Linear-Rational Term Structure Models*

Damir Filipović Martin Larsson Anders Trolle
Swiss Finance Institute @ EPFL

5 October 2013

Abstract

We introduce the class of linear-rational term structure models, where the state price density is modeled such that bond prices become linear-rational functions of the current state. This class is highly tractable with several distinct advantages: i) ensures non-negative interest rates, ii) easily accommodates unspanned factors affecting volatility and risk premia, and iii) admits analytical solutions to swaptions. For comparison, affine term structure models can match either i) or ii), but not both simultaneously, and never iii). A parsimonious specification of the model with three term structure factors and one, or possibly two, unspanned factors has a very good fit to both interest rate swaps and swaptions since 1997. In particular, the model captures well the dynamics of the term structure and volatility during the recent period of near-zero interest rates.

1 Introduction

The current environment with near-zero interest rates creates difficulties for many existing term structure models, most notably Gaussian models that invariably place

*PRELIMINARY VERSION. The authors wish to thank seminar participants at the 10th German Probability and Statistics Days in Mainz, the Conference on Stochastic Analysis and Applications in Lausanne, The Seventh Bachelier Colloquium on Mathematical Finance and Stochastic Calculus in Metabief, Conference of Current Topics in Mathematical Finance in Vienna, the Princeton-Lausanne Workshop on Quantitative Finance in Princeton, and the 29:th European Meeting of Statisticians in Budapest for their comments. The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n. 307465-POLYTE.

large probabilities on negative future rates. Models that respect the zero lower bound on interest rates exist but are often restricted in their ability to accommodate unspanned factors affecting volatility and risk premia and to price many interest rate derivatives. In light of these limitations, the purpose of this paper is twofold: First, we introduce a new class of term structure models, the *linear-rational*, which is highly tractable and i) ensures non-negative interest rates, ii) easily accommodates unspanned factors affecting volatility and risk premia, and iii) admits analytical solutions to swaptions—an important class of interest rate derivatives that underlie the pricing and hedging of mortgage-backed securities, callable agency securities, life insurance products, and a wide variety of structured products. Second, we perform an extensive empirical analysis, focusing in particular on the recent period of near-zero interest rates.

The first contribution of the paper is to introduce the class of linear-rational term structure models. A sufficient condition for the absence of arbitrage opportunities in a model of a financial market is the existence of a *state price density*: a positive adapted process ζ_t such that the price $\Pi(t,T)$ at time t of any time T cash-flow, C_T say, is given by

$$\Pi(t,T) = \frac{1}{\zeta_t}\mathbb{E}[\zeta_T C_T \mid \mathcal{F}_t], \tag{1}$$

where we suppose there is a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ on which all random quantities are defined. Following Constantinides (1992), our approach to modeling the term structure is to directly specify the state price density. Specifically, we assume a multivariate factor process with a drift that is affine in the current state, and a state price density, which is also an affine function of the current state. In this case, zero-coupon bond prices and the short rate become linear-rational functions of the current state, which is why we refer to the framework as linear-rational. One attractive feature of the framework is that one can easily ensure non-negative interest rates. Another attractive feature is that the martingale component of the factor process does not affect the term structure. This implies that one can easily allow for factors that affect prices of interest rate derivatives without affecting bond prices. Assuming that the factor process has diffusive dynamics, we show that the state vector can be partitioned into factors that affect the term structure, factors that affect interest rate volatility but not the term structure (unspanned stochastic volatility, or USV, factors), and factors that neither affect the term structure nor interest rate volatility but may nevertheless have an indirect impact on interest rate derivatives. Assuming further that the factor process is of the square-root type, we show how swaptions can be priced analytically. This specific model is termed the linear-rational square-root (LRSQ) model.

The second contribution of the paper is an extensive empirical analysis of the LRSQ model. We utilize a panel data set consisting of term structures of swap rates and swaption implied volatilities. The sample period is from January 1997 to November 2012. Previous research has shown that a large fraction of the variation in interest rate volatility is largely unrelated to variation in the term structure.[1] Here, we provide an important qualification to this result: volatility becomes gradually more level-dependent as the underlying interest rate approaches the zero lower bound. For instance, conditional on the 1-year swap rate being between zero and one percent, a regression of weekly changes in the implied volatility of the 1-year swap rate on weekly changes in the 1-year swap rate itself produces a highly significant and positive regression coefficient and an R^2 of 0.46. For comparison, unconditionally, the regression coefficient is much lower and the R^2 only 0.05.

The model is estimated by maximum likelihood in conjunction with the Kalman filter. We show that a specification of the model with three term structure factors and one, or possibly two, USV factors gives a very good fit to both interest rate swaps and swaptions simultaneously. This holds true also for the part of the sample period where short-term rates were very close to the zero lower bound. Moreover, we show that the model captures the increasing level-dependence in volatility as interest rates approach the zero lower bound.

