Dimension-Invariant Dynamic Term Structures*

Laurent E. Calvet
HEC Paris and
National Bureau of Economic Research

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

Liuren Wu
Zicklin School of Business
Baruch College

October 5, 2010

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 short-rate cascade, a parsimonious recursive structure that naturally ranks the latent state variables by their rates of mean reversion, each revolving around the next lowest-frequency factor. With appropriate assumptions on volatilities and risk premia, the model overcomes the curse of dimensionality associated with general affine models. Using a panel of 15 LIBOR and swap rates, we estimate models using from one to 15 factors and only five parameters. The in-sample fit of high-dimensional specifications is near exact, with absolute pricing errors averaging less than one basis point, permitting yield-curve stripping in an arbitrage-free, dynamically consistent environment. Cross-maturity correlations accurately reflect empirical evidence, and out-of-sample interest rate forecasts significantly improve on prior benchmarks.

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

Keywords: Term-structure of interest rates; cascade model; dimension-invariance; interest rate forecasting; yield curve stripping; forward rate correlations.

*We thank for helpful comments David Backus, Peter Carr, Anna Cieslak, Antonio Diez de los Rios, Bjorn Flesaker, Mike Gallmeyer, Robert Goldstein, Jeremy Graveline, Dmitry Kreslavskiy, Markus Leippold, Vadim Linetsky, Andrei Lyashenko, Fabio Mercurio, Per Mykland, 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, the University of Chicago, the University of Zurich, Ziff Brothers Investments, the 2010 McGill University Risk Management Conference, the 3rd Triple Crown Conference at Rutgers University, the Financial Econometrics Workshop at the Fields Institute in Toronto, the 2010 International Symposium on Business and Industrial Statistics, the 16th International Conference on Computing in Economics and Finance, the 2010 Econometric Society World Congress, and the 2010 Northern Finance Association Meetings.

1 Introduction

The development of the class of arbitrage-free dynamic term structure models (DTSMs) ranks as one of the most important achievements of modern asset pricing theory. This framework connects the cross-sectional relation between bonds of different maturities to the evolution of the term structure, while maintaining considerable scope for variation in modelling choices. For example the widely-used affine specifications, first characterized by Duffie and Kan (1996), allow very general factor structures to govern term structure dynamics.[1] Theoretical advances in the study of these flexible, tractable models paved the way for an explosion in empirical work, including exhaustive efforts to explore this class and pin down appropriate specifications in pioneering work such as Dai and Singleton (2000) and Duffee (2002).[2]

Despite these efforts, a large subset of DTSMs remains relatively unexplored. In particular, high-dimensional models – meaning roughly those having a larger number of factors than the traditional three – are more difficult to investigate empirically because of the well-known "curse of dimensionality." For example a generic affine three-factor DTSM calls for more than twenty parameters, and specification requirements grow rapidly with the size of the state space. High-dimensional models are therefore generally difficult to identify and estimate.

In this paper, we study *dimension-invariant* DTSMs in which the number of parameters is independent of the size of the state space. We build on the idea that the interest rate term structure responds to shocks of many frequencies, as is evident from the numerous macroeconomic, monetary, and microstructure events that impact bond prices. Many influential studies including Ang and Piazzesi (2003), Piazzesi (2005), Gallmeyer, Hollifield, and Zin (2005), and Diebold, Rudebusch, and Aruba (2006) document these links and relate a wide range of economic time series to models of the term structure.[3] We pursue a standard latent-state variables approach and build an empirically tractable model that permits an arbitrary number of unobservable shocks, each operating at a different frequency and therefore impacting different parts of the term structure.

[1] Important early contributions in the dynamic term structure literature include Vasicek (1977), Brennan and Schwartz (1979), Cox, Ingersoll, and Ross (1985), Constantinides (1992), and Longstaff and Schwartz (1992). Duffie, Pan, and Singleton (2000) and Duffie, Filipović, and Schachermayer (2003) progressively generalize the class of affine models. Leippold and Wu (2002) introduce the quadratic class of arbitrage-free dynamic term structure models.

[2] Other notable empirical contributions include Balduzzi, Das, Foresi, and Sundaram (1996), Dai and Singleton (2002), and Backus, Foresi, Mozumdar, and Wu (2001).

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

The basic building block of our model is a cascade for short-rate dynamics. We posit a lowest-frequency component that provides a central tendency to which the next-lowest factor mean reverts. The dynamics of the remaining factors, which may be arbitrary in number, then follow a recursion whereby each component mean-reverts around the next-lowest frequency in the cascade. We take the highest-frequency element of this cascade to be the short rate, which implies that all lower-frequency factors act as state variables for the term structure. The approach thus substantially extends Balduzzi, Das, and Foresi (1998), who develop a two-factor stochastic central tendency model.

The cascade offers several benefits in implementing high-dimensional models. First, factors naturally separate according to their rates of mean-reversion, eliminating the need to rotate factors and normalize parameters as is typical in general affine models.[4] Second, by adding additional functional form assumptions to specify the progression of mean-reversion rates, volatilities, and risk premia across frequencies, the number of parameters required for identification is independent of the number of factors. In the simplest case, which we take as our base model, the mean-reversion frequencies follow a geometric progression. As few as five parameters then govern the term structure and its dynamics, irrespective the number of factors. This dimension-invariance feature allows us to estimate low and high-dimensional models with equal ease and accuracy.[5]

To assess empirical performance, we use 13 years of data on a broad cross-section of interest-rate securities: 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 cascade models with from one to 15 factors, and find that the 15-factor model significantly outperforms lower-dimensional models both statistically and economically. Notably, the high-dimensional model overcomes several important challenges that have been observed in prior literature. First, Dai and Singleton (2002) show that standard three-factor models generate substantially higher cross-correlations between bonds of different maturities than are observed in the data, implying

[4] Dai and Singleton (2000) provide a complete discussion of factor rotation and normalizations in empirical implementation of DTSMs.

[5] Ease of implementation has not historically been a hallmark of empirical research on DTSMs. For recent discussion of the issues commonly faced, see Duffee and Stanton (2007) and Duffee (2009). Of particular importance is the need to carry out a highly nonlinear optimization over a large parameter space that in many cases is poorly identified. Recently, Joslin, Singleton, and Zhu ("JSZ," 2010) provide a normalization that permits convenient two-step estimation of general affine models. Their approach concentrates out of the likelihood parameters that can be isolated in the mean and autoregressive dynamics of observable factors under the objective density. The method offers the greatest computational benefit when there are no restrictions on risk premia or objective-density factor dynamics, since in this case the mean and autoregressive parameters under the risk-neutral and objective densities are unrelated. In our case where risk premia and factor dynamics are constrained in order to obtain dimension-invariance of the parameter vector, estimation is already straightforward and there is no particular computational benefit of the JSZ normalization. Their approach thus provides an advance in the implementation of maximally flexible affine DTSMs, among other contributions, while we develop dimension-invariant specifications and study empirically tractable high-dimensional DTSMs.

that sources of cross-sectional variation are not captured. Our high-dimensional DTSMs overcome this difficulty, accurately reflecting the cross-correlations observed in the data. Second, Duffee (2002), Ang and Piazzesi (2003), Bali, Heidari, and Wu (2009) and others show that in many cases traditional DTSMs forecast future interest rates no better than a simple random walk. The 15-factor model significantly outperforms both the random walk and an autoregressive specification in predicting interest rate movements from one to twelve-month maturities. Third, the mean absolute pricing errors of the 15-factor model are less than one basis point, an order of magnitude smaller than its three-factor counterpart and very close to the range of common bid-ask spreads.

The cascade model is unique in the literature in offering a cross-sectional fit that is for all practical purposes perfect, while also maintaining dynamic consistency and arbitrage-free pricing. In comparison, while the fitting errors of traditional three-factor models can be small relative to observed yields, they are nonetheless typically large relative to the tiny bid ask spreads in the highly liquid fixed income markets. In many applications such as fixed-income option pricing, this mismatch is problematic because derivative contracts are written on observed rather than model-implied interest rates. As a consequence the option-pricing performance of standard arbitrage-free DTSMs is poor as shown by Dai and Singleton (2002), Li and Zhao (2006), and Heidari and Wu (2009). Alternative approaches, such as the forward rate models of Ho and Lee (1986), Hull and White (1993), and Heath, Jarrow, and Morton (1992), take observed interest rates as given and price options based exclusively on the volatility specification. The forward-rate models begin by fitting the current cross-section perfectly, but typically imply strong restrictions on dynamics leading to frequent recalibrations when the implications are not realized. Ideally, one would like a single unified framework to match the cross-section perfectly while also maintaining dynamic consistency. The high-dimensional cascade achieves these objectives.[6]

One particularly important application of our model is yield-curve "stripping." In the fixed income markets, financial institutions including banks and the Federal Reserve use stripping procedures to obtain the zero-coupon term structure and forward-rate curve from observed Treasury or swap rates. Stripped forward curves are a necessary building block for pricing a variety of instruments, and are needed for example as an input when using a forward-rate model. Common methods of stripping include ad hoc piecewise linear approximation and the "yield-curve models," developed by McCulloch (1975), Nelson and Siegel (1987), Svensson (1995) and others. A problem, however, is that standard yield-curve models are

[6]We show how the model can be combined with volatility specifications in the final section of the paper, but leave empirical investigation of this extension for future research.

not consistent with absence of arbitrage (Filipović (1999)). Recent literature therefore places new emphasis on arbitrage-free approaches to yield-curve stripping, as in Christensen, Diebold, and Rudebusch (2008). An additional difficulty for standard stripping procedures is that the loss of information from compressing a large number of maturities into a low-dimensional factor structure can be significant (Cochrane and Piazzesi (2008)). Our cascade model overcomes these problems, and is therefore an ideal candidate to provide stripped forward-rate curves. The high dimensionality accommodates flexibility and near-perfect fit across the diverse term-structure shapes encountered in the data. Moreover, the smooth stripped term structure and forward rates of the model are consistent with absence of arbitrage.

Relating to prior work, the extreme parsimony of our model and invariance of the number of parameters to the size of the state space builds on Calvet and Fisher (2001, 2004, 2007, 2008). This earlier research models the volatility term structure using a multiplicative cascade, whereas the additive stochastic-tendency cascade we develop here fits interest rate dynamics and provides analytical tractability.

Our paper advances the concept of "dimension-invariance" in two ways. First, we develop term-structure models in which the number of parameters is invariant to the number of factors. Second, we provide conditions under which, for a given *fixed* parameter vector, the sequence of DTSMs obtained by increasing the number of factors n weakly converges to a well-defined, infinite-dimensional DTSM. We thus build a bridge between standard, finite-dimensional DTSMs and the non-degenerate, infinite-dimensional term structures developed using the mathematics of random fields by Kennedy (1994, 1997), Goldstein (2000), and Santa-Clara and Sornette (2001).[7] Our models are therefore simple and parsimonious, but yet can approximate arbitrarily closely the rich dynamics of infinite-state frameworks.

The remainder of the paper is organized as follows. Section 2 develops the general cascade term structure model. Section 3 provides assumptions that produce dimension invariance. Section 4 describes the data, estimation methodology, and estimation results. Section 5 compares the performance of low- and high-dimensional models. Section 6 carries out a variety of specification tests. Section 7 discusses dimension-invariant extensions to stochastic volatility and time-varying risk premia. Section 8 concludes. All proofs are in the Appendix.

[7] See also Collin-Dufresne and Goldstein (2003). Related empirical applications include Longstaff, Santa-Clara, and Schwartz (2001) and Han (2007).

2 A Cascade Model of the Interest Rate Term Structure

We consider a filtered complete probability space $\{\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\geq 0}\}$ that satisfies the usual technical conditions. Let $P(t,\tau)$ denote the time-t value of a zero-coupon bond with one dollar par value and expiry date $t+\tau$ and let r_t denote the instantaneous interest rate defined by continuity:

$$r_t \equiv \lim_{\tau \downarrow 0} \frac{-\ln P(t,\tau)}{\tau}. \tag{1}$$

2.1 The multifrequency cascade

We model the short-term interest rate under the statistical measure $\mathbb{P}$ via an n-factor cascade. Let $W_t = (W_{1,t}, \cdots, W_{n,t})^\top$ denote a standard Wiener process with independent components, and let θ_r denote the long-run level of the short rate. We specify the factors by $x_{0,t} = \theta_r$ and the cascade of diffusions:

$$dx_{j,t} \quad = \quad \kappa_j (x_{j-1,t} - x_{j,t})\, dt + \sigma_j dW_{j,t}, \quad j = 1, 2, \cdots, n. \tag{2}$$

The first component $x_{1,t}$ is a standard mean-reverting diffusion, revolving around the long-run mean θ_r. Each level $x_{j,t}$ higher in the cascade mean-reverts around the level $x_{j-1,t}$ below. The short interest rate is set equal to the last factor:

$$r_t \quad = \quad x_{n,t}. \tag{3}$$

The parameters κ_j control the speed of adjustment of the factors. We assume that $\kappa_1 < \kappa_2 < \cdots < \kappa_n$, so the factors have shorter degrees of persistence as j increases.

We denote by $X_t \equiv (x_{1,t} \cdots x_{n,t})^\top$ the n-dimensional state vector. The statistical dynamics in matrix notation are

$$dX_t = \kappa(\theta - X_t)\, dt + \Sigma^{1/2} dW_t, \tag{4}$$

where the mean-reversion speed matrix κ has diagonal elements $\kappa_{j,j} = \kappa_j$, the off-diagonal elements $\kappa_{j,j-1} = -\kappa_j$, and all other elements are zero. The long-run mean vector θ has identical elements equal to θ_r, and Σ is the diagonal matrix with elements $\sigma_1^2, ..., \sigma_n^2$. We consider stochastic volatility extensions in Section 7.

In general hidden-state models such as the affine models of Duffie and Kan (1996) and the quadratic models of Leippold and Wu (2002), the mean-reversion and covariance matrices are constrained only by technical conditions. Many parameters are not easily identifiable, and factors can rotate making their economic meaning ambiguous. As a consequence, researchers must carry out careful specification analysis as

in Dai and Singleton (2000) to resolve identification issues. The cascade structure in (2), or equivalently the block diagonal form of κ in (4), eliminates factor rotation by ranking the factors based on their frequencies.

To understand how the different frequency components contribute to movements of the instantaneous interest rate, we derive an alternative representation as a sum of stochastic integrals.

Proposition 1 ***(Factor representation of the short rate).*** *Under the n-factor cascade in (4), the instantaneous interest rate can be written as the sum of weighted integrals of previous shocks from the n frequencies,*

$$r_t = \theta_r + \sum_{j=1}^{n} a_j(t)(x_{j,0} - \theta_r) + \sum_{j=1}^{n} \sigma_j \int_0^t a_j(t-s) dW_{j,s}. \tag{5}$$

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

$$a_j(\tau) = (K_j * ... * K_n)(\tau)/\kappa_j, \tag{6}$$

where $*$ denotes the convolution operation, and $K_i(\tau) = \kappa_i e^{-\kappa_i \tau}$ if $\tau \geq 0$ and $K_i(\tau) = 0$ if $\tau < 0$.

A response function $a_j(\tau)$ quantifies the impact of a unit shock from factor j at time $t-\tau$ on the instantaneous interest rate $r_t = x_{n,t}$. We observe that $a_j(\tau)$ depends on both j and n, but for expositional simplicity we keep this dependence implicit throughout the main text.

The response functions $a_j(\tau)$ are rescaled convolution products of exponential densities, and therefore take positive values for every $\tau \in (0, \infty)$. The function $a_n(\tau) = e^{-\kappa_n \tau}$, which quantifies the response to shocks from the highest frequency component $W_{n,s}$, starts at one when $\tau = 0$ and decays exponentially with the time horizon, where the decay speed is governed by κ_n. This property is consistent with the assumption that the short rate is equal to the n^{th} factor: $r_{n,0} = x_{n,0}$.

The response to the $(n-1)^{th}$ factor is determined by convolving two exponential densities:

$$a_{n-1}(\tau) = \frac{(K_{n-1} * K_n)(\tau)}{\kappa_{n-1}} = \frac{\kappa_n}{\kappa_n - \kappa_{n-1}} \left(e^{-\kappa_{n-1}\tau} - e^{-\kappa_n \tau} \right). \tag{7}$$

The response at $\tau = 0$ is zero, $a_{n-1}(0) = 0$, indicating that lower-frequency components are not immediately incorporated into the short rate $r_t = x_{n,t}$. The response function increases for a period of time, reaches a maximum at $\overline{\tau}_{n-1} = (\ln \kappa_n - \ln \kappa_{n-1})/(\kappa_n - \kappa_{n-1})$, and then decreases toward zero.

