^{\tau} b_j(s) ds = \sum_{i=1}^{n} \alpha_{i,j} \left( \tau - \frac{1 - e^{-\kappa_i \tau}}{\kappa_i} \right),
\]
and therefore
\[
\sum_{j=1}^{n} \gamma_j \sigma_j^2 \int_{0}^{\tau} b_j(s) ds = \sum_{j=1}^{n} \gamma_j \sigma_j^2 \sum_{i=j}^{n} \alpha_{i,j} \left( \tau - \frac{1 - e^{-\kappa_i \tau}}{\kappa_i} \right).
\]

Third,
\[
\int_{0}^{\tau} b_j^2(s) ds = \int_{0}^{\tau} \left[ \sum_{i=j}^{n} \alpha_{i,j} (1 - e^{-\kappa_i s}) \right]^2 ds
\]
\[
= \sum_{i=j}^{n} \sum_{k=j}^{n} \alpha_{i,j} \alpha_{k,j} \int_{0}^{\tau} (1 - e^{-\kappa_i s}) (1 - e^{-\kappa_k s}) ds.
\]

We conclude that the Proposition holds.
A.4. Proof of Proposition

A.5. Proof that the response functions are translated versions of each other

Let \( \{E_n\} \) denote a sequence of independent, exponentially distributed random variables. The p.d.f. of \( E_n \) is \( \kappa_n e^{-\kappa_n x} \) for all \( x > 0 \). Let \( Z_{j,n} = E_j + \ldots + E_n \). The Fourier transform of \( Z_{j,n} \) is

\[
\mathbb{E} e^{-2i\pi\xi Z_{j,n}} = \prod_{j'=j}^{n} \frac{\kappa'}{\kappa' + 2i\pi\xi}.
\]

For a fixed \( j \),

\[
\ln \left( \mathbb{E} e^{-2i\pi\xi Z_{j,n}} \right) = -\sum_{j'=j}^{n} \ln \left( 1 + \frac{2i\pi\xi}{\kappa'} \right)
\]

has a limit when \( n \to \infty \) if \( \sum_{j'=j}^{\infty} \frac{1}{\kappa'} < \infty \). We then infer that the random variable \( Z_{j,n} = E_j + \ldots + E_n \) and the function \( K_j \ast \ldots \ast K_n \) both have well-defined limits as \( n \to \infty \).

We now consider that \( j \to \infty \) and \( n - j \to \infty \). Equation (62) implies that

\[
\ln \left( \mathbb{E} e^{-2i\pi\xi \kappa_j Z_{j,n}} \right) \approx -2i\pi\xi \kappa_j \sum_{j'=j}^{n} \frac{1}{\kappa'} - 2i\pi^2\xi^2 \kappa_j^2 \sum_{j'=j}^{n} \frac{1}{\kappa'^2}.
\]

When \( \kappa_j = \kappa_j b^{-1} \) for all \( j \), we know that

\[
\ln \left( \mathbb{E} e^{-2i\pi\xi \kappa_j Z_{j,n}} \right) \to \frac{2i\pi\xi}{1 - b^{-1}} - \frac{2i\pi^2\xi^2}{1 - b^{-2}}.
\]

Hence

\[
\kappa_j Z_{j,n} \to \mathcal{N} \left( \frac{1}{1 - b^{-1}}, \frac{1}{1 - b^{-2}} \right).
\]

The p.d.f. of \( \kappa_j Z_{j,n} \) is \( \kappa_j^{-1} K_j \ast \ldots \ast K_n (\tau/\kappa_j) \). Hence

\[
a_j(\tau) \approx \phi \left( \kappa_j \tau; \frac{1}{1 - b^{-1}}, \frac{1}{1 - b^{-2}} \right),
\]

where \( \phi (\cdot; \mu, \sigma^2) \) denotes the pdf of a normal with mean \( \mu \) and variance \( \sigma^2 \). In Figure 1, the response functions...
are therefore translated versions of each other when $\tau$ is in log scale.
References


Table 1
Summary statistics of LIBOR and swap rates
The data consist of weekly observations (Wednesday closing mid-quotes) on LIBOR at maturities of one, two, three, six, nine, and 12 months, and swap rates at maturities of two, three, four, five, seven, ten, 15, 20, and 30 years. Each series contains 678 weekly observations from January 4, 1995 to December 26, 2007. Entries report the sample average (Mean), standard deviation (Std), skewness (Skew), excess kurtosis (Kurt), and weekly autocorrelations of orders one, five, 10, and 20, respectively, for each series.