A special case of our general linear-rational framework is the model considered by Carr, Gabaix, and Wu (2009). However, the factor process in their model is time-inhomogeneous and non-stationary, while the LRSQ model that we evaluate empirically is time-homogeneous and stationary. Furthermore, the volatility structure in their model is very different from the one in the LRSQ model.[2]

The affine framework, see, e.g., Duffie and Kan (1996) and Dai and Singleton (2000), is arguably the dominant one in the term structure literature. In the affine framework one can either ensure non-negative interest rates (which requires all factors to be of the square-root type) or accommodate USV (which requires at least one conditionally Gaussian factor), but not both.[3] Furthermore, no affine model admits analytical solutions to swaptions. In contrast, the linear-rational framework accommodates all three features.

[1]See, e.g., Collin-Dufresne and Goldstein (2002), Heidari and Wu (2003), Andersen and Benzoni (2010), Li and Zhao (2006), Li and Zhao (2009), Trolle and Schwartz (2009), and Collin-Dufresne, Goldstein, and Jones (2009).

[2]More generally, the linear-rational framework is related to the frameworks in Rogers (1997) and Flesaker and Hughston (1996).

[3]Alternatively, the "shadow rate" model of Black (1995) ensures non-negative interest rates; see, e.g., Kim and Singleton (2012), Bauer and Rudebusch (2013), and Christensen and Rudebusch (2013) for recent applications of this framework.

The paper is structured as follows. Section 2 lays out the general framework, leaving the martingale term of the factor process unspecified. Section 3 specializes to the case where the factor process has diffusive dynamics. Section 4 further specializes to the case where the factor process is of the square-root type. Section 5 describes the data and the estimation approach. Section 6 presents the empirical results. Section 7 concludes. All proofs are given in the appendix.

2 The Linear-Rational Framework

In this section the linear-rational framework is introduced, and explicit formulas for zero-coupon bond prices and short rate are presented. We then discuss how unspanned factors arise in this setting, and how the factor process after a change of coordinates can be decomposed into spanned and unspanned components. We then describe interest rate swaptions, and derive a swaption pricing formula. Finally, the linear-rational framework is compared and contrasted with existing models.

2.1 Term Structure Specification

A linear-rational term structure model consists of two components: a multivariate factor process X_t whose state space is some subset $E \subset \mathbb{R}^d$, and a state price density ζ_t given as a deterministic function of the current state. The linear-rational class becomes tractable due to the interplay between two basic structural assumptions we impose on these components: the factor process is assumed to have a drift that is affine in the current state, and the state price density is similarly required to be an affine function of the current state. More specifically, we assume that X_t is of the form

$$\mathrm{d}X_t = \kappa(\theta - X_t)\mathrm{d}t + \mathrm{d}M_t \tag{2}$$

for some $\kappa \in \mathbb{R}^{d\times d}$, $\theta \in \mathbb{R}^d$, and some martingale M_t.[4] Typically X_t will follow Markovian dynamics, although this is not necessary for this section. Next, the state price density is assumed to be given by

$$\zeta_t = \mathrm{e}^{-\alpha t}\left(\phi + \psi^\top X_t\right), \tag{3}$$

[4] One could replace the drift $\kappa(\theta - X_s)$ in (2) with the slightly more general form $b + \beta X_s$ for some $b \in \mathbb{R}^d$ and $\beta \in \mathbb{R}^{d\times d}$. The gain in generality is moderate (the two parameterizations are equivalent if b lies in the range of β) and is trumped by the gain in notational clarity that will be achieved by using the form (2). The latter form also has the advantage of allowing for a "mean-reversion" interpretation of the drift.

for some $\phi \in \mathbb{R}$ and $\psi \in \mathbb{R}^d$ such that $\phi + \psi^\top x > 0$ for all $x \in E$, and some $\alpha \in \mathbb{R}$. As we discuss below, the role of the parameter α is to ensure that the short rate stays nonnegative.

The affine drift of the factor process implies that conditional expectations have the following simple form, as can be seen from Lemma A.1:

$$\mathbb{E}[X_T \mid \mathcal{F}_t] = \theta + \mathrm{e}^{-\kappa(T-t)}(X_t - \theta), \qquad t \leq T. \tag{4}$$

An immediate consequence is that the zero-coupon bond prices and the short rate become linear-rational functions of the current state, which is is why we refer to this framework as linear-rational. Indeed, the basic pricing formula (1) with $C = 1$ shows that the zero-coupon bond prices are given by $P(t,T) = F(T-t, X_t)$, where

$$F(\tau, x) = \frac{(\phi + \psi^\top \theta)\mathrm{e}^{-\alpha\tau} + \psi^\top \mathrm{e}^{-(\alpha+\kappa)\tau}(x - \theta)}{\phi + \psi^\top x}. \tag{5}$$

The short rate is then obtained via the formula $r_t = -\partial_T \log P(t,T)|_{T=t}$, and is given by

$$r_t = \alpha - \frac{\psi^\top \kappa(\theta - X_t)}{\phi + \psi^\top X_t}. \tag{6}$$