The response functions $a_j(\tau)$ corresponding to shocks of lower frequencies are obtained through convolutions of more exponential densities, which can be solved in closed form.

Proposition 2 ***(Response functions).*** *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:*

$$a_j(\tau) = \sum_{i=j}^{n} \alpha_{i,j} \kappa_i e^{-\kappa_i \tau}, \quad \textit{where} \quad \alpha_{i,j} = \frac{\kappa_j \cdots \kappa_n}{\kappa_i \kappa_j \prod_{k=j,k\neq i}^{n} (\kappa_k - \kappa_i)}. \tag{8}$$

Furthermore,

$$0 \leq \sum_{j=1}^{n} a_j(\tau) \leq 1 \tag{9}$$

for all $\tau \geq 0$.

The humped shape of the response functions $a_j(\tau)$, $j < n$, is a direct result of the cascade. Instantaneously, only the highest-frequency shock $W_{n,t}$ enters the short rate. A lower-frequency shock impacts the short-rate only by first affecting dynamics of the next highest-frequency factor, which then must impact the next component, and so on. As a consequence, lower levels of the cascade impact the short rate at progressively longer horizons. Consequently, each factor should drive different segments of the term structure.

2.2 Risk premia and the term structure of interest rates

We assume that there are no arbitrage opportunities, which implies the existence of a risk-adjusted measure $\mathbb{Q}$ under which zero-coupon bond values are given by $P(t,\tau) = \mathbb{E}_t^{\mathbb{Q}}[\exp(-\int_t^{t+\tau} r_s ds)]$. We allow each factor to have its own potentially time-varying price of risk $\gamma_{j,t}$, and specify $\mathbb{Q}$ by the Radon-Nikodým derivative:

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

The state dynamics of the factors $x_{j,t}$ under the risk-neutral measure are therefore:

$$dx_{j,t} = -\gamma_{j,t}\sigma_j^2 dt + \kappa_j (x_{j-1,t} - x_{j,t})\, dt + \sigma_j dW_{j,t}^{\mathbb{Q}}, \tag{10}$$

following from Girsanov's Theorem.

We permit risk premia to be affine in X_t:

$$\gamma_{j,t} = \gamma_j + \lambda_j^{\top} X_t, \tag{11}$$

where γ_j is a scalar and λ_j an $n \times 1$ column vector, which provides for convenient pricing. The dynamics of the state X_t can be written in vector notation:

$$dX_t = (\mu - \kappa^* X_t)dt + \Sigma^{1/2} dW_t^{\mathbb{Q}},$$

where $\mu = (\kappa_1\theta_r - \gamma_1\sigma_1^2, -\gamma_2\sigma_2^2, \cdots, -\gamma_n\sigma_n^2)^\top$, Λ is a matrix with row vectors $\lambda_1^\top, \cdots, \lambda_n^\top$, and $\kappa^* = \kappa + \Sigma\Lambda$. The state dynamics can equivalently be written:

$$dX_t = \kappa^*(\theta^{\mathbb{Q}} - X_t)dt + \Sigma^{1/2}dW_t^{\mathbb{Q}}, \tag{12}$$

where $\theta^{\mathbb{Q}} = (\kappa^*)^{-1}\mu$ is the long-run mean of the state vector under $\mathbb{Q}$.

Bond prices are easily solved.

Proposition 3 ***(Zero-coupon bond prices).*** *The price at date t of a zero-coupon bond with maturity τ is given by*

$$P(X_t, \tau) = \exp[-b(\tau)^\top X_t - c(\tau)]. \tag{13}$$

The coefficients $b(\tau)$ and $c(\tau)$ satisfy the system of ordinary differential equations:

$$b'(\tau) = \mathbf{e}_n - \kappa^{*\top} b(\tau), \tag{14}$$

$$c'(\tau) = b(\tau)^\top \kappa^* \theta^{\mathbb{Q}} - \frac{1}{2} b(\tau)^\top \Sigma b(\tau), \tag{15}$$

with initial conditions $b(\tau) = 0$ and $c(\tau) = 0$, where $\mathbf{e}_n$ denotes a vector with the value one in the nth position and zeros elsewhere.

In the special case of constant risk premia: $\gamma_{j,t} = \gamma_j$ for all j, the functions $b_j(\tau)$ and $c(\tau)$ are given by

$$b_j(\tau) = \int_0^\tau a_j(\tau)d\tau = \sum_{i=j}^{n} \alpha_{i,j}\left(1 - e^{-\kappa_i\tau}\right), \tag{16}$$

$$\begin{aligned} c(\tau) = {} & \theta_r\kappa_1 \sum_{i=1}^{n} \alpha_{i,1}\left(\tau - \frac{1-e^{-\kappa_i\tau}}{\kappa_i}\right) - \sum_{j=1}^{n} \gamma_j\sigma_j^2 \sum_{i=j}^{n} \alpha_{i,j}\left(\tau - \frac{1-e^{-\kappa_i\tau}}{\kappa_i}\right) \\ & - \sum_{j=1}^{n} \frac{\sigma_j^2}{2} \sum_{i=j}^{n}\sum_{k=j}^{n} \alpha_{i,j}\alpha_{k,j}\left(\tau - \frac{1-e^{-\kappa_i\tau}}{\kappa_i} - \frac{1-e^{-\kappa_k\tau}}{\kappa_i} + \frac{1-e^{-(\kappa_i+\kappa_k)\tau}}{\kappa_i+\kappa_k}\right). \end{aligned} \tag{17}$$

Furthermore, the long-run level of the state vector under $\mathbb{Q}$ is $\theta^{\mathbb{Q}} = (\theta_r - \gamma_1\sigma_1^2/\kappa_1, \theta_r - \gamma_1\sigma_1^2 - \gamma_2\sigma_2^2/\kappa_2, ..., \theta_r - \sum_{i=1}^n \gamma_i\sigma_i^2/\kappa_i)^\top$.

The proposition provides simple pricing under general assumptions. In the general case we easily check that the price loading satisfies:

$$b(\tau) = (\kappa^{*\top})^{-1}\mathbf{e}_n - (\kappa^{*\top})^{-1}e^{-\tau\kappa^{*\top}}\mathbf{e}_n.$$

The intercept $c(\tau)$ can then be computed by numerical integration of (15).

When risk premia are constant, the bond price is available analytically.[8] The price loadings $b_j(\tau)$ on the factors $x_{j,t}$ are integrals of the response functions $a_j(\tau)$, consistent with the fact that the response functions describe the effect of a shock to $x_{j,t}$ on future short rates. The intercept $c(\tau)$ has three components. The first is proportional to the long-run mean θ_r of the state vector. The second depends on the risk premia γ_j and therefore represents risk adjustments. The final term is driven by convexity adjustments, appearing as the quadratic form in equation (15). We show in the Appendix that the function $c(\tau)$ simplifies to

$$c(\tau) = y_\infty \tau + \sum_{i=1}^{n} c_i \frac{1-e^{-\kappa_i \tau}}{\kappa_i} - \sum_{j=1}^{n} \frac{\sigma_j^2}{2} \sum_{i=j}^{n} \sum_{k=j}^{n} \alpha_{i,j}\alpha_{k,j} \frac{1-e^{-(\kappa_i+\kappa_k)\tau}}{\kappa_i+\kappa_k}, \tag{18}$$

where

$$y_\infty = \theta_r - \sum_{j=1}^{n} \frac{\sigma_j^2}{\kappa_j^2}\left(\gamma_j \kappa_j + \frac{1}{2}\right) \tag{19}$$

and $c_i = -\theta_r \kappa_1 \alpha_{i,1} + \sum_{j=1}^{i} \alpha_{i,j}\sigma_j^2\left(\gamma_j + 1/\kappa_j\right)$. In the absence of risk premia ($\gamma_j = 0$), the yield on a zero-coupon bond with a distant maturity, $y_\infty = \lim_{\tau\to+\infty}[b(\tau)^\top X_t + c(\tau)]/\tau$, is lower than the long-run level of the short rate, $\theta_r = \lim_{\tau\to+\infty} r_{t+\tau}$, due to the convexity terms $\sigma_j^2/(2\kappa_j^2)$; the average yield curve is correspondingly downward-sloping. If the risk premium coefficients γ_j are sufficiently negative, however, the long run yield y_∞ is larger than θ_r, and the average term structure is upward sloping.

The instantaneous forward rate is affine in the state vector:

$$f(X_t,\tau) \equiv -\frac{\partial \ln P}{\partial \tau}(X_t,\tau) = b'(\tau)^\top X_t + c'(\tau). \tag{20}$$

In the case of constant risk premia, we know from Proposition 3 that the forward rate loadings and the short-rate response coefficients are identical: $b'(\tau) = a(\tau)$, indicating a tight correspondence between the cross-section of the forward rate curve and time-series dynamics of the short rate.[9] If on the other hand we allow the presence of time-varying risk premia $\Lambda \neq 0$, the coefficient $b'(\tau) = e^{-\tau\kappa^{*\top}}\mathbf{e}_n$ generally differs from $a(\tau) = e^{-\tau\kappa^\top}\mathbf{e}_n$. The correlation between forward rates and expected future short rates is therefore

[8] As is noted in the proof of the proposition, closed form pricing is more generally available when the affine risk premia $\gamma_{j,t}$ defined in (11) depend only on lower-frequency state variables $x_{i,t}$, $i \leq j$.

[9] As in standard dynamic term structure models, this result is consistent with the fact that a form of the expectations hypothesis holds when risk premia are constant. Similar to Fama and Bliss (1987), consider regressing the future short rate $r_{t+\tau}$ on a constant and the instantaneous forward rate $f_{t,\tau}$, i.e.

$$r_{t+\tau} = \phi_0 + \phi_1 f_{t,\tau} + \xi_t.$$

The expectations hypothesis holds if $\phi_0 = 0$ and $\phi_1 = 1$ and the biased expectations hypothesis holds if $\phi_0 \neq 0$ with $\phi_1 = 1$. In our model, the forward rate and the expected short rate satisfy:

$$\begin{aligned} \mathbb{E}(r_{t+\tau}|X_t) &= a(\tau)^\top X_t + [1 - \iota^\top a(\tau)]\theta_r, \\ f_{t+\tau} &= b'(\tau)X_t + c'(\tau) \end{aligned}$$

where $\iota = (1,...,1)^\top$. When risk premia are constant, we know that $b'(\tau) = a(\tau)$ and therefore $\phi_1 = 1$. The intercept ϕ_0 generally differs from zero due to both convexity terms and risk premia. See Piazzesi (2010) for a general discussion.

imperfect. Our framework thus allows for considerable flexibility in the link between forward rates and expected future spot rates, depending on the specification of risk premia. Identifying risk premia typically requires longer time series as in Cochrane and Piazzesi (2005), and therefore the benefits of including this added richness in the model will generally depend on the particular application.

3 Dimension-Invariant Term Structures

This section develops a version of the cascade model in which a fixed and finite parameter vector describes the dynamics of the term-structure for an arbitrary number of factors.

3.1 A Dimension-Invariant Specification

When the rates of mean reversion κ_j, volatilities σ_j, and risk premia γ_j are specific to each factor, the n-factor cascade requires $3n+1$ parameters: $(\theta_r, \kappa_1, \cdots, \kappa_n, \sigma_1, \cdots, \sigma_n, \gamma_1, \cdots, \gamma_n)$. In order to obtain a model where the number of parameters is independent of the size of the state space, we choose parsimonious functional forms to describe how κ_j, σ_j, and γ_j depend on the position j in the cascade.

To guide our choice of the frequency specification, we note that the progression of the most liquid maturities observed in the market is approximately geometric. For example, the most commonly quoted maturities for interest rate swaps are at two, three, five, ten, fifteen, and thirty years, with increasingly wider spacing at longer maturities. That is, each liquid maturity is a multiple of between 1.5 and 2 times the next lowest liquid maturity. Given that these most liquid maturities endogenously arise in the market, we may presume that their spacing in some sense reflects an optimal allocation that balances the capture of information across different parts of the yield curve. This in turn suggests that a geometric spacing of frequencies in the cascade could be effective.

Our base model therefore assumes that the sequence of mean reversion rates follows a geometric progression:

$$\kappa_j = \kappa_1 b^{j-1}, \quad j = 1, 2, \cdots, n, \tag{21}$$

where κ_1 determines the mean reversion speed of the lowest frequency component $x_{1,t}$ and the coefficient $b > 1$ controls the spacing between different frequency components. Two parameters thus control the mean reversion speeds of all factors.

To verify our intuition about the regions of the term structure that each factor will affect under (21), we examine the response functions $a_j(\tau)$. These have a particularly appealing form, as shown in Figure 1 for a numerical example with $n = 15$ components having durations ranging from one week to 30 years. Except for the highest frequency which has exponential decay, all response functions are hump-shaped with maxima progressively increasing and approximately evenly spaced on a log scale, consistent with the progression of liquid maturities observed in the market. In the Appendix we explain in greater detail why the $a_j(\tau)$ are approximately translated versions of each other in log scale.

Building on the idea that each of the factors in the model should carry a roughly equal amount of information relative to its neighbors, our second assumption is that factor volatilities are constant across components. Given this choice, it is natural to also assume that risk premia are identical:

$$\sigma_j = \sigma_1, \quad \text{and} \quad \gamma_j = \gamma_1 \ \text{ for all } \quad j = 2, \cdots, n. \tag{22}$$

The long-run level of the first factor under the risk-adjusted measure is $\theta_r^{\mathbb{Q}} = \theta_r - \gamma_1\sigma_1^2/\kappa_1$. In the empirical work, we find it convenient to estimate $\theta_r^{\mathbb{Q}}$ and back out the price of risk

$$\gamma_1 = \kappa_1(\theta_r - \theta_r^{\mathbb{Q}})/\sigma_1^2.$$

The term structure and its dynamics are fully determined for an arbitrary number of factors by five parameters

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

We investigate the empirical performance of this model in Sections 4-5, and test the functional form assumptions in Section 6. In Section 7, we discuss dimension-invariant extensions that permit time-varying risk premia and stochastic volatility.

Under these assumptions, the long-term forward rate

$$f_\infty = y_\infty = \theta_r - \frac{\sigma_1^2}{\kappa_1^2}\left[\gamma_1\kappa_1\frac{1-b^{-n}}{1-b^{-1}} + \frac{1-b^{-2n}}{2(1-b^{-2})}\right], \tag{23}$$

is higher than the short rate if the risk premium γ_1 is sufficiently negative to overcome the convexity effect, $-\gamma_1 > (1+b^{-n})/[2\kappa_1(1+b^{-1})]$. We also observe that for large n,

$$f_\infty \longrightarrow \theta_r - \frac{\sigma_1^2}{\kappa_1^2}\left[\frac{\gamma_1\kappa_1}{1-b^{-1}} + \frac{1}{2(1-b^{-2})}\right].$$

This suggests that the cascade converges, as we now investigate.

3.2 Limiting Behavior

We give conditions under which both the short rate and term structure weakly converge as the number of factors n grows. Using the property of the response functions in (9), we know that $a_j(\tau) \leq 1$ and thus $a_j^2(\tau) \leq a_j(\tau)$. The unconditional variance of the instantaneous interest rate therefore satisfies:

$$Var(r_t) = \sum_{j=1}^{n} \sigma_j^2 \int_0^\infty a_j^2(s)ds \leq \sum_{j=1}^{n} \sigma_j^2 \int_0^\infty a_j(s)ds = \sum_{j=1}^{n} \frac{\sigma_j^2}{\kappa_j}. \tag{24}$$

Using (21) and (22), we can rewrite the upper bound as

$$Var(r_t) \leq \frac{\sigma_1^2}{\kappa_1} \frac{1-b^{-n}}{1-b^{-1}}. \tag{25}$$

Therefore, as long as $\kappa_1 > 0$ and $b > 1$, the unconditional variance of the instantaneous interest rate remains bounded as $n \to \infty$. Consider an arbitrary $T > 0$, and let L^2 denote the Hilbert space of adapted square integrable processes defined on $[0,T]$.