<table>
<thead>
<tr>
<th>Maturity</th>
<th>Mean</th>
<th>Std</th>
<th>Skew</th>
<th>Kurt</th>
<th>Autocorrelation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>1 m</td>
<td>4.335</td>
<td>1.798</td>
<td>-0.714</td>
<td>-1.050</td>
<td>0.998</td>
</tr>
<tr>
<td>2 m</td>
<td>4.370</td>
<td>1.803</td>
<td>-0.722</td>
<td>-1.038</td>
<td>0.998</td>
</tr>
<tr>
<td>3 m</td>
<td>4.405</td>
<td>1.808</td>
<td>-0.720</td>
<td>-1.025</td>
<td>0.998</td>
</tr>
<tr>
<td>6 m</td>
<td>4.475</td>
<td>1.803</td>
<td>-0.713</td>
<td>-0.970</td>
<td>0.998</td>
</tr>
<tr>
<td>9 m</td>
<td>4.547</td>
<td>1.789</td>
<td>-0.689</td>
<td>-0.910</td>
<td>0.997</td>
</tr>
<tr>
<td>12 m</td>
<td>4.631</td>
<td>1.769</td>
<td>-0.653</td>
<td>-0.854</td>
<td>0.996</td>
</tr>
<tr>
<td>2 y</td>
<td>4.877</td>
<td>1.570</td>
<td>-0.529</td>
<td>-0.699</td>
<td>0.994</td>
</tr>
<tr>
<td>3 y</td>
<td>5.093</td>
<td>1.414</td>
<td>-0.407</td>
<td>-0.663</td>
<td>0.992</td>
</tr>
<tr>
<td>4 y</td>
<td>5.260</td>
<td>1.298</td>
<td>-0.291</td>
<td>-0.693</td>
<td>0.991</td>
</tr>
<tr>
<td>5 y</td>
<td>5.395</td>
<td>1.209</td>
<td>-0.187</td>
<td>-0.748</td>
<td>0.990</td>
</tr>
<tr>
<td>7 y</td>
<td>5.595</td>
<td>1.091</td>
<td>-0.023</td>
<td>-0.850</td>
<td>0.988</td>
</tr>
<tr>
<td>10 y</td>
<td>5.798</td>
<td>0.994</td>
<td>0.126</td>
<td>-0.949</td>
<td>0.987</td>
</tr>
<tr>
<td>15 y</td>
<td>6.009</td>
<td>0.909</td>
<td>0.228</td>
<td>-1.020</td>
<td>0.986</td>
</tr>
<tr>
<td>20 y</td>
<td>6.103</td>
<td>0.870</td>
<td>0.254</td>
<td>-1.020</td>
<td>0.985</td>
</tr>
<tr>
<td>30 y</td>
<td>6.136</td>
<td>0.851</td>
<td>0.295</td>
<td>-0.949</td>
<td>0.986</td>
</tr>
<tr>
<td>Average</td>
<td>5.135</td>
<td>1.398</td>
<td>-0.316</td>
<td>-0.896</td>
<td>0.992</td>
</tr>
</tbody>
</table>
Table 2
Parameter estimates, standard errors, and log likelihoods.
Entries report the maximum likelihood estimates and their standard errors (in parentheses) of the model parameters. Each row represents one set of parameter estimates under the assumption of \( n \) frequency components, with \( n = 1, 2, \ldots, 15 \). The column under \( L \) reports the maximized aggregate log likelihood value for each model. The last column under \( \mathcal{V} \) reports the Vuong likelihood ratio test statistics between the 15-factor model and the other 14 models. Asymptotically, the statistic has a standard normal distribution.