The latter expression clarifies the role of the parameter α; provided that the short rate is bounded from below, we may guarantee that it stays nonnegative by choosing α large enough. This leads to an intrinsic choice of α as the smallest value that yields a nonnegative short rate. In other words, we define

$$\alpha^* = \sup_{x \in E} \frac{\psi^\top \kappa(\theta - x)}{\phi + \psi^\top x} \quad \text{and} \quad \alpha_* = \inf_{x \in E} \frac{\psi^\top \kappa(\theta - x)}{\phi + \psi^\top x}, \tag{7}$$

and set $\alpha = \alpha^*$, provided this is finite. The short rate then satisfies

$$r_t \in [0, \alpha^* - \alpha_*] \qquad (r_t \in [0, \infty) \text{ if } \alpha_* = -\infty).$$

Notice that α^* and α_* depend on the parameters of the process X_t, which are determined through calibration. A crucial step of the model validation process is therefore to verify that the range of possible short rates is sufficiently wide. Finally, notice that whenever the eigenvalues of κ have nonnegative real part, one easily verifies the equality

$$\lim_{\tau \to \infty} -\frac{1}{\tau} \log F(\tau, x) = \alpha,$$

valid for any $x \in E$. In other words, α can be interpreted as an infinite-maturity forward rate.

2.2 Unspanned Factors

Our focus is now to describe the directions $\xi \in \mathbb{R}^d$ such that the term structure remains unchanged when the state vector moves along ξ. It is convenient to carry out this discussion in terms of the kernel of a function.[5]

Definition 2.1. *The* term structure kernel*, denoted by* $\mathcal{U}$*, is given by*

$$\mathcal{U} = \bigcap_{\tau \geq 0} \ker F(\tau, \cdot).$$

That is, $\mathcal{U}$ consists of all $\xi \in \mathbb{R}^d$ such that $\nabla F(\tau, x)^\top \xi = 0$ for all $\tau \geq 0$ and all $x \in E$.[6] Therefore the location of the state X_t along the direction ξ cannot be recovered solely from knowledge of the time t bond prices $P(t, t+\tau)$, $\tau \geq 0$. In this sense the term structure kernel is unspanned by the term structure. In Section 3.1 we will discuss how this notion relates to spanning in the sense of bond market completeness. The following result characterizes $\mathcal{U}$ in terms of the model parameters.

Proposition 2.2. *Assume the term structure is not trivial.*[7] *Then*

$$\mathcal{U} = \bigcap_{p=0}^{d-1} \ker \psi^\top \kappa^p. \tag{8}$$

In the case where κ is diagonalizable, this leads to the following corollary.

Corollary 2.3. *Assume κ is diagonalizable with real eigenvalues, i.e. $\kappa = S^{-1}\Lambda S$ with Λ diagonal and real. Then $\mathcal{U} = \{0\}$ if and only if all eigenvalues of κ are distinct and all components of $S^{-\top}\psi$ are nonzero.*

---

[5] We define the *kernel* of a differentiable function f on E by

$$\ker f = \left\{\xi \in \mathbb{R}^d : \nabla f(x)^\top \xi = 0 \text{ for all } x \in E\right\}.$$

This notion generalizes the standard one: if $f(x) = v^\top x$ is linear, for some $v \in \mathbb{R}^d$, then $\nabla f(x) = v$ for all $x \in E$, so $\ker f = \ker v^\top$ coincides with the usual notion of kernel.

[6] Here and in the sequel, $\nabla F(\tau, x)$ denotes the gradient with respect to the x variables.

[7] We say that the term structure is *trivial* if the short rate r_t is constant. In view of (6), this happens if and only if ψ is an eigenvector of $\kappa^\top$ with eigenvalue λ satisfying $\lambda(\phi + \psi^\top \theta) = 0$. In this case, we have $r_t \equiv \alpha + \lambda$ and $\mathcal{U} = \mathbb{R}^d$, while the right side of (8) equals $\ker \psi^\top$. The assumption that the term structure be not trivial will be in force throughout the paper.

We now transform the state space so that the unspanned directions correspond to the last components of the state vector. To this end, first let S be any invertible linear transformation on $\mathbb{R}^d$. The transformed factor process $\widehat{X}_t = SX_t$ satisfies the affine drift dynamics

$$\mathrm{d}\widehat{X}_t = \widehat{\kappa}(\widehat{\theta} - \widehat{X}_t)\mathrm{d}t + \mathrm{d}\widehat{M}_t,$$

where

$$\widehat{\kappa} = S\kappa S^{-1}, \qquad \widehat{\theta} = S\theta, \qquad \widehat{M}_t = SM_t. \tag{9}$$

Defining also

$$\widehat{\phi} = \phi, \qquad \widehat{\psi} = S^{-\top}\psi, \tag{10}$$

we have $\zeta_t = \mathrm{e}^{-\alpha t}(\widehat{\phi} + \widehat{\psi}^\top \widehat{X}_t)$. This gives a linear-rational term structure model that is equivalent to the original one. Suppose now that S maps the term structure kernel into the standard basis of $\mathbb{R}^d = \mathbb{R}^m \times \mathbb{R}^n$,

$$S(\mathcal{U}) = \{0\} \times \mathbb{R}^n \tag{11}$$

where $n = \dim \mathcal{U}$ and $m = d - n$. Decomposing the transformed factor process accordingly, $\widehat{X}_t = (Z_t, U_t)$, our next result and the subsequent discussion will show that Z_t affects the term structure, while U_t does not.