Proposition 4 ***(Convergence of the term structure).*** *Consider a fixed set of parameters $(\kappa_1, b, \sigma_1, \theta_r)$. When the number of factors in the cascade $n \to \infty$, the sequence of short rate processes $r_{n,t} = x_{n,t}$ converges in L^2 to a limit process with continuous sample paths and a finite variance. When the parameter $\theta_r^{\mathbb{Q}}$ is also fixed, the term structure of zero-coupon bond yields also converges in L^2 as $n \to \infty$.*

Convergence holds more generally when the volatilities σ_j^2 are heterogeneous across frequencies and satisfy $\sum_{j=1}^{\infty} \sigma_j^2/\kappa_j < \infty$, under the maintained assumptions that (21) holds and $\gamma_j = \gamma$ for all j.

We emphasize two implications of the convergence result. First, the proposition builds a bridge between standard, finite-factor DTSMs and the infinite-factor models described using the mathematics of random fields by Kennedy (1994,1997), Goldstein (2000), and Santa-Clara and Sornette (2001). This bridge is useful because our model can be implemented using very simple empirical methods, yet approximates arbitrarily closely the rich dynamics of genuinely infinite-state term structures. Second, convergence guarantees that as we add factors a non-degenerate limit does in fact exist. Thus in empirical work, we expect that our parameter estimates and estimated likelihood will converge as we increase the number of factors. We are now in position to estimate with confidence our first dimension-invariant DTSM.

4 Data and Estimation

We estimate the model on a panel 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 January 4, 1995 to December 26, 2007. This gives a total of 678 weekly observations for each series.

4.1 Summary statistics of LIBOR and swap rates

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

$$LIBOR(X_t,\tau) = \frac{100}{\tau}\left(\frac{1}{P(X_t,\tau)} - 1\right), \tag{26}$$

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

$$SWAP(X_t,\tau) = 100h \times \frac{1 - P(X_t,\tau)}{\sum_{i=1}^{h\tau} P(X_t, i/h)}, \tag{27}$$

where τ denotes the swap maturity in years and h denotes the number of payments per year. The swap contracts make semi-annual payments ($h = 2$) and follow 30/360 day-count convention.

In Table 1, we report for each series the sample mean, standard deviation, skewness, kurtosis, and weekly autocorrelations of order 1, 5, 10, and 20. The average term structure is upward-sloping. The interest rates of all maturities exhibit small skewness and excess kurtosis. They are also highly persistent with first-order autocorrelations ranging from 0.9885 to 0.999.

Panel A of Figure 2 illustrate the time series of the interest rates. The short-term LIBOR started at 6% in 1995, varied between 5 and 6% in the 1990s, dropped to about 1% in 2003, and moved upward from mid-2004 to 2006. Panel B plots the term structure of the LIBOR/swap rates at different dates, showing a wide variety of shapes including upward- and downward-sloping, hump-shaped, and flat. The data should therefore provide a meaningful challenge.

4.2 Estimation and likelihood tests

We cast the dimension-invariant DTSM into a state space form and estimate the parameters using quasi-maximum likelihood. The state propagation equation is a discrete-time analog of the statistical dynamics

(4):

$$X_{t+1} = A + \Phi X_t + \sqrt{\Sigma_x}\varepsilon_{t+1}, \tag{28}$$

where $\Delta t = 1/52$, $\Phi = \exp(-\kappa\Delta t)$, I_n denotes an n-dimensional identity matrix, $A = (I_n - \Phi)\theta$, $\{\varepsilon_{t+1}\}$ is i.i.d. $\mathcal{N}(0, I_n)$, and $\Sigma_x = \sigma_1^2 \Delta t I_n$.

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

$$y_t = h(X_t) + e_t, \tag{29}$$

where y_t denotes the data, $h(X_t)$ denotes model values of the LIBOR and swap rates as a function of the state X_t, and e_t denotes a vector of measurement errors. We assume that the measurement errors are normally distributed iid random variables with zero mean and variance σ_e^2.

In systems where the state variables are Gaussian and the measurement equations are linear, the Kalman (1960) filter provides efficient least-squares updating. In our application, the state propagation equation (28) satisfies Gaussian linearity, but the measurement equations are nonlinear functions of the state variables. We therefore use the unscented Kalman filter as described in the Appendix to obtain quasi-maximum likelihood estimates of the parameter vector $(\kappa_1, b, \sigma_1, \theta_r, , \theta_r^{\mathbb{Q}}, \sigma_e^2)^\top$, composed of the five model parameters and the pricing error variance σ_e^2.

Using this procedure, we estimate models in which the number of frequency components n range from one to fifteen. Table 2 reports parameter estimates, standard errors, and the maximized log likelihood for the 15 models. With $n = 1$ 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. Comparing the parameter estimates as n varies helps to build intuition about the model. The estimates for the mean-reversion speed of the lowest frequency component κ_1 are similar across n. The estimation thus identifies low-frequency movements first, and as we add factors higher frequencies are captured as well. The estimated scaling parameter b falls with n, implying a finer frequency spacing as we add factors, which is intuitive. Finally, the measurement error variance σ_e^2 declines with n, suggesting that fit improves as we add factors. Confirming this result, the log likelihood ($\mathcal{L}$) rises monotonically with the number of frequencies. The likelihood increase is rapid initially but levels off as we add more factors, consistent with the weak convergence demonstrated in Section 3.

Focusing on the best-performing 15-factor model, the lowest frequency is $\kappa_1 = 0.0572$, corresponding to a time horizon ($1/\kappa_1$) of about 17.5 years. The scaling coefficient $b = 1.74$ gives the highest frequency

a mean-reversion speed of about two days. The estimate for the statistical long-run mean is close to zero, and the risk-neutral mean is 5.59%. Given the parsimony of the model and the large amount of data, all parameters are estimated with small standard errors in all specifications.

To assess the statistical significance of the improvement in fit shown by the 15-factor model, we use the Vuong (1989) test, a standard method to assess the significance of the likelihood difference between two non-nested models. The Appendix provides details of the test. We calculate the Vuong statistic for the difference in likelihood between the 15-factor model and each of the other 14 models, and report the results in the last column of Table 2. Asymptotically, the statistic has a standard normal distribution, and the large values of the statistics indicate rejection at the 1% level of the hypothesis that the 15-factor model has a lower likelihood than the comparison models. The fit of the 15-factor model is thus significantly better than for all lower-frequency models.

5 Comparing High- versus Low-Dimensional Models

We compare other aspects of the performance of high- versus low-dimensional specifications, both in- and out-of-sample. To simplify exposition, we will refer to the 15-factor cascade as the high-dimensional model and the three-factor cascade as the low-dimensional benchmark.

5.1 In-sample fit and yield curve stripping

Table 3 reports the summary statistics of the pricing errors from the two representative models. The summary statistics of 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 6 basis points. Since the bid-ask spreads for swap rates average around half to one basis point, these pricing errors although small are economically significant. By contrast, the 15-factor model fits observed interest rates to near perfection, with a root mean squared pricing error of less than one basis point and explained variations close to 100% for all series.

An important practical application of our model is to generate stripped 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 and between different observed maturities. Common basis functions 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 (1995)).

Among these approaches, the exponential functional forms of Nelson and Siegel (1987) and Svensson (1995) have become popular choices in the literature and are widely used to generate data used in a variety of studies (e.g., Gurkaynak, Sack, and Wright (2006)). A difficulty, however, is that the original Nelson-Siegel basis functions cannot be made consistent with absence of arbitrage under any interest rate dynamics (Filipović (1999)). In using data generated by such a stripping procedure, Cochrane and Piazzesi (2008) therefore offer the strong caveat that, "this functional form cannot be generated by standard yield curve models. Since the asymptotic ($n \to \infty$) forward rate and yield vary over time, there is an asymptotic arbitrage opportunity."

Christensen, Diebold, and Rudebusch (2008) propose a modification of the original Nelson and Siegel approach to make it consistent with a three-factor dynamic term structure model. While the three-factor structure does capture major variations of the yield curve (Diebold and Li (2006)), the remaining fitting errors can be economically significant and the functional forms overly restrictive in some instances. Indeed, Cochrane and Piazzesi (2008) argue that the loss of information can be substantial even when using the six-factor Svensson model to compress the original data points described by fifteen separate maturities.[10] They further explain that because the fit is approximate, "Yield curves will be evaluated by how well they match the functional form, not necessarily how well they match the underlying data."

Our approach solves all of these problems, permitting near-perfect arbitrage-free yield-curve stripping for an arbitrary number of factors. Following the forward rate expression (20), the basis functions for the forward rate curve are the response functions $\{a_j(\tau)\}_{j=1}^n$. The factors $\{x_{j,t}\}_{j=1}^n$ act as time-varying weights for the basis functions, and the rich state space provides the flexibility to match closely virtually all observed term structure shapes.

Some of the features of a standard stripping approach versus our procedure can be seen in Figure 3. In Panel A, we show the standard method used in industry of assuming a piece-wise constant step function for the forward rate and backing out the levels of the steps sequentially from low to high maturity. The stripped forward rates from our model are shown in Panel B, calculated using the forward-rate function $f(\hat{X}_t, \tau)$,

[10] They comment, "Regressions using 15 maturities on the right hand-side, generated from a six-factor model, are obviously hopeless." Further, "because excess return forecasts imply multiple differences of the underlying price data... small amounts of smoothing have the potential to lose a lot of information in forecasting exercises."

where $\hat{X}_t$ denotes the filtered state values at each date t. While both approaches match the observed LIBOR and swap rates well, our model-generated curves offer important advantages over the piece-wise constant approach. In particular, our forward curves are dynamically consistent and exclude arbitrage opportunities. The piece-wise constant assumption by design matches observed rates exactly. However, the discontinuities in the forward curves indicate potentially large mispricing for any maturities not explicitly used in the construction of the forward curves. These discontinuities can induce instabilities when used as inputs to the forward rate model of Heath, 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 smooth, dynamically consistent forward rate curves that match to near perfection the cross-section of observed LIBOR and swap rates.

5.2 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 has the promise of generating interest rate fair values that match the cross-correlations observed in the data.

To measure cross-correlations between non-overlapping forward rates, one must first strip the swap rates. The estimates would thus depend on the particular stripping method and the basis functional forms. 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 15×15 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 and dashed lines respectively represent correlation estimates from the $n = 15$ and $n = 3$ factor model implied values. We observe that the correlations from the 15-factor model match the

data well, while the 3-factor model generates correlations that are too high, consistent with prior evidence. Thus, by using a high-dimensional structure, we readily overcome a major limitation of low-dimensional DTSMs.

5.3 Interest rate forecasting

Several studies, e.g., Duffee (2002), Ang and Piazzesi (2003), and Bali, Heidari, and Wu (2009), find that low-dimensional DTSMs fare little better, and often worse, than a simple random walk in forecasting interest rate movements. We hypothesize that two limitations of typical DTSMs inhibit forecasting performance. First, general three-factor models involve over twenty parameters, many of which cannot be estimated with statistical significance. Therefore, no matter how well a traditional three-factor model fits the data in-sample, we should expect the out-of-sample performance to deteriorate substantially due to parameter instability. Second, traditional three-factor models do not fit observed interest rates closely in-sample. The fitting errors carry over to out-of-sample forecasts, because all model-based forecasts use as a starting point the fitted yield curve.

The random walk (RW) hypothesis implies that the best forecast of a future spot rate is the current spot rate of the same maturity. While naive, the RW has the advantage of starting from the correct value of the current spot rate. Over a short enough forecasting horizon, it is therefore likely to outperform any model that does not match well the current term structure. Furthermore, the RW hypothesis involves no parameter estimation, so there is no distinction between in- and out-of-sample performance.

The 15-factor cascade seems to address the two main difficulties traditional three-factor DTSMs face in forecasting. First, the cascade is parsimonious and well-identified. This should improve out-of-sample stability, and cause less degradation between in-sample and out-of-sample performance. Second, our high-dimensional cascade fits the yield curve nearly exactly in-sample, and its forecasts start at essentially the same place as the random walk. For these reasons, our cascade is likely to be a strong performer out-of-sample.

To verify these ideas, we compare our fifteen-factor cascade to: (i) its three-factor counterpart, (ii) the random walk, and (iii) a first-order autoregression (AR). To demonstrate the differential effects of in-sample fit versus out-of-sample stability on forecasting performance, we first calculate the prediction errors of each model using in-sample parameter estimates, and then investigate out-of-sample performance.

In-sample analysis. In this subsection, the entire sample period is used to both estimate parameters and calculate prediction errors. For the AR strategy we estimate an AR(1) regression for each interest rate series j for each forecasting horizon h,

$$y_{j,t+h} = a + by_{j,t} + e_{j,t+h}, \quad j = 1, \cdots, 15, \tag{30}$$

where $y_{j,t}$ denotes the time-t observed value of the j-th interest rate (LIBOR or swap rate) series, and the forecasting horizons are $h = 1, 2, 3$, and 4 weeks. We run separate regressions for different maturities j and horizons h, so that the estimated coefficients a and b depend on both j and h. Given the full-sample estimates $(\widehat{a}, \widehat{b})$, we generate predicted values $\overline{y}_{j,t+h} = \widehat{a} + \widehat{b}y_{j,t}$, and define the errors as $\hat{e}_{j,t+h} = y_{j,t+h} - \overline{y}_{j,t+h}$.

For the cascade, we use the filtered state values at each date $\widehat{X}_t$ and the statistical factor dynamics to predict future values of the state over different horizons h,

$$\overline{X}_{t+h} = A_h + \Phi_h \widehat{X}_t, \tag{31}$$

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. We predict values of the LIBOR and swap rates according to (26) and (27), using the predicted the state vector $\overline{X}_{t+h}$.

We compare the in-sample prediction errors $\hat{e}_{j,t+h}$ of each strategy to the errors from the random walk hypothesis, $y_{j,t+h} - y_{j,t}$. 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:

$$PV_j = 1 - \frac{\sum_{t=1}^{N-h} \left(\hat{e}_{j,t+h}\right)^2}{\sum_{t=1}^{N-h} \left(y_{j,t+h} - y_{j,t}\right)^2}. \tag{32}$$

The predictive variation is positive when the strategy has smaller in-sample prediction errors than the random walk.

In Table 4, we report the in-sample predictive variation estimates. Panel A shows the predictive variations for the AR strategy, which are all positive because the AR strategy nests the random walk. The predictive variations range from 17.53% to 68.12% for the LIBOR series, and are about 10% for the less-predictable swap series. By contrast, Panel B shows that the three-factor model performs worse than the random walk at the one-week horizon for all LIBOR and swap rates, with mixed results at two to four week horizons. Panel C reports predictive variations from the 15-factor model. These are close to zero for the random-walk-like swap rates, but show significant improvements on the random walk for all LIBOR series across all four forecasting horizons.

Thus, even in-sample the 3-factor model cannot match predictions from the random walk over short intervals. Since out-of-sample predictive variations can only be worse than in-sample predictive variations because of parameter instability, the results in Table 4 are sufficient to discard the three-factor model as a viable candidate to outperform the random-walk in out-of sample forecasts. The term-structure model will suffer some degradation of performance out-of-sample due to parameter instability while the random walk hypothesis involves no parameter estimation. By contrast, both the AR specifications and the 15-factor cascade provide better in-sample prediction errors than the random walk. However, overfitting may be more of a concern with the AR specifications as we estimate a total of thirty parameters for the 15 interest-rate series, whereas the 15-factor cascade ties together the dynamics of the entire term-structure of interest rates using only five parameters. We therefore continue to evaluate the out-of-sample forecasting performances of both the AR model and the 15-factor cascade relative to the random walk.

Out-of-sample performance. We re-estimate the autoregressive coefficients and the model parameters at each date t using the data up to that date, starting from January 7, 1998. We generate forecasts based on the coefficient estimates on the date t. Table 5 reports the out-of-sample results.

Panel A reports the predictive variation of the AR(1) regression. Although the AR strategy showed the best performance in-sample, its predictive power deteriorates dramatically out-of-sample, indicating parameter instability and in-sample overfitting. The out-of-sample performance is worse than the random walk hypothesis across all interest rate series and over all four forecasting horizons.

Panel B reports the out-of-sample performance of the 15-factor model, showing substantial improvements over the random walk for all LIBOR series over all forecasting horizons. To help explain this good performance, we note that the out-of-sample predictive variations are very close to the corresponding in-sample estimates. This confirms that parameter instability is not a problem for our specification, which uses only five parameters to control the dynamics and term structure of the 15 interest rate series.