<table>
<thead>
<tr>
<th>( n )</th>
<th>( \kappa_r )</th>
<th>( \theta_r )</th>
<th>( \sigma_r )</th>
<th>( \theta_i^2 )</th>
<th>( b )</th>
<th>( \sigma_i^2 )</th>
<th>( L )</th>
<th>( \mathcal{V} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.2092</td>
<td>0.0436</td>
<td>0.0065</td>
<td>0.0688</td>
<td>0.0000</td>
<td>0.1574</td>
<td>4086</td>
<td>47.91</td>
</tr>
<tr>
<td>2</td>
<td>0.0603</td>
<td>0.0131</td>
<td>0.0111</td>
<td>0.0695</td>
<td>6.1387</td>
<td>0.0187</td>
<td>13967</td>
<td>30.79</td>
</tr>
<tr>
<td>3</td>
<td>0.0526</td>
<td>0.0000</td>
<td>0.0101</td>
<td>0.0662</td>
<td>7.3138</td>
<td>0.0047</td>
<td>19928</td>
<td>20.70</td>
</tr>
<tr>
<td>4</td>
<td>0.0366</td>
<td>0.0000</td>
<td>0.0116</td>
<td>0.0653</td>
<td>4.2707</td>
<td>0.0019</td>
<td>23276</td>
<td>17.33</td>
</tr>
<tr>
<td>5</td>
<td>0.0441</td>
<td>0.0000</td>
<td>0.0125</td>
<td>0.0507</td>
<td>2.8266</td>
<td>0.0010</td>
<td>25551</td>
<td>15.99</td>
</tr>
<tr>
<td>6</td>
<td>0.0383</td>
<td>0.0000</td>
<td>0.0123</td>
<td>0.0497</td>
<td>3.0267</td>
<td>0.0005</td>
<td>27527</td>
<td>10.60</td>
</tr>
<tr>
<td>7</td>
<td>0.0283</td>
<td>0.0000</td>
<td>0.0129</td>
<td>0.0419</td>
<td>2.6150</td>
<td>0.0004</td>
<td>27898</td>
<td>11.93</td>
</tr>
<tr>
<td>8</td>
<td>0.0275</td>
<td>0.0000</td>
<td>0.0133</td>
<td>0.0632</td>
<td>2.5271</td>
<td>0.0004</td>
<td>28445</td>
<td>11.00</td>
</tr>
<tr>
<td>9</td>
<td>0.0278</td>
<td>0.0000</td>
<td>0.0141</td>
<td>0.0650</td>
<td>2.2351</td>
<td>0.0003</td>
<td>28801</td>
<td>9.18</td>
</tr>
<tr>
<td>10</td>
<td>0.0313</td>
<td>0.0000</td>
<td>0.0140</td>
<td>0.0507</td>
<td>2.2010</td>
<td>0.0003</td>
<td>28972</td>
<td>6.68</td>
</tr>
<tr>
<td>11</td>
<td>0.0305</td>
<td>0.0000</td>
<td>0.0144</td>
<td>0.0966</td>
<td>1.9603</td>
<td>0.0003</td>
<td>29036</td>
<td>6.06</td>
</tr>
<tr>
<td>12</td>
<td>0.0359</td>
<td>0.0000</td>
<td>0.0147</td>
<td>0.0876</td>
<td>1.9130</td>
<td>0.0002</td>
<td>29194</td>
<td>4.41</td>
</tr>
<tr>
<td>13</td>
<td>0.0383</td>
<td>0.0000</td>
<td>0.0149</td>
<td>0.0833</td>
<td>1.8953</td>
<td>0.0002</td>
<td>29283</td>
<td>3.33</td>
</tr>
<tr>
<td>14</td>
<td>0.0409</td>
<td>0.0000</td>
<td>0.0151</td>
<td>0.0781</td>
<td>1.8757</td>
<td>0.0002</td>
<td>29332</td>
<td>2.32</td>
</tr>
<tr>
<td>15</td>
<td>0.0572</td>
<td>0.0000</td>
<td>0.0156</td>
<td>0.0559</td>
<td>1.7400</td>
<td>0.0002</td>
<td>29377</td>
<td>—</td>
</tr>
</tbody>
</table>
Table 3
Summary statistics of pricing errors
Entries report the summary statistics of the pricing errors on the LIBOR and swap rates from the power-law scaled cascade term structure models with three (panel A) and 15 (panel B) factors, respectively. The pricing errors are measured as the difference in basis points between the observed interest rates and the model-implied fair values. The statistics include the sample average of the error (Mean), root mean squared error (Rmse), the first-order weekly autocorrelation of the error (Auto), the maximum absolute error (Max), and the explained variation (VR) (in percentages), defined as one minus the ratio of the pricing error variance to the variance of the original interest rate series.