Theorem 2.4. *Let $m, n \geq 0$ be integers with $m + n = d$. Then* (11) *holds if and only if the transformed model parameters* (9)–(10) *satisfy:*

(i) $\widehat{\psi} = (\widehat{\psi}_Z, 0) \in \mathbb{R}^m \times \mathbb{R}^n$;

(ii) $\widehat{\kappa}$ *has block lower triangular structure,*

$$\widehat{\kappa} = \begin{pmatrix} \widehat{\kappa}_{ZZ} & 0 \\ \widehat{\kappa}_{UZ} & \widehat{\kappa}_{UU} \end{pmatrix} \in \mathbb{R}^{(m+n)\times(m+n)};$$

(iii) *The upper left block $\widehat{\kappa}_{ZZ}$ of $\widehat{\kappa}$ satisfies*

$$\bigcap_{p=0}^{m-1} \ker \widehat{\psi}_Z^\top \widehat{\kappa}_{ZZ}^p = \{0\}.$$

In this case, the dimension of the term structure kernel $\mathcal{U}$ equals n.

Assuming that (11) holds, and writing $Sx = (z, u) \in \mathbb{R}^m \times \mathbb{R}^n$ and $\widehat{\theta} = (\widehat{\theta}_Z, \widehat{\theta}_U)$, we now see that

$$\widehat{F}(\tau, z) = F(\tau, x) = \frac{(\widehat{\phi} + \widehat{\psi}_Z^\top \widehat{\theta}_Z)\mathrm{e}^{-\alpha\tau} + \widehat{\psi}_Z^\top \mathrm{e}^{-(\alpha + \widehat{\kappa}_{ZZ})\tau}(z - \widehat{\theta}_Z)}{\widehat{\phi} + \widehat{\psi}_Z^\top z}$$

does not depend on u. Hence the bond prices are given by $P(t, T) = \widehat{F}(T - t, Z_t)$. This gives a clear interpretation of the components of U_t as *unspanned factors*: their values do not influence the current term structure. As a consequence, a snapshot of the term structure at time t does not provide any information about U_t. The sub-vector Z_t, on the other hand, directly impacts the term structure, and can be reconstructed from a snapshot of the term structure at time t, under mild technical conditions. For this reason we refer to the components of Z_t as *term structure factors*. The following proposition formalizes the above discussion.

Proposition 2.5. *The term structure $\widehat{F}(\tau, z)$ is injective if and only if $\widehat{\kappa}_{ZZ}$ is invertible and $\widehat{\phi} + \widehat{\psi}_Z^\top \widehat{\theta}_Z \neq 0$.*[8]

In view of Theorem 2.4, the dynamics of $\widehat{X}_t = (Z_t, U_t)$ can be decomposed into term-structure dynamics

$$\mathrm{d}Z_t = \widehat{\kappa}_{ZZ}(\widehat{\theta}_Z - Z_t)\mathrm{d}t + \mathrm{d}\widehat{M}_{Zt} \tag{12}$$

and unspanned factor dynamics

$$\mathrm{d}U_t = \Big(\widehat{\kappa}_{UZ}(\widehat{\theta}_Z - Z_t) + \widehat{\kappa}_{UU}(\widehat{\theta}_U - U_t)\Big)\,\mathrm{d}t + \mathrm{d}\widehat{M}_{Ut}$$

where we denote $\widehat{M}_t = (\widehat{M}_{Zt}, \widehat{M}_{Ut})$. Moreover, the state price density can be written

$$\zeta_t = \mathrm{e}^{-\alpha t}(\widehat{\phi} + \widehat{\psi}_Z^\top Z_t). \tag{13}$$

Now, since the process Z_t has an affine drift that depends only on Z_t itself, and since the state price density also depends only on Z_t, we can view Z_t as the factor process of an m-dimensional linear-rational term structure model (12)–(13), which is equivalent to (2)–(3). In view of Proposition 2.2, this leads to an interpretation of Theorem 2.4(iii): the model (12)–(13) is minimal in the sense that its own term structure kernel is trivial.

[8] Injectivity means that if $\widehat{F}(\tau, z) = \widehat{F}(\tau, z')$ for all $\tau \geq 0$, then $z = z'$. In other words, if $\widehat{F}(\tau, Z_t)$ is known for all $\tau \geq 0$, we can back out the value of Z_t.