To obtain a broader perspective on this forecasting analysis, consider the idea of exploiting information in the cross-section of interest rates by using a general VAR(1). While this approach has a natural appeal, by our prior results we can already see that there is little prospect of finding success in this approach. Overfitting and parameter instability was shown to be a substantial difficulty even when using a set of AR(1) regressions, which dramatically restricts the general VAR approach by shutting down all off-diagonal elements. Obviously, a general VAR(1) system would have too many free parameters to be estimated with any accuracy, and out-of-sample instability would be an even larger problem than with the simple AR(1)

approach.[11] By contrast, our model essentially builds a VAR(1) system on the forward rates that exploits information from the whole term structure, but the model requires only five parameters. The ability to parsimoniously incorporate information from the entire term-structure is a necessity for generating good forecasts, and drives our results.

The 15-factor model succeeds because it is parsimonious with parameters, yet flexibile due to its rich state space. This combination differs from existing literature, which typically specifies few factors with many parameters. From this perspective, our approach reflects an effort to be as simple as possible in aspects that permit it, but not simpler than required to accurately match the data. The combination of a DTSM that is rich in states but thrifty with parameters is thus important to producing excellent in-sample fit and improved out-of sample forecasts.

6 Specification Analysis

In this section, we investigate the empirical validity of the key assumptions in our benchmark dimension-invariant model.

6.1 The frequency specification

To verify the empirical validity of our assumed geometric progression of mean reversion rates in (21), we estimate an extended version of the 15-factor model by letting κ_j be a free parameter for each frequency j. The total number of parameters increases from the original 6 to 19. In Figure 5, we plot the logarithm of the κ_j estimates in circles, and as a solid line the linear relation implied by (21) with $\kappa_1 = 0.0572$ and $b = 1.74$. The estimates for the free parameters κ_j vary around the solid line, suggesting that our frequency specification holds reasonably well.

Using a likelihood ratio test to refine this analysis, restriction (21) is strongly rejected.[12] Nevertheless, the added freedom does not improve forecasting performance. When we try to perform rolling estimation of the unrestricted model, we often experience convergence issues. When we use the in-sample parameter estimates instead to study prediction errors, the grand average of the percentage predicted variation on the

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

[12] The maximized log likelihood values of the unrestricted model is 29,997, whereas the restricted log likelihood is 29,376, which produces a rejection at the 1% level given 13 degrees of freedom.

six LIBOR rates over the 13 different horizons is 26.80%. In comparison the restricted model gives an improvement in predictive variation relative to the random walk of 26.98%, marginally better. Thus, in practical applications the cascade that assumes a geometric progression of frequencies performs at least as well as the cascade with unconstrained frequencies.

6.2 Risks and risk premia

Following the idea that our specification of mean-reversion rates κ_j approximately matches the near-geometric progression of maturities in the most widely quoted fixed-income instruments, our base model makes use of the additional simplifying assumption that the volatilities σ_j are identical across factors. This assumption builds on the idea that each factor in the cascade model should carry a roughly equal amount of information relative to its neighbors. Given this choice, we also assume in our base specification that risk premia are identical across components: $\gamma_j = \gamma_1$.

To investigate the specification of risks and risk premia across frequencies, we extend the dimension-invariant model to permit separate scaling parameters for volatilities and the market premia across factors. We specify the instantaneous variance of the components as

$$\sigma_j^2 = \sigma_1^2 b^{(j-1)s_\sigma}, \tag{33}$$

and we accommodate a more flexible risk premium specification by

$$\gamma_{j,t}\sigma_j^2 = \gamma_{0,j}\sigma_j^2 - \gamma_{1,j}\sigma_j^2 x_{j-1,t} + \gamma_{2,j}\sigma_j^2 x_{j,t}, \tag{34}$$

where the three components satisfy

$$\begin{aligned} \gamma_{0,j}\sigma_j^2 &= \gamma_0\sigma_1^2 b^{(i-1)s_0}, \\ \gamma_{1,j}\sigma_j^2 &= \gamma_1\sigma_1^2 b^{(j-1)s_1}, \\ \gamma_{2,j}\sigma_j^2 &= \gamma_2\sigma_1^2 b^{(j-1)s_2}. \end{aligned} \tag{35}$$

This relaxed specification adds two more risk premium coefficients (γ_1, γ_2) and four additional scaling parameters $(s_\sigma, s_0, s_1, s_2)$. A zero estimate for a scaling exponent, say s_σ, indicates that σ_i does not vary across frequencies, while a positive estimate indicates that the component is larger for higher frequency factors.

Table 6 reports the additional parameter estimates for the model with flexible volatilities and risk premia across factors. All three risk premium coefficients are negative, suggesting that the risk premium increases in the 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 is smaller at higher frequencies. However, the estimates are much smaller than one in absolute magnitude. Thus, the instantaneous variance changes much more slowly across frequencies than the mean reversion speed. The estimate of s_0 is positive, implying that the risk premium is larger in absolute magnitude for higher-frequency factors. Again, however, the estimate is much smaller than one. The estimate of s_1 is virtually zero, and s_2 is negative, suggesting that the risk premium depends on the risk level, but this dependence becomes small at high frequencies.

Overall, the risks and risk premiums do not vary nearly as much across frequencies as does the mean reversion speed. The assumption of stability across frequencies can thus be viewed as a convenient simplification that can be relaxed in particular applications, but offers the benefits of parsimony and robust identification.

6.3 Positive interest rates

The linear Gaussian structure of the benchmark model provides analytical tractability but permits interest rates to have positive probabilities of becoming negative. To exclude negative interest rates, we consider two alternative dimension-invariant specifications.

First, we specify the instantaneous volatility of each factor to be proportional to the square root of the corresponding factor itself, $\sigma_j = \sigma_1\sqrt{x_{j,t}}$. The one-factor version of this model thus corresponds to Cox, Ingersoll, and Ross (1985). We further assume that the market price of risk is proportional to the square root of the risk level $\gamma_1\sqrt{x_{j,t}}$ so that the risk premium is proportional to the risk level, and $\kappa^{\mathbb{Q}}_{j,j} = \kappa_{j,j} + \gamma_1\sigma_1^2$. The spot-rate loading coefficients can no longer be solved analytically, but can be solved numerically from the following set of ordinary differential equations,

$$\begin{aligned} b'(\tau) &= \mathbf{e}_n - (\kappa^{\mathbb{Q}})^\top b(\tau) - \tfrac{1}{2}\sigma_1^2\left[b(\tau) * b(\tau)\right], \\ c'(\tau) &= b_1(\tau)\kappa_1\theta_r, \end{aligned} \tag{36}$$

starting at $b(0) = 0$ and $c(0) = 0$, where $*$ denotes element-by-element multiplication. When we estimate this alternative specification with 15 factors, the maximized log likelihood is 29,377, almost identical to the likelihood of the benchmark linear Gaussian model. The in-sample prediction errors from the six LIBOR rates generates a grand average of 21.16%, again similar to the value from the benchmark linear Gaussian specification (21.10%).

The second alternative maintains the linear Gaussian structure of the factors but sets the instantaneous

interest rate to the square of the highest-frequency component $r_t \equiv x_{n,t}^2$, implying that zero-coupon bond prices become exponential quadratic functions of the factors as in Leippold and Wu (2002). Bond prices are given by

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

where the coefficients solve the system of ordinary differential equations:

$$\begin{aligned} B'(\tau) &= \mathcal{Z}_n - B(\tau)\kappa - \kappa^\top B(\tau) - 2B(\tau)^2\sigma_1^2, \\ b'(\tau) &= 2B(\tau)\kappa\theta^{\mathbb{Q}} - \kappa^\top b(\tau) - 2B(\tau)b(\tau)\sigma_1^2, \\ c'(\tau) &= b(\tau)^\top \kappa\theta^{\mathbb{Q}} + tr[B(\tau)]\sigma_1^2 - \frac{\sigma_1^2}{2} b(\tau)^\top b(\tau), \end{aligned}$$

starting at $B(0) = 0$, $b(0) = 0$, and $c(0) = 0$, where $\mathcal{Z}_n$ is a matrix of zeros with a one in the (n,n)th position and tr denotes the trace operator. When we estimate this model with 15 factors, we obtain a log likelihood of $29,742$, slightly higher than the linear Gaussian benchmark model. The in-sample predictive variation from the six LIBOR rates over all horizons generates a grand average of 21.05%, slightly lower than the benchmark linear Gaussian specification.

Since both of these alternative specifications generate very similar performance to the benchmark linear Gaussian model, we recommend the benchmark model for its simplicity and analytical tractability. When the possibility of negative interest rates is an important concern, either of the two alternatives can be chosen while maintaining the same dimension-invariance properties as the base model.

7 Dimension-Invariant Extensions: Stochastic Volatility and Time-varying Risk Premia

It is natural to view our efforts in this paper as the first step in a more general dimension-invariant approach. We now extend it to accommodate volatility and risk premium dynamics, allowing future applications to interest rate options and the analysis of excess bond returns.

7.1 Stochastic volatility and fixed-income option pricing

Prior literature shows that the stochastic volatility impacting interest rates is largely "unspanned" by bond prices themselves (e.g., Collin-Dufresne and Goldstein, 2003). The specification of stochastic volatility should therefore not have a large impact on the term structure, and our benchmark model therefore assumes

constant volatilities for simplicity. For other applications, such as the pricing of interest rate options, the specification of stochastic volatility is critical. For example, current practice for interest rate option pricing takes observed interest rates as given and focuses on modelling volatility dynamics.

The following candidate specification provides a natural dimension-invariant extension of our term structure model to include an analogous stochastic volatility cascade suitable for option pricing:

$$d(\sigma_{j,t}^2) = \kappa_j^v(\sigma_{j-1,t}^2 - \sigma_{j,t}^2)dt + \omega\sigma_{j,t}dZ_{j,t}, \qquad j = 1,...,n, \tag{37}$$

$$\sigma_{0,t}^2 = \theta_v, \tag{38}$$

$$\kappa_j^v = \beta^{j-1}\kappa_1^v, \qquad \beta > 1, \tag{39}$$

$$\rho = \mathbb{E}[dW_{j,t}dZ_{j,t}]/dt. \tag{40}$$

The specification thus permits an m-dimensional stochastic volatility cascade in which each component mean reverts at a geometrically increasing rate around the next-lowest frequency component. In addition to the geometric progression of mean-reversion speeds, the specification achieves dimension invariance by assuming a constant and identical coefficient ω describing the volatility of volatility, and identical correlations between the interest rate and variance innovations.

We leave empirical investigation of option pricing for future research. One possible procedure for estimation can be as follows. One first estimates our five-parameter term structure model with as many frequency components as needed to match 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 variance dynamics using interest rate options data. This estimation procedure extracts the interest-rate frequency components from the first step and the volatility factors from the second step.

7.2 Time-varying bond risk premia

Extending the concept of a dimension-invariant term structure to time-varying risk premia is straightforward. For example, Cochrane and Piazzessi (2005, 2008) suggest that risk premia affecting bonds of different maturities are all driven by a single factor, approximated by a tent-shaped function of forward rates. Within our framework this idea can be captured by allowing that $\lambda_j = l_j\lambda$, $j = 1,\cdots,n$ where l_j are scalars and λ is an $n \times 1$ column vector, where the λ_j are the risk-premium loadings on the factors defined in (11).

Alternatively, a more direct approach to investigating bond risk premia is to use our dynamic term structure model to extract the complete set of frequency components $\{x_{j,t}\}_{j=1}^n$ and use the variance dynamics

specified in (37-40) to extract the variance components $\{v_{k,t}\}_{k=1}^{m}$ from interest rate options. One can then easily investigate how bond returns over different investment horizons depend on the level of these different frequency components:

$$\text{Excess Return}_{t+\Delta t} = a + \sum_{j=1}^{n} b_j x_{j,t} + \sum_{k=1}^{m} c_k v_{k,t} + e_{t+\Delta t}.$$

Through such regression analysis, one can analyze whether the coefficients $\{b_j\}$ are tent-shaped, whether the excess returns also depend on the interest rate variances, and how the dependence structure varies across different maturities and investment horizons.

8 Conclusion

We develop a class of dynamic term structure models that are extremely parsimonious, with parameter requirements that are independent of the number of factors. Our base model uses merely five parameters to govern the time-series and cross-section of interest rates. The approach eliminates the well-known curse of dimensionality in general specifications, and allows us to estimate with equal ease and accuracy specifications with arbitrarily large numbers of factors.

We show that genuinely high-dimensional specifications improve on the traditional three-factor structure that predominates in the existing literature. A 15-factor model achieves near perfect fit to a broad cross-section of LIBOR and swap rates, is dynamically consistent, accurately matches the correlation between bonds of different maturities, and generates superior out-of-sample forecasts relative to lower-dimensional specifications.

The dimension-invariant approach we develop in this paper can naturally be extended to accommodate stochastic volatility, allowing the pricing of interest-rate options, and time-varying risk premia. We leave empirical investigations of these extensions for future research.

A Appendix

Throughout the appendix, we make explicit the dependence of the response function on the number of factors n.

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 $d(e^{\kappa_1 t} x_{1,t}) = \kappa_1 e^{\kappa_1 t} x_{1,t} dt + e^{\kappa_1 t} d(x_{1,t})$. Since $d(x_{1,t}) = \kappa_1(\theta_r - x_{1,t})dt + \sigma_1 dW_{1,t}$, we have

$$d(e^{\kappa_1 t} x_{1,t}) = e^{\kappa_1 t}(\kappa_1 \theta_r dt + \sigma_1 dW_{1,t}).$$

Integrating both sides and then dividing by $e^{\kappa_1 t}$, we obtain equation (5) for $n = 1$.

We now assume that property (5) holds for an $(n-1)$-factor structure,

$$x_{n-1,t} = \theta_r + \sum_{j=1}^{n-1}(x_{j,0} - \theta_r) a_{j,n-1}(t) + \sum_{j=1}^{n-1} \sigma_j \int_0^t a_{j,n-1}(t-s) dW_{j,s}, \tag{41}$$

where $a_{j,n-1} = (K_j * ... * K_{n-1})/\kappa_j$. Ito's lemma implies $d(e^{\kappa_n t} x_{n,t}) = \kappa_n e^{\kappa_n t} x_{n-1,t} dt + \sigma_n e^{\kappa_n t} dW_{n,t}$. Integrating both sides and then dividing both sides by $e^{\kappa_n t}$, we have

$$x_{n,t} = e^{-\kappa_n t} x_{n,0} + \int_0^t \kappa_n e^{-\kappa_n (t-s)} x_{n-1,s} ds + \sigma_n \int_0^t e^{-\kappa_n (t-s)} dW_{n,s}.$$

Substitute out $x_{n-1,s}$ according to equation (41),

$$\begin{aligned} \int_0^t \kappa_n e^{-\kappa_n (t-s)} x_{n-1,s} ds &= \theta_r (1 - e^{-\kappa_n t}) + \sum_{j=1}^{n-1} (x_{j,0} - \theta_r) \int_0^t \kappa_n e^{-\kappa_n (t-s)} a_{j,n-1}(s) ds \\ &\quad + \sum_{j=1}^{n-1} \sigma_j \int_0^t \kappa_n e^{-\kappa_n (t-s)} \left[\int_0^s a_{j,n-1}(s-u) dW_{j,u} \right] ds. \end{aligned}$$

Let $a_{n,n}(t) = e^{-\kappa_n t}$, and $a_{j,n}(t) = \int_0^t \kappa_n e^{-\kappa_n (t-s)} a_{j,n-1}(s) ds$ for all $j \leq n-1$. We observe that

$$\begin{aligned} \int_0^t \kappa_n e^{-\kappa_n (t-s)} \left[\int_0^s a_{j,n-1}(s-u) dW_{j,u} \right] ds &= \int_0^t \left[\int_u^t \kappa_n e^{-\kappa_n (t-s)} a_{j,n-1}(s-u) ds \right] dW_{j,u} \\ &= \int_0^t \left[\int_0^{t-u} \kappa_n e^{-\kappa_n (t-u-s')} a_{j,n-1}(s') ds' \right] dW_{j,u} \\ &= \int_0^t a_{j,n}(t-u) dW_{j,u} \end{aligned}$$

Thus, the proposition holds for the n-factor structure, and we conclude that it holds for all n.

The coefficients are defined recursively by $a_{j,n} = K_n * a_{j,n-1}$ starting with $a_{j,j} = K_j/\kappa_j$ for $t > 0$. Hence

$$a_{j,n} = K_n * ... * K_{j+1} * a_{j,j} = K_n * K_{n-1} * \cdots K_j/\kappa_j, \tag{42}$$

which proves equation (6) holds.