<table>
<thead>
<tr>
<th>Model</th>
<th>A. Three-factor model</th>
<th>B. 15-factor model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maturity</td>
<td>Mean</td>
<td>Rmse</td>
</tr>
<tr>
<td>1 m</td>
<td>-0.68</td>
<td>7.47</td>
</tr>
<tr>
<td>2 m</td>
<td>0.63</td>
<td>3.82</td>
</tr>
<tr>
<td>3 m</td>
<td>1.61</td>
<td>5.03</td>
</tr>
<tr>
<td>6 m</td>
<td>0.39</td>
<td>6.78</td>
</tr>
<tr>
<td>9 m</td>
<td>-1.74</td>
<td>6.88</td>
</tr>
<tr>
<td>12 m</td>
<td>-3.06</td>
<td>6.74</td>
</tr>
<tr>
<td>2 y</td>
<td>2.11</td>
<td>6.17</td>
</tr>
<tr>
<td>3 y</td>
<td>1.97</td>
<td>6.90</td>
</tr>
<tr>
<td>4 y</td>
<td>0.87</td>
<td>6.32</td>
</tr>
<tr>
<td>5 y</td>
<td>-0.21</td>
<td>5.85</td>
</tr>
<tr>
<td>7 y</td>
<td>-1.89</td>
<td>5.55</td>
</tr>
<tr>
<td>10 y</td>
<td>-2.35</td>
<td>5.17</td>
</tr>
<tr>
<td>15 y</td>
<td>0.88</td>
<td>3.87</td>
</tr>
<tr>
<td>20 y</td>
<td>1.91</td>
<td>5.35</td>
</tr>
<tr>
<td>30 y</td>
<td>-0.76</td>
<td>9.67</td>
</tr>
<tr>
<td>Average</td>
<td>-0.02</td>
<td>6.11</td>
</tr>
</tbody>
</table>
Table 4  
In-sample forecasting performance

Entries report the predictive variation (in percentage points) on each interest rate series over four predicting horizons \( h \) at one, two, three, and four weeks from (i) a first-order autoregressive regression (panel A), (ii) the three-factor cascade model (panel B), and (iii) the 15-factor cascade model (panel C). The predictive variation is defined as one minus the ratio of mean squared predicting error to mean squared interest rate change, which can be regarded as the mean squared predicting error under the random walk hypothesis. All forecasting exercises are performed in sample. The autoregressive coefficients and the model parameters are estimated using the whole sample period. The predicting error statistics are also computed over the whole sample period from January 1995 to December 2007.

<table>
<thead>
<tr>
<th>Model</th>
<th>A. AR(1)</th>
<th>B. Three-factor model</th>
<th>C. 15-factor model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>( h ) (weeks)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LIBOR/swap maturity:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 m</td>
<td>25.85</td>
<td>43.84</td>
<td>57.50</td>
</tr>
<tr>
<td>2 m</td>
<td>23.83</td>
<td>36.65</td>
<td>47.28</td>
</tr>
<tr>
<td>3 m</td>
<td>22.82</td>
<td>32.19</td>
<td>41.34</td>
</tr>
<tr>
<td>6 m</td>
<td>20.85</td>
<td>25.00</td>
<td>31.90</td>
</tr>
<tr>
<td>9 m</td>
<td>20.22</td>
<td>19.35</td>
<td>23.79</td>
</tr>
<tr>
<td>2 y</td>
<td>4.26</td>
<td>7.93</td>
<td>9.65</td>
</tr>
<tr>
<td>3 y</td>
<td>3.64</td>
<td>6.81</td>
<td>8.64</td>
</tr>
<tr>
<td>4 y</td>
<td>4.75</td>
<td>7.23</td>
<td>9.00</td>
</tr>
<tr>
<td>5 y</td>
<td>3.35</td>
<td>6.41</td>
<td>8.36</td>
</tr>
<tr>
<td>7 y</td>
<td>3.44</td>
<td>6.43</td>
<td>8.08</td>
</tr>
<tr>
<td>10 y</td>
<td>3.53</td>
<td>6.19</td>
<td>7.87</td>
</tr>
<tr>
<td>15 y</td>
<td>2.71</td>
<td>5.06</td>
<td>6.72</td>
</tr>
<tr>
<td>20 y</td>
<td>2.42</td>
<td>4.90</td>
<td>6.69</td>
</tr>
<tr>
<td>30 y</td>
<td>3.09</td>
<td>5.44</td>
<td>7.16</td>
</tr>
</tbody>
</table>

53
Table 5  
Out-of-sample forecasting performance

Panel A reports the out-of-sample predictive variation on each interest rate series over four forecasting horizons from a first-order autoregressive regression. The predictive variation is defined as one minus the ratio of mean squared predicting error to mean squared interest rate change. Panel B reports the corresponding out-of-sample predictive variation from the 15-factor model. It also reports the t-statistics on the performance difference between the 15-factor model and the random walk hypothesis. In performing the out-of-sample forecasting exercise, we start from January 7, 1998, re-estimate the model parameters and the autoregressive coefficients at each date \( t \) using the data up to that date, and generate predictions based on estimates on that date. The statistics are computed based on the out-of-sample predicting errors from January 1998 to December 2008.