Carrying this observation further, we see that if the unspanned factors U_t do not enter into the dynamics of $\widehat{M}_{Zt}$, then Z_t is a fully autonomous Markov process, assuming that X_t is Markovian. In this case U_t would be redundant and play no role in the model. However, if U_t does enter into the dynamics of $\widehat{M}_{Zt}$, then the unspanned factors would not be redundant. This situation is what gives rise to USV, and is discussed in Section 3.2.

Finally, note that even if the term structure kernel is trivial, $\mathcal{U} = \{0\}$, the short end of the term structure may nonetheless be insensitive to movements of the state along certain directions. In view of Proposition 2.2, for $d \geq 3$ we can have $\mathcal{U} = \{0\}$ while still there exists a non-zero vector ξ such that $\psi^\top \xi = \psi^\top \kappa \xi = 0$. This implies that the short rate function is constant along ξ, see (6). On the other hand, we can still, in the generic case, recover X_t from a snapshot of the term structure, see Proposition 2.5.

2.3 Swaps and Swaptions

The linear-rational term structure models have the important advantage of allowing for tractable swaption pricing.

A fixed versus floating interest rate swap is specified by a tenor structure of reset and payment dates $T_0 < T_1 < \cdots < T_n$, where we take $\Delta = T_i - T_{i-1}$ to be constant for simplicity, and a pre-determined annualized rate K. At each date T_i, $i = 1, \ldots, n$, the fixed leg pays ΔK and the floating leg pays LIBOR accrued over the preceding time period.[9] From the perspective of the fixed-rate payer, the value of the swap at time $t \leq T_0$ is given by[10]

$$\Pi_t^{\text{swap}} = P(t, T_0) - P(t, T_n) - \Delta K \sum_{i=1}^{n} P(t, T_i). \tag{14}$$

The time-t forward swap rate, S_t, is the strike rate K that makes the value of the

[9]For expositional ease, we assume that the payments on the fixed and floating legs occur at the same frequency. In reality, in the USD market fixed-leg payments occur at a semi-annual frequency, while floating-leg payments occur at a quarterly frequency. However, only the frequency of the fixed-leg payments matter for the valuation of the swap.

[10]This valuation equation, which was the market standard until a few years ago, implicitly assumes that payments are discounted with a rate that incorporates the same credit and liquidity risk as LIBOR. In reality, swap contracts are virtually always collateralized, which makes swap (and swaption) valuation significantly more involved; see, e.g., Johannes and Sundaresan (2007) and Filipovic and Trolle (2013). In the present paper we simplify matters by adhering to the formula (14).

swap equal to zero. It is given by

$$S_t = \frac{P(t,T_0) - P(t,T_n)}{\sum_{i=1}^n \Delta P(t,T_i)}. \tag{15}$$

The forward swap rate becomes the spot swap rate at time T_0.

A payer swaption is an option to enter into an interest rate swap, paying the fixed leg at a pre-determined rate and receiving the floating leg. A European payer swaption expiring at T_0 on a swap with the characteristics described above has a value at expiration of

$$C_{T_0} = \left(\Pi_{T_0}^{\mathrm{swap}}\right)^+ = \left(\sum_{i=0}^n c_i P(T_0,T_i)\right)^+ = \frac{1}{\zeta_{T_0}}\left(\sum_{i=0}^n c_i \mathbb{E}\left[\zeta_{T_i} \mid \mathcal{F}_{T_0}\right]\right)^+,$$

for coefficients c_i that can easily be read off the expression (14).

In a linear-rational term structure model, the conditional expectations $\mathbb{E}\left[\zeta_{T_i} \mid \mathcal{F}_{T_0}\right]$ are affine functions of X_{T_0}, with coefficients that are explicitly given in terms of the model parameters, see Lemma A.1. Specifically, we have

$$C_{T_0} = \frac{1}{\zeta_{T_0}} p_{\mathrm{swap}}(X_{T_0})^+,$$

where p_{swap} is the explicit affine function

$$p_{\mathrm{swap}}(x) = \sum_{i=0}^n c_i \mathrm{e}^{-\alpha T_i}\left(\phi + \psi^\top \theta + \psi^\top \mathrm{e}^{-\kappa(T_i - T_0)}(x-\theta)\right).$$

The swaption price at time $t \le T_0$ is then obtained by an application of the fundamental pricing formula (1), which yields

$$\Pi_t^{\mathrm{swpt}} = \frac{1}{\zeta_t}\mathbb{E}[\zeta_{T_0} C_{T_0} \mid \mathcal{F}_t] = \frac{1}{\zeta_t}\mathbb{E}\left[p_{\mathrm{swap}}(X_{T_0})^+ \mid \mathcal{F}_t\right]. \tag{16}$$

To compute the price one has to evaluate the conditional expectation on the right side of (16). If the conditional distribution of X_{T_0} given $\mathcal{F}_t$ is known, this can be done via direct numerical integration over $\mathbb{R}^d$. This appears to be a challenging problem in general; fortunately there is an alternative approach based on Fourier transform methods that tends to perform better in practice.