A.2 Proof of Proposition 2

Property 1. *If $\tau > 0$ is a local optimum of the response function $a_{j,n}$, then $a''_{j,n}(\tau) = \kappa_{j+1} a'_{j+1,n}(\tau)$.*

Proof. Since $a_{j,n} = (K_j * \cdots * K_n)/\kappa_j$, we infer that

$$a_{j,n}(\tau) = \frac{\kappa_{j+1}}{\kappa_j} \int_0^\tau K_j(\tau - s) a_{j+1,n}(s) ds.$$

We differentiate this relation with respect to τ

$$a'_{j,n}(\tau) = \frac{\kappa_{j+1}}{\kappa_j} \left[\int_0^\tau -\kappa_j K_j(\tau - s) a_{j+1,n}(s) ds + K_j(0) a_{j+1,n}(\tau) \right].$$

This implies the relations: $a'_{j,n}(\tau) = \kappa_{j+1} a_{j+1,n}(\tau) - \kappa_j a_{j,n}(\tau)$ and

$$a''_{j,n}(\tau) = \kappa_{j+1} a'_{j+1,n}(\tau) - \kappa_j a'_{j,n}(\tau).$$

An interior local optimum of $a_{j,n}$ therefore satisfies $a''_{j,n}(\tau) = \kappa_{j+1} a'_{j+1,n}(\tau)$ since $a'_{j,n}(\tau) = 0$. ■

We now show by backward induction that for all $j = n-1, \cdots, 1$, the function $a_{j,n}(\tau)$ is single peaked and reaches a maximum at $\overline{\tau}_{j,n}$. Furthermore, $\overline{\tau}_{1,n} \geq \cdots \geq \overline{\tau}_{n,n}$.

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

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

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

Inequality (9) can be proved by a forward recursion. Starting at $n=1$, the condition holds since $a_{1,1}(t) = e^{-\kappa_1 t} \leq 1$ for $t \geq 0$. We now assume that the inequality holds for an $(n-1)$-factor structure. We infer that

$$\begin{aligned}\sum_{j=1}^{n} a_{j,n}(t) &= e^{-\kappa_n t} + \sum_{j=1}^{n-1} \int_0^t \kappa_n e^{-\kappa_n (t-s)} a_{j,n-1}(s) ds \\ &= e^{-\kappa_n t} + \int_0^t \kappa_n e^{-\kappa_n (t-s)} \sum_{j=1}^{n-1} a_{j,n-1}(s) ds.\end{aligned}$$

Since $\sum_{j=1}^{n-1} a_{j,n-1}(s) \leq 1$ for all $s \geq 0$, 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.

A.3 Proof of Proposition 3

Derivation of the system of ordinary differential equations $(14)-(15)$**.** Ito's lemma implies that the expected return on the bond under $\mathbb{Q}$ is:

$$\mathbb{E}^{\mathbb{Q}}\left(\frac{1}{P_t}\frac{dP_t}{dt}\right) = -b_n(\tau)^\top \kappa^* (\theta^{\mathbb{Q}} - X_t) + \frac{1}{2} b_n(\tau)^\top \Sigma b_n(\tau) + c_n'(\tau) + b_n'(\tau)^\top X_t.$$

Since this expectation is equal to the interest rate $r_t = x_{n,t}$, we infer that

$$-b_n(\tau)^\top \kappa^* \theta^{\mathbb{Q}} + \frac{1}{2} b_n(\tau)^\top \Sigma b_n(\tau) + c_n'(\tau) = 0, \tag{43}$$

$$b_n(\tau)^\top \kappa^* X_t + b_n'(\tau)^\top X_t = \mathbf{e}_n^\top X_t. \tag{44}$$

Equation (43) is equivalent to (15). Equation (44) implies that $b_n(\tau)^\top \kappa^* + b_n'(\tau)^\top = \mathbf{e}_n^\top$, and we conclude that (14) holds.

Constant risk premia. When risk premia are constant: $\gamma_{j,t} = \gamma_j$ for all j, we can solve the system of ordinary differential equations $(14)-(15)$ in closed-form. By (14), the function $b_n(\tau)$ satisfies

$$\begin{aligned} b_{1,n}'(\tau) &= -\kappa_1 b_{1,n}(\tau) + \kappa_2 b_{2,n}(\tau), \\ &\vdots \\ b_{n-1,n}'(\tau) &= -\kappa_{n-1} b_{n-1,n}(\tau) + \kappa_n b_{n,n}(\tau), \\ b_{n,n}'(\tau) &= 1 - \kappa_n b_{n,n}(\tau). \end{aligned}$$

The last equation implies that

$$b_{n,n}(\tau) = \int_0^\tau a_{n,n}(s)ds = \frac{1-e^{-\kappa_n \tau}}{\kappa_n}.$$

The penultimate equation implies that

$$\frac{d}{d\tau}\left[e^{\kappa_{n-1}\tau} b_{n-1,n}(\tau)\right] = e^{\kappa_{n-1}\tau}\kappa_n b_{n,n}(\tau).$$

Hence $b_{n-1,n}(\tau) = (\kappa_n/\kappa_{n-1})K_{n-1}(\tau) * b_{n,n}(\tau)$, and therefore

$$b'_{n-1,n}(\tau) = \frac{\kappa_n}{\kappa_{n-1}} K_{n-1}(\tau) * b'_{n,n}(\tau) = \frac{\kappa_n}{\kappa_{n-1}} K_{n-1}(\tau) * a_{n,n}(\tau) = a_{n-1,n}(\tau).$$

We infer that

$$b_{n-1,n}(\tau) = \int_0^\tau a_{n-1,n}(s)ds.$$

More generally, we infer that $b_{j,n} = (\kappa_n/\kappa_j)K_j * ... * K_{n-1} * b_{n,n}$ and therefore (16) holds for all $j \in \{1,...,n-1\}$.

By (15), the function $c_n(\tau)$ satisfies

$$c'_n(\tau) = \kappa_1\theta_r b_{1,n}(\tau) - \sum_{j=1}^n \gamma_j\sigma_j^2 b_{j,n}(\tau)\,dsb - \frac{1}{2}\sum_{j=1}^n \sigma_j^2 b_{j,n}^2(\tau).$$

Hence

$$c_n(\tau) = \kappa_1\theta_r \int_0^\tau b_{1,n}(\tau)\,d\tau - \sum_{j=1}^n \gamma_j\sigma_j^2 \int_0^\tau b_{j,n}(s)\,ds - \frac{1}{2}\sum_{j=1}^n \sigma_j^2 \int_0^\tau b_{j,n}^2(s)\,ds.$$

We infer from (16) that

$$\int_0^\tau b_{j,n}(s)\,ds = \sum_{i=j}^n \alpha_{i,j,n}\left(\tau - \frac{1-e^{-\kappa_i\tau}}{\kappa_i}\right).$$

Similarly, (16) implies

$$\begin{aligned}\int_0^\tau b_{j,n}^2(s)ds &= \sum_{i=j}^n\sum_{k=j}^n \alpha_{i,j,n}\alpha_{k,j,n}\int_0^\tau (1-e^{-\kappa_i s})(1-e^{-\kappa_k s})ds \\ &= \sum_{i=j}^n\sum_{k=j}^n \alpha_{i,j,n}\alpha_{k,j,n}\left[\tau - \frac{1-e^{-\kappa_i\tau}}{\kappa_i} - \frac{1-e^{-\kappa_k\tau}}{\kappa_k} + \frac{1-e^{-(\kappa_i+\kappa_k)\tau}}{\kappa_i+\kappa_k}\right].\end{aligned}$$

and we conclude that equation (17) holds.

We easily verify that

$$\kappa^{-1} = \begin{pmatrix} \kappa_1^{-1} & 0 & \cdots & 0 \\ \kappa_1^{-1} & \kappa_2^{-1} & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ \kappa_1^{-1} & \kappa_2^{-1} & \cdots & \kappa_n^{-1} \end{pmatrix},$$

and infer that $\theta^{\mathbb{Q}} = (\theta_r - \gamma_1\sigma_1^2/\kappa_1, \theta_r - \gamma_1\sigma_1^2 - \gamma_2\sigma_2^2/\kappa_2, ..., \theta_r - \sum_{i=1}^n \gamma_i\sigma_i^2/\kappa_i)^\top$.

Derivation of equation (18). Since $a_{j,n} = K_n * K_{n-1} * \cdots K_j/\kappa_j$, we infer that

$$\int_0^{+\infty} a_{j,n}(\tau)d\tau = \frac{1}{\kappa_j}.$$

By (16),

$$\sum_{i=j}^n \alpha_{i,j,n} = \frac{1}{\kappa_j}.$$

Hence

$$\sum_{j=1}^n \gamma_j\sigma_j^2 \sum_{i=j}^n \alpha_{i,j,n}\left(\tau - \frac{1-e^{-\kappa_i\tau}}{\kappa_i}\right) = \left(\sum_{j=1}^n \frac{\gamma_j\sigma_j^2}{\kappa_j}\right)\tau - \sum_{i=1}^n \left(\sum_{j=1}^i \alpha_{i,j,n}\gamma_j\sigma_j^2\right)\frac{1-e^{-\kappa_i\tau}}{\kappa_i}.$$

Similarly,

$$\begin{aligned} & \sum_{i=j}^n \sum_{k=j}^n \alpha_{i,j,n}\alpha_{k,j,n}\left(\tau - \frac{1-e^{-\kappa_i\tau}}{\kappa_i} - \frac{1-e^{-\kappa_k\tau}}{\kappa_k} + \frac{1-e^{-(\kappa_i+\kappa_k)\tau}}{\kappa_i+\kappa_k}\right) \\ = & \frac{\tau}{\kappa_j^2} - \frac{2}{\kappa_j}\sum_{i=j}^n \alpha_{i,j,n}\frac{1-e^{-\kappa_i\tau}}{\kappa_i} + \sum_{i=j}^n \sum_{k=j}^n \alpha_{i,j,n}\alpha_{k,j,n}\frac{1-e^{-(\kappa_i+\kappa_k)\tau}}{\kappa_i+\kappa_k}. \end{aligned}$$

Equation (18) therefore holds.

Generalization. More generally, we note that we can solve the equation:

$$b_n'(\tau) = \mathbf{e}_n - \kappa^{*\top} b_n(\tau) +^\top X_t$$

as long as the matrix $\kappa^{*\top}$ is upper triangular. Since $\kappa^* = \kappa + \Sigma\Lambda$, we conclude that we can solve for $b_n(\tau)$ explicitly when Λ is lower triangular, that is if the risk premium of any factor j is only affected by factors $i \leq j$.

A.4 Proof that the response functions are translated versions of each other

Let $\{E_j\}_{j=1}^\infty$ denote a sequence of independent, exponentially distributed random variables. The p.d.f. of E_j is $\kappa_j e^{-\kappa_j x}$, where κ_j satisfies (21). The Fourier transform of $E_j + ... + E_n$ is

$$\mathbb{E}e^{-2i\pi\xi(E_j+...+E_n)} = \prod_{\ell=j}^n \frac{\kappa_\ell}{\kappa_\ell + 2i\pi\xi}.$$

For a fixed j, we infer that

$$\psi_{j,n}(\xi) = \ln\left[\mathbb{E}e^{-2i\pi\xi(E_j+...+E_n)}\right] = -\sum_{\ell=j}^n \ln\left(1 + \frac{2i\pi\xi}{\kappa_\ell}\right) \tag{45}$$

has a limit when $n \to \infty$ since $\sum_{\ell=j}^{\infty}(1/\kappa_\ell) < \infty$. The random variable $E_j + ... + E_n$ and its density $K_j * \cdots * K_n$ therefore have well-defined limits as $n \to \infty$.

We note that

$$\psi_{j,n}(\xi) = -\sum_{\ell=j}^{n} \ln\left(1 + \frac{2i\pi b\xi}{\kappa_1 b^j}\right) = \psi_{j+1,n+1}(b\xi).$$

The inverse Fourier transform of $\psi_{j,n}(\xi)$ is $\kappa_j a_{j,n}(\tau)$, while the inverse Fourier transform of $\psi_{j+1,n+1}(b\xi)$ is $\kappa_{j+1}a_{j+1,n+1}(\tau/b)/b = \kappa_j a_{j+1,n+1}(\tau/b)$. Hence

$$a_{j,n}(\tau) \equiv a_{j+1,n+1}(\tau/b).$$

Let $A_j = \lim_{n\to\infty} a_{j,n}(\tau)$. We know that $A_j(\tau) = A_{j+1}(\tau/b)$, that is the $A_j(\tau)$ are translated versions of each other in log scale. We infer that $a_{j,n}(\tau) \approx a_{j+1,n}(\tau/b)$ for large n.

A.5 Proof of Proposition 4

We denote by $\|y\|_2^2 = \mathbb{E}\left(\int_0^T y_t^2 dt\right)$ the norm of an adapted square integrable process $y \in L^2$. The convergence proof is based on the relation:

$$x_{n,t} = \theta_r + \sum_{j=1}^{n} \sigma_j \int_{-\infty}^{t} a_{j,n}(t-s) dW_{j,s}.$$

By Proposition 2, the factor loadings satisfy $a_{j,n} = \sum_{i=j}^{n} \alpha_{i,j,n} K_i$ for all $j \geq 1$, where

$$\alpha_{i,j,n} = \frac{\kappa_j ... \kappa_n}{\kappa_i \kappa_j \prod_{k=j,k\neq i}^{n}(\kappa_k - \kappa_i)}.$$

When $\kappa_i = \kappa_1 b^{i-1}$, the coefficients can be rewritten as

$$\alpha_{i,j,n} = \frac{1}{\kappa_j} \frac{(-1)^{i-j} b^{-(i-j)(i-j+1)/2}}{F(n-i)F(i-j)},$$

where $F(0) = 1$ and $F(k) = (1-b^{-1})...(1-b^{-k})$ for all $k \geq 1$. The sequence $\{F(k)\}$ is decreasing and converges to a strictly positive limit F_∞. We observe that

$$F(k) - F_\infty = F(k)\left[1 - \prod_{i=k+1}^{+\infty}\left(1-b^{-i}\right)\right] \leq F(k) \sum_{i=k+1}^{+\infty} b^{-i}$$

and therefore

$$\frac{F(k) - F_\infty}{F(k)} \leq \frac{b^{-k}}{b-1}$$

for all k.

As n goes to infinity, the coefficient $\alpha_{i,j,n}$ converges to

$$\bar{\alpha}_{i,j} = \frac{1}{\kappa_j F_\infty} \frac{(-1)^{i-j} b^{-(i-j)(i-j+1)/2}}{F(i-j)}.$$

We show:

Property 2. *There exists a finite constant C such that $\sum_{i=j}^{+\infty} |\bar{\alpha}_{i,j}| \, \|K_i\|_2 \leq C/b^{j/2}$ for all j.*

Proof. Since $||K_i||_2 \leq \sqrt{\kappa_i}$, we infer that

$$\sum_{i=j}^{+\infty} |\bar{\alpha}_{i,j}| \, \|K_i\|_2 \leq \frac{1}{F_\infty^2 \sqrt{\kappa_j}} \sum_{i=j}^{+\infty} b^{-(i-j)(i-j+1)/2} \sqrt{\frac{\kappa_i}{\kappa_j}} = \frac{1}{F_\infty^2 \sqrt{\kappa_j}} \sum_{i=j}^{+\infty} b^{-(i-j)(i-j-1)/2}.$$

Letting $C = F_\infty^{-2} \sqrt{b/\kappa_1} \sum_{i=0}^{+\infty} b^{-i(i-1)/2}$, we conclude that $\sum_{i=j}^{+\infty} |\bar{\alpha}_{i,j}| \, \|K_i\|_2 \leq C/b^{j/2}$. ■

The function $\bar{a}_j = \sum_{i=j}^{+\infty} \bar{\alpha}_{i,j} K_i$ is therefore well-defined and square-integrable, and its norm satisfies $\|\bar{a}_j\|_2 \leq C/b^{j/2}$. We also show:

Property 3. *There exists a finite constant C' such that $||a_{j,n} - \bar{a}_j||_2 \leq C' b^{(j/2-n)}$ for all $j \leq n$.*

Proof of Property 3. The identity $a_{j,n} - \bar{a}_j = \sum_{i=j}^{n} (\alpha_{i,j,n} - \bar{\alpha}_{i,j}) K_i - \sum_{i=n+1}^{+\infty} \bar{\alpha}_{i,j} K_i$ implies that

$$||a_{j,n} - \bar{a}_j||_2 \leq \sum_{i=j}^{n} |\alpha_{i,j,n} - \bar{\alpha}_{i,j}| \sqrt{\kappa_i} + \sum_{i=n+1}^{+\infty} |\bar{\alpha}_{i,j}| \sqrt{\kappa_i}$$

We observe that

$$\begin{aligned} |\alpha_{i,j,n} - \bar{\alpha}_{i,j}| \sqrt{\kappa_i} &\leq \frac{1}{\sqrt{\kappa_j} F_\infty^2} b^{-(i-j)(i-j-1)/2} \frac{F(n-i) - F_\infty}{F(n-i)} \\ &\leq \frac{1}{\sqrt{\kappa_j} F_\infty^2} b^{-(i-j)(i-j-1)/2} \frac{b^{-(n-i)}}{b-1}. \end{aligned}$$

and therefore

$$\sum_{i=j}^{n} |\alpha_{i,j,n} - \bar{\alpha}_{i,j}| \sqrt{\kappa_i} \leq \frac{\sqrt{b}}{(b-1)\sqrt{\kappa_1} F_\infty^2} \sum_{i=j}^{n} b^{-(i-j)(i-j-3)/2} b^{(j/2-n)} = C_1' b^{(j/2-n)}.$$

where $C_1' = \sqrt{b} \left[\sum_{i=0}^{\infty} b^{-i(i-3)/2} \right] / \left[(b-1)\sqrt{\kappa_1} F_\infty^2 \right]$.