<table>
<thead>
<tr>
<th>Model</th>
<th>A. AR(1)</th>
<th>B. 15-factor model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistics</td>
<td>Predictive variation</td>
<td>Predictive variation</td>
</tr>
<tr>
<td>h (weeks)</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>LIBOR/swap maturity:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 m</td>
<td>-1.57</td>
<td>-3.50</td>
</tr>
<tr>
<td>2 m</td>
<td>-1.50</td>
<td>-3.64</td>
</tr>
<tr>
<td>3 m</td>
<td>-1.98</td>
<td>-4.19</td>
</tr>
<tr>
<td>9 m</td>
<td>-4.52</td>
<td>-7.90</td>
</tr>
<tr>
<td>12 m</td>
<td>-4.90</td>
<td>-8.45</td>
</tr>
<tr>
<td>2 y</td>
<td>-3.00</td>
<td>-5.91</td>
</tr>
<tr>
<td>3 y</td>
<td>-2.44</td>
<td>-4.78</td>
</tr>
<tr>
<td>4 y</td>
<td>-2.11</td>
<td>-4.13</td>
</tr>
<tr>
<td>5 y</td>
<td>-1.84</td>
<td>-3.54</td>
</tr>
<tr>
<td>7 y</td>
<td>-1.43</td>
<td>-2.74</td>
</tr>
<tr>
<td>10 y</td>
<td>-1.07</td>
<td>-2.06</td>
</tr>
<tr>
<td>15 y</td>
<td>-0.89</td>
<td>-1.68</td>
</tr>
<tr>
<td>20 y</td>
<td>-0.76</td>
<td>-1.52</td>
</tr>
<tr>
<td>30 y</td>
<td>-0.63</td>
<td>-1.30</td>
</tr>
</tbody>
</table>
Table 6
Scaling in risks and risk premiums
Entries report the maximum likelihood estimates and their standard errors (in parentheses) of the model parameters that govern the scaling of risks and risk premiums across the different frequency components in a 15-factor structure.

<table>
<thead>
<tr>
<th>Θ</th>
<th>Estimates</th>
<th>Standard Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_r$</td>
<td>0.0276</td>
<td>(0.0001)</td>
</tr>
<tr>
<td>$\gamma_0 \sigma_r^2$</td>
<td>-0.0019</td>
<td>(0.0001)</td>
</tr>
<tr>
<td>$\gamma_1 \sigma_r^2$</td>
<td>-0.0520</td>
<td>(0.0004)</td>
</tr>
<tr>
<td>$\gamma_2 \sigma_r^2$</td>
<td>-0.1634</td>
<td>(0.0006)</td>
</tr>
<tr>
<td>$b$</td>
<td>1.7276</td>
<td>(0.0032)</td>
</tr>
<tr>
<td>$s_\sigma$</td>
<td>-0.2408</td>
<td>(0.0021)</td>
</tr>
<tr>
<td>$s_0$</td>
<td>0.2532</td>
<td>(0.0111)</td>
</tr>
<tr>
<td>$s_1$</td>
<td>0.0010</td>
<td>(0.0064)</td>
</tr>
<tr>
<td>$s_2$</td>
<td>-1.7617</td>
<td>(0.0038)</td>
</tr>
</tbody>
</table>
Figure 1
LIBOR and swap time series and term structure.
The top panel plots the time series of the 15 LIBOR/swap rate series. The bottom panel plots the term structure at each date.
**Figure 2**

**Factor loading across the term structure.**

The solid line in the top panel plots the mean forward rate curve. The lines in the bottom panel denote the contemporaneous response across the term structure of instantaneous forward rates to unit shocks to each of the 15 factors. The solid line represents the response to the lowest frequency component, the dashed line represents the response to the highest frequency component. The responses to the intermediate frequencies are in dotted lines. To differentiate the different frequencies, we use log scale on the maturity in the bottom panel. Both panels are generated from the estimated 15-factor model.
Figure 3

Term structure of forward rates stripped from LIBOR and swap rates

Lines plot the term structure of forward rates generated from the piece-wise constant assumption in the top panel and from the estimated 15-factor model in the bottom panel.
Figure 4
Cross-correlation between weekly changes in six-month LIBOR and other interest rate series.
Circles denote the cross-correlation estimates between weekly changes in the six-month LIBOR and weekly changes in other interest rate series. The solid line denotes estimates from model values generated from the 15-factor model. The dashed line denotes estimates from model values generated from the three-factor model.
Figure 5
The scaling of $\kappa_j$.
The circles are estimated as free parameters. The solid line is generated from the benchmark model with the scaling $\kappa_i = \kappa_r b^{i-1}$. 