Theorem 2.6. *Define $\widehat{q}(z) = \mathbb{E}\left[\exp\left(z\, p_{\mathrm{swap}}(X_{T_0})\right) \mid \mathcal{F}_t\right]$ for every $z \in \mathbb{C}$ such that the conditional expectation is well-defined. Pick any $\mu > 0$ such that $\widehat{q}(\mu) < \infty$. Then the swaption price is given by*

$$\Pi_t^{\mathrm{swpt}} = \frac{1}{\zeta_t\, \pi}\int_0^\infty \mathrm{Re}\left[\frac{\widehat{q}(\mu + \mathrm{i}\lambda)}{(\mu + \mathrm{i}\lambda)^2}\right] \mathrm{d}\lambda.$$

Theorem 2.6 reduces the problem of computing an integral over $\mathbb{R}^d$ to that of computing a simple line integral. Of course, there is a price to pay: we now have to evaluate $\widehat{q}(\mu+\mathrm{i}\lambda)$ efficiently as λ varies through $\mathbb{R}_+$. This problem can be approached in various ways depending on the specific class of factor processes under consideration. In our empirical evaluation we focus on affine factor processes, for which computing $\widehat{q}(z)$ amounts to solving a system of ordinary differential equations, see Section 4.2.

It is often more convenient to represent swaption prices in terms of implied volatilities. In the USD market, the market standard is the "normal" (or "absolute" or "basis point") implied volatility, which is the volatility parameter that matches a given price when plugged into the pricing formula that assumes a normal distribution for the underlying forward swap rate.[11] When the swaption strike is equal to the forward swap rate ($K = S_t$, see (15)), there is a particularly simple relation between the swaption price and the normal implied volatility, $\sigma_{N,t}$, given by

$$\Pi_t^{\mathrm{swpt}} = \sqrt{T_0 - t}\frac{1}{\sqrt{2\pi}}\left(\sum_{i=1}^{n} \Delta P(t,T_i)\right)\sigma_{N,t}; \tag{17}$$

see, e.g., Corp (2012).

2.4 Comparison with Other Models

When the factor process X_t is Markovian, the linear-rational framework falls in the broad class of models contained under the potential approach laid out in Rogers (1997). There the state price density is modeled by the expression

$$\zeta_t = \mathrm{e}^{-\alpha t} R_\alpha g(X_t),$$

where R_α is the resolvent operator corresponding to the Markov process X_t, and g is a suitable function. In our setting we would have $R_\alpha g(x) = \phi + \psi^\top x$, and thus $g(x) = (\alpha - \mathcal{G})R_\alpha g(x) = \alpha\phi - \psi^\top \kappa\theta + \psi^\top(\alpha+\kappa)x$, where $\mathcal{G}$ is the generator of X_t.

Another related setup which slightly pre-dates the potential approach is the framework of Flesaker and Hughston (1996). The state price density now takes the form

$$\zeta_t = \int_t^\infty M_{tu}\mu(u)\mathrm{d}u,$$

[11] Alternatively, a price may be represented in terms of "log-normal" (or "percentage") implied volatility, which assumes a log-normal distribution for the underlying forward swap rate.

where for each u, $(M_{tu})_{0\leq t\leq u}$ is a martingale. The Flesaker-Hughston framework is related to the potential approach (and thus to the linear-rational framework) via the representation

$$\mathrm{e}^{-\alpha t}R_\alpha g(X_t) = \int_t^\infty \mathbb{E}\left[\mathrm{e}^{-\alpha u} g(X_u) \mid \mathcal{F}_t\right] \mathrm{d}u,$$

which implies $M_{tu}\mu(u) = \mathbb{E}[\mathrm{e}^{-\alpha u} g(X_u) \mid \mathcal{F}_t]$. The linear-rational framework fits into this template by taking $\mu(u) = \mathrm{e}^{-\alpha u}$ and $M_{tu} = \mathbb{E}[g(X_u) \mid \mathcal{F}_t] = \alpha\phi + \alpha\psi^\top\theta + \psi^\top(\alpha+\kappa)\mathrm{e}^{-\kappa(u-t)}(x-\theta)$, where $g(x) = \alpha\phi - \psi^\top\kappa\theta + \psi^\top(\alpha+\kappa)x$ was chosen as above. One member of this class, introduced in Flesaker and Hughston (1996), is the one-factor rational log-normal model. The simplest time-homogeneous version of this model is, in the notation of (2)–(3), obtained by taking ϕ and ψ positive, $\kappa = \theta = 0$, and letting the martingale part M_t of the factor process X_t be geometric Brownian motion.