Similarly,

$$\sum_{i=n+1}^{+\infty} |\bar{\alpha}_{i,j}| \sqrt{\kappa_i} \leq \frac{1}{F_\infty^2 \sqrt{\kappa_j}} \sum_{i=n+1}^{+\infty} b^{-(i-j)(i-j-1)/2} \leq \frac{1}{F_\infty^2 \sqrt{\kappa_j}} \sum_{i=n+1}^{+\infty} b^{-(i-j-1)}$$

Recall that $\sum_{i=n+1}^{+\infty} b^{-(i-j-1)} = b^{-(n-j)}/(1-b^{-1})$, and define $C_2' = \sqrt{b}/[(b-1)\sqrt{\kappa_1}F_\infty^2]$. We infer that

$$\sum_{i=n+1}^{+\infty} |\bar{\alpha}_{i,j}| \sqrt{\kappa_i} \leq C_2' b^{(j/2-n)}.$$

Letting $C' = C_1' + C_2'$, we conclude that $||a_{j,n} - \bar{a}_j||_2 \leq C' b^{(j/2-n)}$. ■

The process

$$x_{\infty,t} = \theta_r + \sum_{j=1}^{+\infty} \sigma_j \int_{-\infty}^{t} \bar{a}_j(t-s) dW_{j,s}.$$

is well-defined when $\sum_{j=1}^{+\infty} \sigma_j^2/\kappa_j < \infty$. Since

$$x_{n,t} - x_{\infty,t} = \sum_{j=1}^{n} \sigma_j \int_{-\infty}^{t} (a_{j,n} - \bar{a}_j)(t-s) dW_{j,s} - \sum_{j=n+1}^{+\infty} \int_{-\infty}^{t} \bar{a}_j(t-s) \sigma_j dW_{j,s},$$

we know that

$$\begin{aligned} ||x_n - x_\infty||_2^2 &= \sum_{j=1}^{n} \sigma_j^2 ||a_{j,n} - \bar{a}_j||_2^2 + \sum_{j=n+1}^{+\infty} \sigma_j^2 ||\bar{a}_j||_2^2 \\ &\leq C'^2 \sum_{j=1}^{n} \sigma_j^2 b^{(j-2n)} + C^2 \sum_{j=n+1}^{+\infty} \frac{\sigma_j^2}{b^j} \end{aligned}$$

Let $[y]$ denote the integral part of a real number y. If $\sum_{j=1}^{\infty} \sigma_j^2/\kappa_j < \infty$, we conclude that

$$\begin{aligned} ||x_n - x_\infty||_2^2 &\leq C'^2 \sum_{j=1}^{[n/2]} \frac{\sigma_j^2}{b^j} b^{2(j-n)} + C'^2 \sum_{j=[n/2]+1}^{n} \frac{\sigma_j^2}{b^j} b^{2(j-n)} + C^2 \sum_{j=n+1}^{+\infty} \frac{\sigma_j^2}{b^j} \\ &\leq C'^2 \left(\sum_{j=1}^{\infty} \frac{\sigma_j^2}{b^j} \right) b^{-n} + (C^2 + C'^2) \left(\sum_{j=[n/2]+1}^{+\infty} \frac{\sigma_j^2}{b^j} \right) \end{aligned}$$

converges to zero when $n \to \infty$. This establishes the convergence of the short rate process.

We next turn to the convergence of the zero-coupon bond yield:

$$y_n(t,\tau) = \frac{\sum_{j=1}^{n} b_{j,n}(\tau) x_{j,t} + c_n(\tau)}{\tau}$$

as $n \to \infty$ for a fixed value of $\tau > 0$. We observe that

$$\left\| \sum_{j=1}^{n} b_{j,n}(\tau) x_{j,t} - \sum_{j=1}^{\infty} \bar{b}_j(\tau) x_{j,t} \right\|_2 \leq \left(\sum_{j=1}^{n} \left| b_{j,n}(\tau) - \bar{b}_j(\tau) \right| + \sum_{j=n+1}^{\infty} \left| \bar{b}_j(\tau) \right| \right) \sup_j \|x_{j,t}\|_2 .$$

Since $\left|\bar{b}_j(\tau)\right| \leq 1/\kappa_j$ and

$$\left| b_{j,n}(\tau) - \bar{b}_j(\tau) \right| \leq \sum_{i=j}^{n} |\alpha_{i,j,n} - \bar{\alpha}_{i,j}| \leq \frac{C''}{\kappa_n},$$

where $C'' = \sum_{i=0}^{\infty} b^{-i(i-2)}/[(b-1)F_\infty^2]$, we conclude that $\sum_{j=1}^{n} b_{j,n}(\tau)x_{j,t}/\tau$ converges to $\sum_{j=1}^{\infty} \bar{b}_j(\tau)x_{j,t}$ in L^2. We also infer from (19) that $c_n(\tau)$ converges to

$$\bar{c}(\tau) = \theta_r - \sum_{j=1}^{\infty} \frac{\sigma_j^2}{\kappa_j^2}\left(\gamma\kappa_j + \frac{1}{2}\right) + \sum_{i=1}^{\infty} \bar{c}_i \frac{1-e^{-\kappa_i\tau}}{\kappa_i\tau} - \sum_{j=1}^{\infty} \frac{\sigma_j^2}{2} \sum_{i=j}^{\infty}\sum_{k=j}^{\infty} \bar{\alpha}_{i,j}\bar{\alpha}_{k,j} \frac{1-e^{-(\kappa_i+\kappa_k)\tau}}{(\kappa_i+\kappa_k)\tau},$$

where $\bar{c}_i = -\theta_r\kappa_1\bar{\alpha}_{i,1} + \sum_{j=1}^{i} \bar{\alpha}_{i,j}\sigma_j^2(\gamma + 1/\kappa_j)$, and conclude that the yield $y_n(t,\tau)$ converges to $\left[\sum_{j=1}^{\infty} \bar{b}_j(\tau)x_{j,t} + \bar{c}(\tau)\right]/\tau$.

A.6 Unscented Kalman filter, maximum likelihood estimation, and Vuong test

Since the measurement function $h(X_t)$ is nonlinear, one possibility is to rely on a Taylor expansion to obtain extended forms of the Kalman Filter (e.g. Baadsgaard, Madsen, and Nielsen (2001), Chen and Scott (2003), Duan and Simonato (1999), and Duffee and Stanton (2008)). Alternatively, Julier and Uhlmann (1997) propose the unscented Kalman filter (UKF) to directly approximate the posterior density using a set of deterministically chosen sample points (sigma points). The UKF is accurate 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,

$$\overline{X}_t = A + \Phi\widehat{X}_{t-1}, \quad \overline{V}_{x,t} = \Phi\widehat{V}_{x,t-1}\Phi^\top + \Sigma_x, \tag{46}$$

where $\overline{X}_t$ and $\overline{V}_{x,t}$ 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 χ_i on the state,

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

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

$$\begin{aligned} \overline{y}_t &= \textstyle\sum_{i=0}^{2k} w_i\zeta_{t,i}, \\ \overline{V}_{y,t} &= \textstyle\sum_{i=0}^{2k} w_i\left[\zeta_{t,i} - \overline{y}_t\right]\left[\zeta_{t,i} - \overline{y}_t\right]^\top + \Sigma_y, \\ \overline{V}_{xy,t} &= \textstyle\sum_{i=0}^{2k} w_i\left[\chi_{t,i} - \overline{X}_t\right]\left[\zeta_{t,i} - \overline{y}_t\right]^\top. \end{aligned} \tag{47}$$

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

$$\widehat{X}_t = \overline{X}_t + K_t(y_t - \overline{y}_t), \quad \widehat{V}_{y,t} = \overline{V}_{y,t} - K_t\overline{V}_{y,t}K_t^\top, \tag{48}$$

where $K_t = \overline{V}_{xy,t} \left(\overline{V}_{y,t}\right)^{-1}$ 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:

$$l_t(\Theta) = -\frac{1}{2}\ln\left|\overline{V}_{y,t}\right| - \frac{1}{2}\left(y_t - \overline{y}_t\right)^{\top}\left(\overline{V}_{y,t}\right)^{-1}\left(y_t - \overline{y}_t\right). \tag{49}$$

We choose model parameters $\Theta \equiv (\kappa_1, b, \sigma_1, \theta_r, , \theta_r^{\mathbb{Q}}, \sigma_e^2)^{\intercal}$ that maximize $\mathcal{L}(\Theta) \equiv \sum_{t=1}^{N} l_t(\Theta)$. We constrain $(\theta_r, \theta_r^{\mathbb{Q}})$ to be positive in the estimation.

The Vuong statistic is given by

$$\mathcal{V}_n = \sqrt{N}(m_n)/s_n, \quad \rho_{n,t} = l_{15,t} - l_{n,t}, \quad n = 1, 2, \cdots, 14, \tag{50}$$

where $m_n = \Sigma_{t=1}^{N}\rho_{n,t}/N$ denotes the sample mean of the weekly log likelihood difference $\rho_{n,t}$ between the 15-factor model and the model with n factors. The sample standard deviation of $\rho_{n,t}$ is denoted by s_n.

References

Akkouchi, M., 2008, "On the Convolution of Exponential Distributions," *Journal of the Chungcheong Mathematical Society*, 21(4), 501–510.

Ang, A., S. Dong, and M. Piazzesi, 2007, "No-Arbitrage Taylor Rules," working paper, Columbia University and University of Chicago.

Ang, A., and M. Piazzesi, 2003, "A No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables," *Journal of Monetary Economics*, 50(4), 745–787.

Ang, A., M. Piazzesi, and M. Wei, 2004, "What Does the Yield Curve Tell Us About GDP Growth?," *Journal of Econometrics*, 131(1-2), 359–403.

Baadsgaard, M., H. Madsen, and J. N. Nielsen, 2001, "Estimating and Testing an Exponential-Affine Term Structure Model by Nonlinear Filtering," working paper, Technical University of Denmark, Denmark.

Backus, D., S. Foresi, A. Mozumdar, and L. Wu, 2001, "Predictable Changes in Yields and Forward Rates," *Journal of Financial Economics*, 59(3), 281–311.

Balduzzi, P., G. Bertola, and S. Foresi, 1997, "A Model of Target Changes and the Term Structure of Interest Rates," *Journal of Monetary Economics*, 39(2), 223–249.

Balduzzi, P., S. Das, and S. Foresi, 1998, "The Central Tendency: A Second Factor in Bond Yields," *Review of Economics and Statistics*, 80(1), 62–72.

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

Bali, T., M. Heidari, and L. Wu, 2009, "Predictability of Interest Rates and Interest-Rate Portfolios," *Journal of Business and Economic Statistics*, 27(4), 517–527.

Bekaert, G., S. Cho, and A. Moreno, 2005, "New-Keynesian Macroeconomics and the Term Structure," working paper, Columbia University.

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

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

———, 2004, "How to Forecast Long-run Volatility: Regime-switching and the Estimation of Multifractal Processes," *Journal of Financial Econometrics*, 2, 49–83.

———, 2007, "Multifrequency News and Stock Returns," *Journal of Financial Economics*, 86, 178–212.

———, 2008, *Multifractal Volatility: Theory, Forecasting, and Pricing*. Academic Press, Burlington, MA.

Chambers, D. R., W. T. Carleton, and D. W. Waldman, 1984, "A New Approach to Estimation of the Term Structure of Interest Rates," *Journal of Financial and Quantitative Analysis*, 19(3), 233–252.

Chen, R.-R., and L. Scott, 2003, "Multi-Factor Cox-Ingersoll-Ross Models of the Term Structure: Estimates and Tests from a Kalman Filter Model," *Journal of Real Estate Finance and Economics*, 27(2), 143–172.

Christensen, J. H. E., F. X. Diebold, and G. D. Rudebusch, 2008, "An Arbitrage-Free Generalized Nelson-Siegel Term Structure Model," working paper, Federal Reserve Bank of San Francisco and University of Pennsylvania.

Cochrane, J. H., and M. Piazzesi, 2008, "Decomposing the Yield Curve," Working paper, University of Chicago and Stanford University.

Collin-Dufresne, P., and R. S. Goldstein, 2003, "Generalizating the Affine Framework to HJM and Random Field Models," working paper, Washington University at St. Louis.

Constantinides, G. M., 1992, "A Theory of the Nominal Term Structure of Interest Rates," *Review of Financial Studies*, 5(4), 531–552.

Cox, J. C., J. E. Ingersoll, and S. R. Ross, 1985, "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53(2), 385–408.

Dai, Q., and K. Singleton, 2000, "Specification Analysis of Affine Term Structure Models," *Journal of Finance*, 55(5), 1943–1978.

———, 2002, "Expectation Puzzles, Time-Varying Risk Premia, and Affine Models of the Term Structure," *Journal of Financial Economics*, 63(3), 415–441.

———, 2003, "Term Structure Dynamics in Theory and Reality," *Review of Financial Studies*, 16(3), 631–678.

Diebold, F. X., and C. Li, 2006, "Forecasting the Term Structure of Government Bond Yields," *Journal of Econometrics*, 130(2), 337–364.

Diebold, F. X., G. D. Rudebusch, and S. B. Aruba, 2006, "The Macroeconomy and the Yield Curve: A Dynamic Latent Factor Approach," *Journal of Econometrics*, 131(1-2), 309–338.

Duan, J.-C., and J.-G. Simonato, 1999, "Estimating and Testing Exponential-Affine Term Structure Models by Kalman Filter," *Review of Quantitative Finance and Accoutning*, 13(2), 111–135.

Duffee, G. R., 2002, "Term Premia and Interest Rate Forecasts in Affine Models," *Journal of Finance*, 57(1), 405–443.

———, 2009, "Forecasting with the Term Structure: The Role of No-Arbitrage," Working paper, Johns Hopkins University.

Duffee, G. R., and R. Stanton, 2007, "Evidence on Simulation Inference for Near Unit-Root Processes with Implications for Term Structure Estimation," *Journal of Financial Econometrics*, 6(1), 108–142.

Duffee, G. R., and R. H. Stanton, 2008, "Evidence on Simulation Inference for Near Unit-Root Processes with Implications for Term Structure Estimation," *Journal of Financial Econometrics*, 6(1), 108–142.

Duffie, D., D. Filipović, and W. Schachermayer, 2003, "Affine Processes and Applications in Finance," *Annals of Applied Probability*, 13, 984–1053.

Duffie, D., and R. Kan, 1996, "A Yield-Factor Model of Interest Rates," *Mathematical Finance*, 6(4), 379–406.

Duffie, D., J. Pan, and K. Singleton, 2000, "Transform Analysis and Asset Pricing for Affine Jump Diffusions," *Econometrica*, 68(6), 1343–1376.

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

Filipović, D., 1999, "A Note on the Nelson-Siegel Family," *Mathematical Finance*, 9(4), 349–359.

Gallmeyer, M., B. Hollifield, F. Palomino, and S. Zin, 2005, "Arbitrage-Free Bond Pricing with Dynamic Macroeconomic Models," *Journal of Monetary Economics*, 52(5), 921–950.