Finally, a more recently introduced set of models that are closely related to those mentioned above is the linearity-generating family studied in Gabaix (2009) and Carr, Gabaix, and Wu (2009). The model considered by Carr, Gabaix, and Wu (2009) falls within the linear-rational class: One sets $\phi = 0$, $\alpha = 0$, $\theta = 0$, and lets the martingale part M_t of the factor process X_t be given by

$$\mathrm{d}M_t = \mathrm{e}^{-\kappa t}\beta \mathrm{d}Z_t,$$

where β is a vector in $\mathbb{R}^d$ and Z_t is an exponential martingale of the form

$$\frac{\mathrm{d}Z_t}{Z_t} = \sum_{i=1}^m \sqrt{v_{it}}\mathrm{d}B_{it},$$

for independent Brownian motions B_{it} and processes v_{it} following square-root dynamics. The factor process in this model is non-stationary due the time-inhomogeneous volatility specification. In fact, assuming the eigenvalues of κ have positive real part (which is the case in Carr, Gabaix, and Wu (2009)), the volatility of X_t tends to zero as time goes to infinity, and the state itself converges to zero almost surely. The models we consider in our empirical analysis are time-homogeneous and stationary. They also have a volatility structure that is very different from the specification in Carr, Gabaix, and Wu (2009).

A common feature of all the above models is that bond prices are given as a ratio of two functions of the state. This is of course an artifact of the form of the pricing equation (1), and the fact that the state price density is the primitive object that is being modeled.

3 Linear-Rational Diffusion Models

We now specialize the linear-rational framework (2)–(3) to the case where the factor process has diffusive dynamics of the form

$$\mathrm{d}X_t = \kappa(\theta - X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}B_t. \tag{18}$$

Here $\sigma : E \to \mathbb{R}^{d\times d}$ is measurable, and B_t is d-dimensional Brownian motion. We denote the diffusion matrix by $a(x) = \sigma(x)\sigma(x)^\top$, and assume that it is differentiable. The goal of this section is two-fold: we first discuss how the notion of spanning in Section 2.2 relates to bond market completeness. Then, in the case where unspanned factors are present, we answer the question of when these unspanned factors give rise to USV.

3.1 Bond Markets

Bond price volatilities will be central to the discussion, so we begin by considering the dynamics of bond prices. To this end, first observe (via a short calculation using Itô's formula) that the dynamics of the state price density can be written

$$\frac{\mathrm{d}\zeta_t}{\zeta_t} = -r_t\mathrm{d}t - \lambda_t^\top \mathrm{d}B_t,$$

where the short rate r_t is given by (6), and $\lambda_t = -\sigma(X_t)^\top\psi/(\phi+\psi^\top X_t)$ is the market price of risk. It then follows that the dynamics of $P(t,T)$ is

$$\frac{\mathrm{d}P(t,T)}{P(t,T)} = \left(r_t + \nu(t,T)^\top\lambda_t\right)\mathrm{d}t + \nu(t,T)^\top\mathrm{d}B_t, \tag{19}$$

where the volatility vector is given by

$$\nu(t,T) = \frac{\sigma(X_t)^\top\nabla F(T-t,X_t)}{F(T-t,X_t)}.$$

It is intuitively clear that a non-trivial term structure kernel gives rise to bond market incompleteness in the sense that not every contingent claim can be hedged using bonds. Conversely, one would expect that whenever the term structure kernel is trivial, bond markets are complete. In this section we confirm this intuition. The following definition of completeness is standard.

Definition 3.1. *We say that* bond markets are complete *if for any $T \geq 0$ and any bounded $\mathcal{F}_T$-measurable random variable C_T, there is a set of maturities $T_1, \ldots, T_m$ and a self-financing trading strategy in the bonds $P(t, T_1), \ldots, P(t, T_m)$ and the money market account, whose value at time T is equal to C_T.*

Our next result clarifies the connection between bond market completeness and the existence of unspanned factors. We assume that the filtration is generated by the Brownian motion, and that the volatility matrix of the factor process itself is almost surely invertible. Otherwise there would be measurable events which cannot be generated by the factor process, and bond market completeness would fail.

Theorem 3.2. *Assume that the filtration $\mathcal{F}_t$ is generated by the Brownian motion B_t, that $\sigma(X_t)$ is invertible $\mathrm{d}t \otimes \mathrm{d}\mathbb{P}$-almost surely, and that $\phi + \psi^\top \theta \neq 0$. Then the following conditions are equivalent:*

(i) *bond markets are complete;*

(ii) $\operatorname{span}\{\nabla F(\tau, X_t) : \tau \geq 0\} = \mathbb{R}^d$, $\mathrm{d}t \otimes \mathrm{d}\mathbb{P}$-*almost surely;*

(iii) $\mathcal{U} = \{0\}$ *and κ is invertible;*

(iv) *The term structure $F(\tau, x)$ is injective.*

3.2 Unspanned Stochastic Volatility Factors

We now refine the discussion in Section 2.2 by singling out those unspanned factors that give rise to USV. To this end we describe directions $\xi \in \mathbb{R}^d$ with the property that movements of the state vector along ξ influence neither the bond return volatilities, nor the covariations between returns on bonds with different maturities.