Gallmeyer, M., B. Hollifield, and S. Zin, 2005, "Taylor Rules, McCallum Rules, and the Term Structure of Interest Rates," *Journal of Monetary Economics*, 52(5), 921–950.

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

Gurkaynak, R. S., B. Sack, and J. H. Wright, 2006, "The U.S. Treasury Yield Curve: 1961 to the Present," Finance and Economics Discussion Series 2006-28, Federal Reserve Board.

Han, B., 2007, "Stochastic Volatilities and Correlations of Bond Yields," *Journal of Finance*, 62, 1491–1524.

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

Heidari, M., and L. Wu, 2009, "Market Anticipation of Fed Policy Changes and the Term Structure of Interest Rates," *Review of Finance*, forthcoming.

Ho, T. S. Y., and S. B. Lee, 1986, "Term Structure Movements and Pricing Interest Rate Contingent Claims," *Journal of Finance*, 41(5), 1011–1030.

Hördahl, P., O. Tristanoi, and D. Vestin, 2006, "A Joint Econometric Model of Macroeconomic and Term Structure Dynamics," *Journal of Econometrics*, 131(1-2), 405–444.

Hull, J., and A. White, 1993, "Bond Option Pricing Based on a Model For the Evolution of Bond Prices," *Advances in Futures and Options Research*, 6, 1–13.

Joslin, S., K. J. Singleton, and H. Zhu, 2010, "A New Perspective on Gaussian Dynamic Term Structure Models," *Review of Financial Studies*, forthcoming.

Julier, S. J., and J. K. Uhlmann, 1997, "A New Extension of the Kalman filter to Nonlinear Systems," in *Proceedings of AeroSense: The 11th International Symposium on Aerospace/Defense Sensing, Simulation and Controls*, Orlando, Florida.

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

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

———, 1997, "Characterizing Gaussian Models of the Term Structure of Interest Rates," *Mathematial Finance*, 7, 107–118.

Leippold, M., and L. Wu, 2002, "Asset Pricing under the Quadratic Class," *Journal of Financial and Quantitative Analysis*, 37(2), 271–295.

Li, H., and F. Zhao, 2006, "Unspanned Stochastic Volatility: Evidence from Hedging Interest Rate Derivatives," *Journal of Finance*, 61(1), 341–378.

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

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

Longstaff, F. A., and E. S. Schwartz, 1992, "Interest-Rate Volatility and the Term Structure: A Two Factor General Equilibrium Model," *Journal of Finance*, 47(4), 1259–1282.

Lu, B., and L. Wu, 2009, "Macroeconomic Releases and the Interest Rate Term Structure," *Journal of Monetary Economics*, 56(6), 872–884.

McCulloch, J. H., 1975, "The Tax-Adjusted Yield Curve," *Journal of Finance*, 30(3), 811–830.

Nelson, C. R., and A. F. Siegel, 1987, "Parsimonious Modeling of Yield Curves," *Journal of Business*, 60(4), 473–489.

Piazzesi, M., 2005, "Bond Yields and the Federal Reserve," *Journal of Political Economy*, 113(2), 311–344.

——— , 2010, "Affine Term Structure Models," in *Handbook of Financial Econometrics*, ed. by Y. Ait-Sahalia, and L. P. Hansen. North Holland Elsevier, Amsterdam, vol. 1, pp. 691–766.

Ronn, E. I., 1987, "A New Linear Programming Approach to Bond Portfolio Management," *Journal of Financial and Quantitative Analysis*, 22(4), 439–466.

Rudebusch, G. D., 2002, "Term Structure Evidence on Interest Rate Smoothing and Monetary Policy Inertia," *Journal of Monetary Economics*, 49(6), 1161–1187.

Rudebusch, G. D., E. T. Swanson, and T. Wu, 2006, "The Bond Yield "Conundrum" from a Macro-Finance Perspective," *Monetary and Economic Studies*, forthcoming.

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

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

Vasicek, O. A., 1977, “An Equilibrium Characterization of the Term Structure,” *Journal of Financial Economics*, 5(2), 177–188.

Vuong, Q., 1989, “Likelihood Ratio Tests for Model Selection and Non-nested Hypotheses,” *Econometrica*, 57, 307–333.

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.

Maturity	Mean	Std	Skew	Kurt	Autocorrelation			
					1	5	10	20
1 m	4.335	1.798	-0.714	-1.050	0.998	0.988	0.971	0.922
2 m	4.370	1.803	-0.722	-1.038	0.998	0.989	0.972	0.923
3 m	4.405	1.808	-0.720	-1.025	0.998	0.989	0.971	0.921
6 m	4.475	1.803	-0.713	-0.970	0.998	0.986	0.967	0.916
9 m	4.547	1.789	-0.689	-0.910	0.997	0.983	0.962	0.908
12 m	4.631	1.769	-0.653	-0.854	0.996	0.979	0.954	0.898
2 y	4.877	1.570	-0.529	-0.699	0.994	0.966	0.932	0.865
3 y	5.093	1.414	-0.407	-0.663	0.992	0.957	0.916	0.842
4 y	5.260	1.298	-0.291	-0.693	0.991	0.950	0.904	0.823
5 y	5.395	1.209	-0.187	-0.748	0.990	0.944	0.893	0.808
7 y	5.595	1.091	-0.023	-0.850	0.988	0.937	0.880	0.787
10 y	5.798	0.994	0.126	-0.949	0.987	0.931	0.870	0.769
15 y	6.009	0.909	0.228	-1.020	0.986	0.928	0.865	0.762
20 y	6.103	0.870	0.254	-1.020	0.985	0.927	0.864	0.761
30 y	6.136	0.851	0.295	-0.949	0.986	0.926	0.862	0.756
Average	5.135	1.398	-0.316	-0.896	0.992	0.959	0.919	0.844

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 $n = 1, 2, \cdots, 15$. The column under $\mathcal{L}$ 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.

n	κ_1	θ_r	σ_1	$\theta_r^{\mathbb{Q}}$	b	σ_e^2	$\mathcal{L}$	$\mathcal{V}$
1	0.2092 (0.0009)	0.0436 (0.0003)	0.0065 (0.0001)	0.0688 (0.0001)	0.0000 (0.0000)	0.1574 (0.0009)	4086	47.91
2	0.0603 (0.0007)	0.0131 (0.0004)	0.0111 (0.0001)	0.0695 (0.0004)	6.1387 (0.0905)	0.0187 (0.0001)	13967	30.79
3	0.0526 (0.0003)	0.0000 (0.0000)	0.0101 (0.0000)	0.0662 (0.0001)	7.3138 (0.0393)	0.0047 (0.0000)	19928	20.70
4	0.0366 (0.0004)	0.0000 (0.0000)	0.0116 (0.0000)	0.0653 (0.0002)	4.2707 (0.0216)	0.0019 (0.0000)	23276	17.33
5	0.0441 (0.0005)	0.0000 (0.0000)	0.0125 (0.0000)	0.0507 (0.0003)	2.8266 (0.0108)	0.0010 (0.0000)	25551	15.99
6	0.0383 (0.0003)	0.0000 (0.0000)	0.0123 (0.0000)	0.0497 (0.0002)	3.0267 (0.0070)	0.0005 (0.0000)	27527	10.60
7	0.0283 (0.0003)	0.0000 (0.0000)	0.0129 (0.0000)	0.0419 (0.0003)	2.6150 (0.0047)	0.0004 (0.0000)	27898	11.93
8	0.0275 (0.0003)	0.0000 (0.0000)	0.0133 (0.0000)	0.0632 (0.0006)	2.5271 (0.0057)	0.0004 (0.0000)	28445	11.00
9	0.0278 (0.0003)	0.0000 (0.0000)	0.0141 (0.0000)	0.0650 (0.0006)	2.2351 (0.0046)	0.0003 (0.0000)	28801	9.18
10	0.0313 (0.0004)	0.0000 (0.0000)	0.0140 (0.0000)	0.0507 (0.0005)	2.2010 (0.0048)	0.0003 (0.0000)	28972	6.68
11	0.0305 (0.0002)	0.0000 (0.0000)	0.0144 (0.0000)	0.0966 (0.0006)	1.9603 (0.0046)	0.0003 (0.0000)	29036	6.06
12	0.0359 (0.0003)	0.0000 (0.0000)	0.0147 (0.0000)	0.0876 (0.0007)	1.9130 (0.0045)	0.0002 (0.0000)	29194	4.41
13	0.0383 (0.0003)	0.0000 (0.0000)	0.0149 (0.0000)	0.0833 (0.0007)	1.8953 (0.0041)	0.0002 (0.0000)	29283	3.33
14	0.0409 (0.0004)	0.0000 (0.0000)	0.0151 (0.0000)	0.0781 (0.0006)	1.8757 (0.0048)	0.0002 (0.0000)	29332	2.32
15	0.0572 (0.0004)	0.0000 (0.0000)	0.0156 (0.0000)	0.0559 (0.0004)	1.7400 (0.0041)	0.0002 (0.0000)	29377	—

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.

Model	A. Three-factor model					B. 15-factor model				
Maturity	Mean	Rmse	Auto	Max	VR	Mean	Rmse	Auto	Max	VR
1 m	-0.68	7.47	0.86	43.93	99.83	0.02	0.62	0.36	5.40	100.00
2 m	0.63	3.82	0.69	37.42	99.96	0.01	1.76	0.52	16.31	99.99
3 m	1.61	5.03	0.85	42.54	99.93	-0.11	1.79	0.60	18.96	99.99
6 m	0.39	6.78	0.93	24.05	99.86	0.04	1.06	0.59	8.78	100.00
9 m	-1.74	6.88	0.89	32.06	99.86	0.38	0.92	0.69	4.31	100.00
12 m	-3.06	6.74	0.79	33.00	99.88	-0.49	1.21	0.06	4.71	100.00
2 y	2.11	6.17	0.81	24.38	99.86	0.28	1.09	-0.02	4.52	100.00
3 y	1.97	6.90	0.88	34.12	99.78	-0.19	0.75	0.36	3.88	100.00
4 y	0.87	6.32	0.90	33.48	99.76	-0.04	0.81	0.16	8.08	100.00
5 y	-0.21	5.85	0.90	27.63	99.76	0.07	0.73	0.20	4.60	100.00
7 y	-1.89	5.55	0.92	17.32	99.77	0.08	0.70	0.35	6.86	100.00
10 y	-2.35	5.17	0.89	18.65	99.78	-0.12	0.95	0.23	9.00	99.99
15 y	0.88	3.87	0.86	13.14	99.82	0.00	0.72	0.29	4.68	99.99
20 y	1.91	5.35	0.90	17.64	99.66	0.08	0.79	0.33	6.90	99.99
30 y	-0.76	9.67	0.95	31.88	98.68	-0.09	0.71	0.23	4.82	99.99
Average	-0.02	6.11	0.87	28.75	99.75	-0.00	0.98	0.33	7.45	99.99

Table 4

In-sample predictive variations

Entries reports 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.

Model	A. AR(1)				B. Three-factor model				C. 15-factor model			
h (weeks)	1	2	3	4	1	2	3	4	1	2	3	4
LIBOR/swap maturity:												
1 m	25.85	43.84	57.50	68.12	-0.71	32.92	42.84	47.58	21.71	40.82	52.16	58.02
2 m	23.83	36.65	47.28	55.05	-1.94	15.23	23.31	28.19	17.65	28.50	37.00	43.13
3 m	22.82	32.19	41.34	47.49	-50.31	-12.95	1.57	9.93	8.78	21.86	29.17	33.70
6 m	20.85	25.00	31.90	36.31	-87.43	-42.16	-24.57	-13.88	5.77	12.56	16.94	19.30
9 m	20.22	19.35	23.79	27.10	-67.23	-38.76	-28.15	-20.99	1.30	4.99	7.06	7.95
12 m	21.45	17.53	20.58	22.84	-39.25	-26.45	-21.32	-17.32	6.85	3.71	3.07	2.53
2 y	4.26	7.93	9.65	11.53	-17.52	-5.21	-2.06	-0.99	-1.12	-1.44	-2.29	-3.05
3 y	3.64	6.81	8.64	10.30	-18.64	-4.28	-0.52	0.85	-1.25	-1.75	-2.30	-2.72
4 y	4.75	7.23	9.00	10.32	-15.20	-3.56	-0.49	0.32	0.53	-0.23	-0.76	-1.30
5 y	3.35	6.41	8.36	9.89	-16.34	-5.99	-3.09	-2.39	-0.48	-0.89	-0.91	-1.24
7 y	3.44	6.43	8.08	9.54	-20.20	-10.37	-8.14	-7.14	-0.56	-0.57	-0.87	-1.03
10 y	3.53	6.19	7.87	9.21	-20.23	-12.00	-10.41	-9.21	0.12	0.12	-0.12	-0.22
15 y	2.71	5.06	6.72	7.93	-11.47	-8.24	-7.60	-6.73	0.92	0.38	0.35	0.19
20 y	2.42	4.90	6.69	7.67	-17.91	-10.09	-7.78	-5.92	0.17	-0.24	-0.15	-0.28
30 y	3.09	5.44	7.16	8.25	-60.10	-25.86	-15.76	-9.32	-0.15	0.15	-0.11	-0.01

Table 5

Out-of-sample forecasting

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.

Model	A. AR(1)				B. 15-factor model							
Statistics	Predictive variation				Predictive variation				*t*-statistics			
h (weeks)	1	2	3	4	1	2	3	4	1	2	3	4
LIBOR/swap maturity:												
1 m	-1.57	-3.50	-5.40	-7.75	24.24	42.22	52.64	58.47	1.73	3.49	4.81	6.34
2 m	-1.50	-3.64	-5.77	-8.29	19.59	31.45	39.98	46.22	1.68	3.48	5.03	6.86
3 m	-1.98	-4.19	-6.28	-8.94	9.80	24.77	32.47	37.76	1.69	4.75	6.33	7.34
6 m	-3.36	-6.34	-8.62	-11.83	8.45	16.58	21.26	24.57	2.46	4.58	5.83	6.57
9 m	-4.52	-7.90	-10.35	-14.11	4.71	8.56	11.15	13.06	2.26	3.53	4.15	4.54
12 m	-4.90	-8.45	-10.93	-14.95	7.94	5.19	5.10	5.40	3.63	2.33	2.02	1.99
2 y	-3.00	-5.91	-8.85	-12.21	-1.08	-1.19	-2.14	-2.87	-0.94	-0.90	-1.33	-1.62
3 y	-2.44	-4.78	-7.26	-10.25	-0.92	-1.44	-2.18	-2.69	-1.11	-1.39	-1.79	-1.96
4 y	-2.11	-4.13	-6.34	-9.02	0.22	-0.73	-1.25	-1.90	0.23	-0.73	-1.15	-1.62
5 y	-1.84	-3.54	-5.49	-7.89	-0.76	-1.25	-1.33	-1.64	-1.10	-1.64	-1.56	-1.73
7 y	-1.43	-2.74	-4.41	-6.39	-0.39	-0.53	-0.84	-1.04	-0.79	-1.15	-1.58	-1.72
10 y	-1.07	-2.06	-3.43	-5.12	-0.40	-0.26	-0.61	-0.74	-0.70	-0.45	-1.00	-1.16
15 y	-0.89	-1.68	-2.97	-4.49	0.30	-0.12	-0.21	-0.49	0.60	-0.28	-0.44	-0.94
20 y	-0.76	-1.52	-2.80	-4.25	0.56	-0.11	-0.39	-0.56	1.00	-0.26	-0.80	-1.04
30 y	-0.63	-1.30	-2.35	-3.65	0.02	-0.05	-0.40	-0.50	0.04	-0.11	-0.81	-0.91

Table 6

Scaling in risks and risk premia

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.

Θ	Estimates	Standard Errors
σ_1	0.0276	(0.0001)
$\gamma_0\sigma_1^2$	-0.0019	(0.0001)
$\gamma_1\sigma_1^2$	-0.0520	(0.0004)
$\gamma_2\sigma_1^2$	-0.1634	(0.0006)
b	1.7276	(0.0032)
s_σ	-0.2408	(0.0021)
s_0	0.2532	(0.0111)
s_1	0.0010	(0.0064)
s_2	-1.7617	(0.0038)

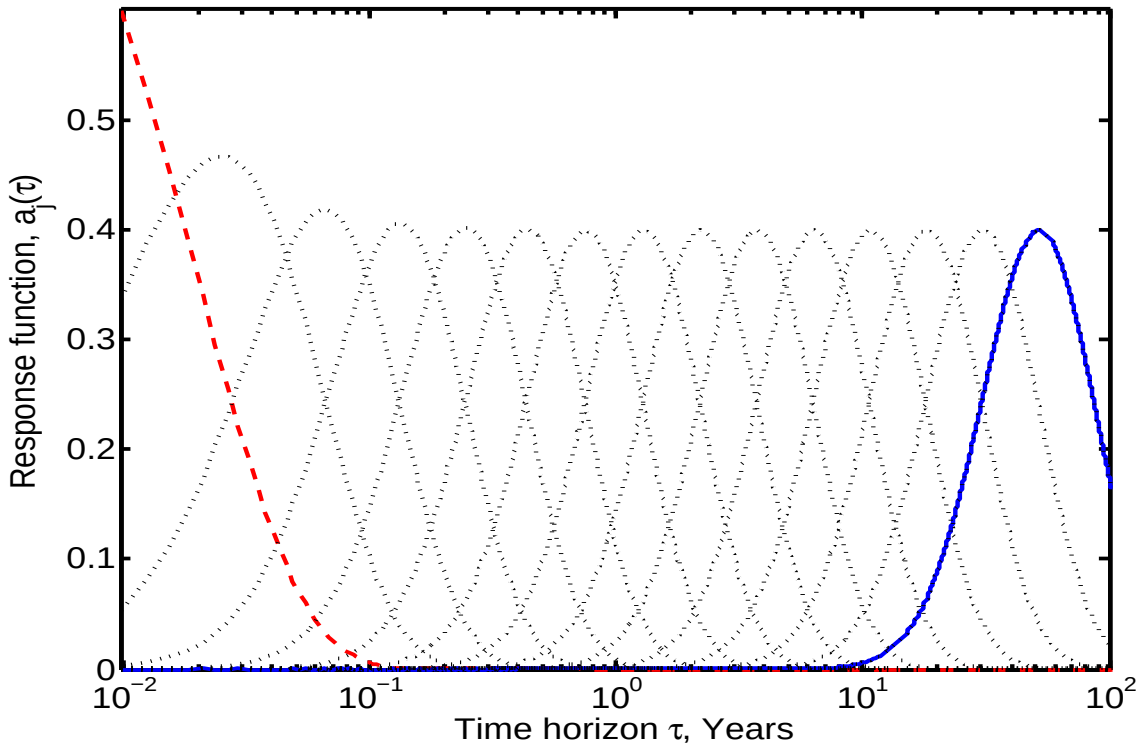

Figure 1

The instantaneous interest rate response to unit shocks from different frequency components.

Lines plot the response of the instantaneous interest rate to unit shocks from each of the 15 frequency components across different time horizons. The solid line denotes response to shocks from the lowest frequency $dW_{1,t}$, the dashed line denotes response to shocks from the highest frequency $dW_{15,t}$. The dotted lines represent responses to intermediate frequency components. The responses are computed with the parameters $\kappa_r = 1/30$, $b = 1.69$, and $n = 15$.

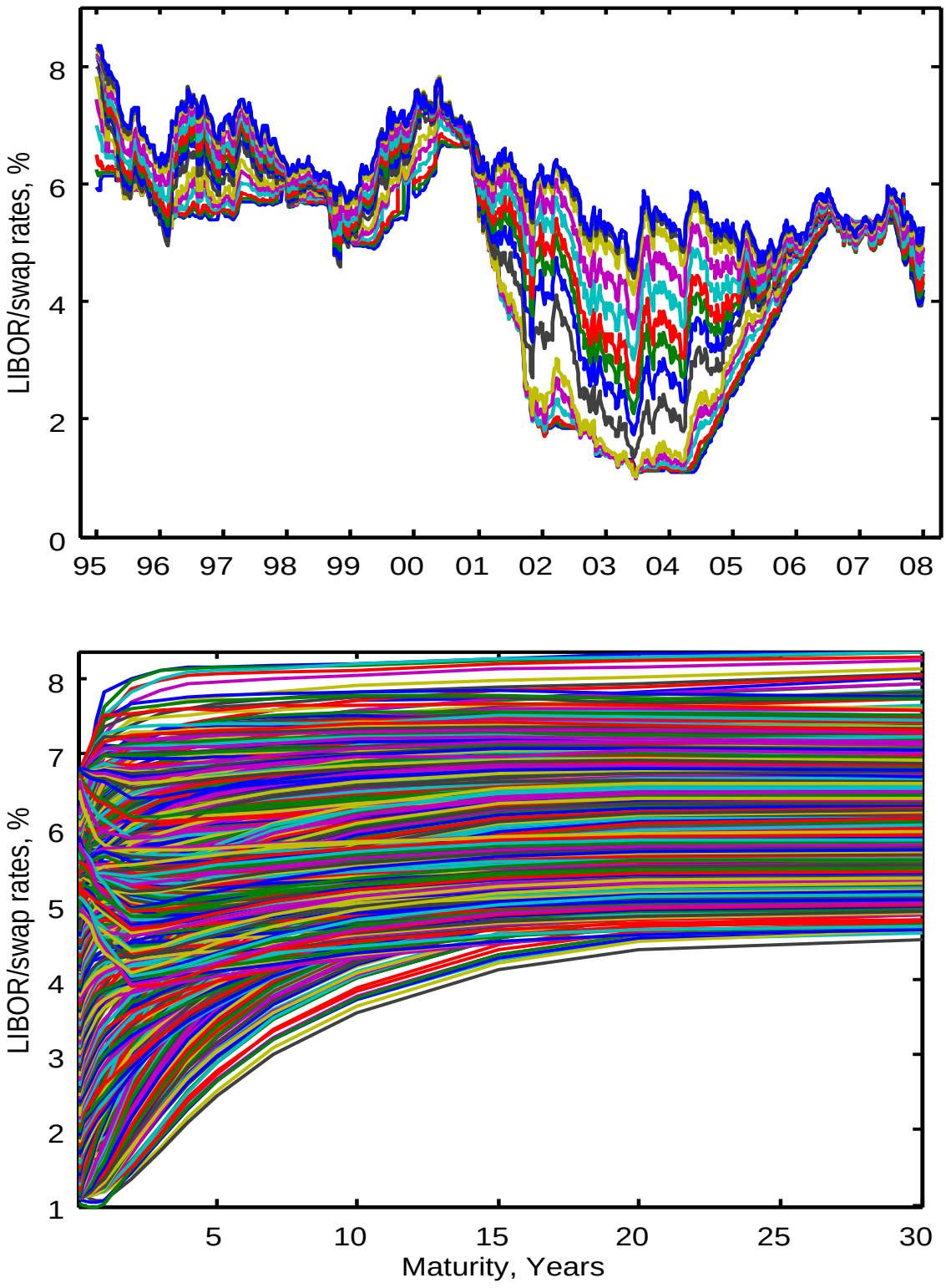

Figure 2

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.

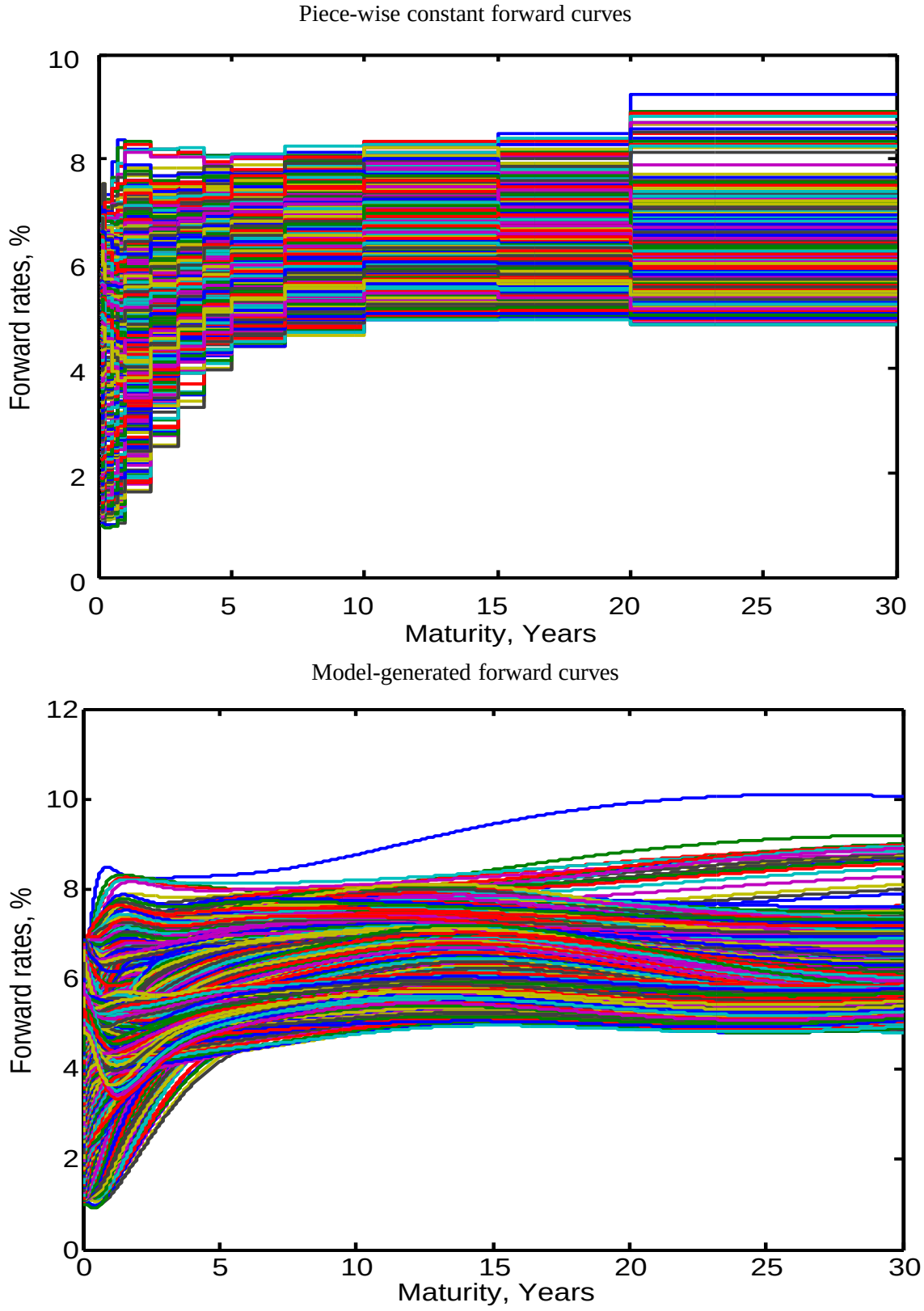

Figure 3

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.

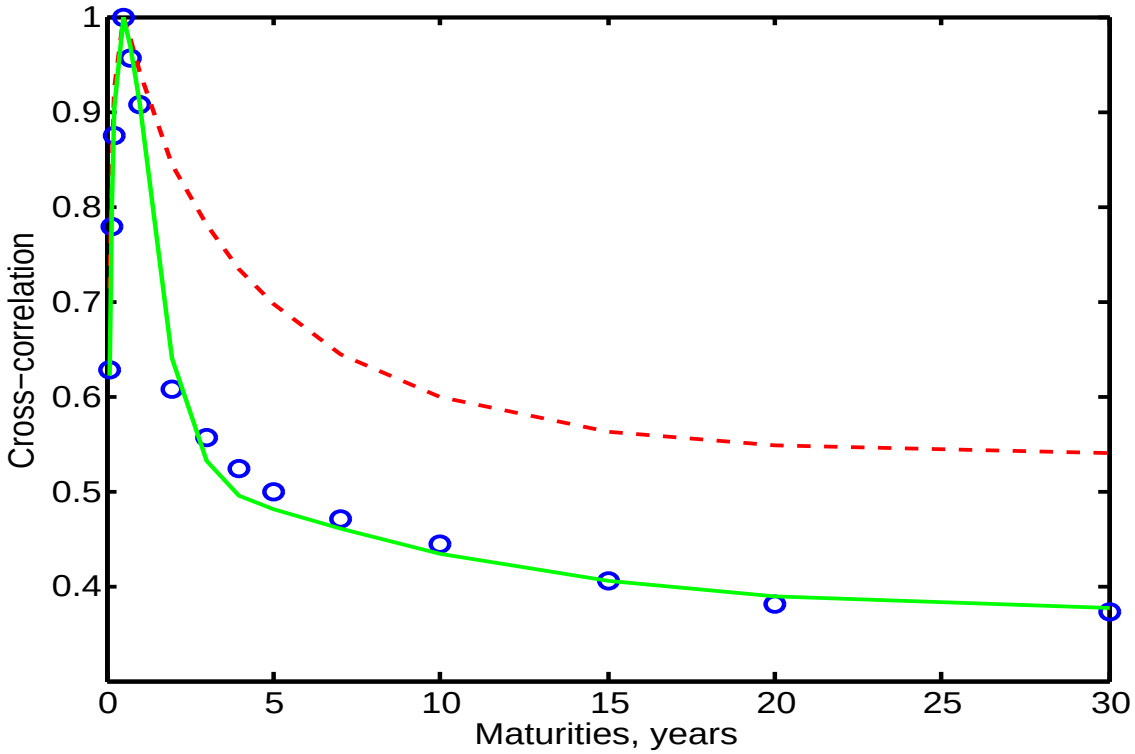

Figure 4

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.

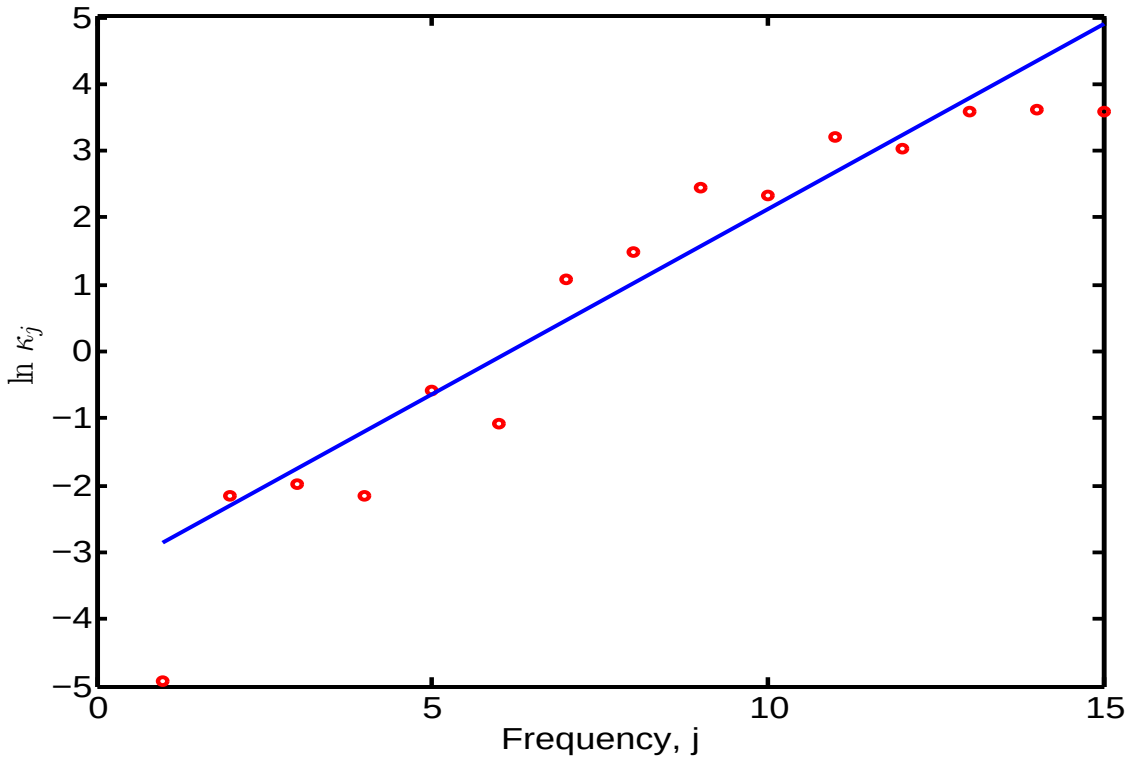

Figure 5

The scaling of κ_j.

The circles are estimated as free parameters. The solid line is generated from the benchmark model with the scaling $\kappa_i = \kappa_1 b^{i-1}$.