According to (19), the covariation at time t between the returns on two bonds with maturities T_1 and T_2 is given by $\nu(t, T_1)^\top \nu(t, T_2) = G(T_1 - t, T_2 - t, X_t)$, where we define

$$G(\tau_1, \tau_2, x) = \frac{\nabla F(\tau_1, x)^\top a(x) \nabla F(\tau_2, x)}{F(\tau_1, x)\, F(\tau_2, x)}.$$

In analogy with Definition 2.1 we introduce the following notion:

Definition 3.3. *The* variance-covariance kernel*, denoted by $\mathcal{W}$, is given by*

$$\mathcal{W} = \bigcap_{\tau_1, \tau_2 \geq 0} \ker G(\tau_1, \tau_2, \cdot).$$

That is, $\mathcal{W}$ consists of all $\xi \in \mathbb{R}^d$ such that $\nabla G(\tau_1, \tau_2, x)^\top \xi = 0$ for all $\tau_1, \tau_2 \geq 0$ and all $x \in E$. We say that the model exhibits USV if there are elements of the term structure kernel that do not lie in the variance-covariance kernel—i.e., if $\mathcal{U} \setminus \mathcal{W} \neq \emptyset$.

Analogously to Section 2.2 we may now transform the state space so that the intersection $\mathcal{U} \cap \mathcal{W}$ of the term structure kernel and variance-covariance kernel corresponds to the last components of the state vector. To this end, let S be an invertible linear transformation satisfying (11), with the additional property that

$$S(\mathcal{U} \cap \mathcal{W}) = \{0\} \times \{0\} \times \mathbb{R}^q,$$

where $q = \dim \mathcal{U} \cap \mathcal{W}$, and $p + q = n = \dim \mathcal{U}$. The unspanned factors then decompose accordingly into $U_t = (V_t, W_t)$. Movements of W_t affect neither the term structure, nor bond return volatilities or covariations. In contrast, movements of V_t, while having no effect on the term structure, do impact bond return volatilities or covariations. For this reason we refer to V_t as *USV factors*, whereas W_t is referred to as *residual factors*. Note that the residual factors W_t may still have an indirect impact on the distribution of future bond prices. Example A.9 in the appendix illustrates this fact.

Whether a given linear-rational term structure model exhibits USV depends on how σ interacts with the other parameters of the model. Proposition A.7 in the appendix gives a description of the variance-covariance kernel $\mathcal{W}$, which facilitates checking the presence of USV. As a corollary we obtain the following useful sufficient condition for USV. This condition is, for example, satisfied for the square root model discussed in Section 4. It is stated in terms of the diffusion matrix $\widehat{a}(z, u)$ of the transformed factor process $\widehat{X}_t = (Z_t, U_t)$, given by $\widehat{a}(z, u) = Sa(S^{-1}(z, u))S^\top$.

Corollary 3.4. *Assume for every $j \in \{1, \ldots, n\}$, there exists $i \in \{1, \ldots, m\}$ such that $\widehat{a}_{ii}(z, u)$ is not constant in u_j. Then $\mathcal{U} \cap \mathcal{W} = \{0\}$, and therefore every unspanned factor is in fact a USV factor.*

4 The Linear-Rational Square-Root Model

The primary example of a linear-rational diffusion model (18) with state space $E = \mathbb{R}^d_+$ is the *linear-rational square-root (LRSQ)* model. It is based on a multivariate square-root factor process of the form

$$\mathrm{d}X_t = \kappa(\theta - X_t)\mathrm{d}t + \mathrm{Diag}\left(\sigma_1\sqrt{X_{1t}}, \ldots, \sigma_d\sqrt{X_{dt}}\right)\mathrm{d}B_t, \tag{20}$$

with parameters $\sigma_i > 0$. In this section we consider this model, focusing on how unspanned stochastic volatility can be incorporated, and how swaption pricing can be done efficiently. This lays the groundwork for our empirical analysis.

4.1 Unspanned Stochastic Volatility

The aim is now to construct a large class of LRSQ specifications with m term structure factors and n USV factors. Other constructions are possible, but the one given here is more than sufficient for the applications we are interested in.

As a first step we show that the LRSQ model admits a canonical representation.

Theorem 4.1. *The short rate* (6) *is bounded from below if and only if, after a coordinatewise scaling of the factor process* (20)*, we have* $\zeta_t = \mathrm{e}^{-\alpha t}(1 + \mathbf{1}^\top X_t)$*. In this case, the extremal values in* (7) *are given by* $\alpha^* = \max \mathcal{S}$ *and* $\alpha_* = \min \mathcal{S}$ *where*

$$\mathcal{S} = \left\{ \mathbf{1}^\top \kappa\theta, -\mathbf{1}^\top \kappa_1, \ldots, -\mathbf{1}^\top \kappa_d \right\}.$$

In accordance with this result, we always let the state price density be given by $\zeta_t = \mathrm{e}^{-\alpha t}(1 + \mathbf{1}^\top X_t)$ when considering the LRSQ model.

Now fix nonnegative integers $m \geq n$ with $m + n = d$, representing the desired number of term structure and USV factors, respectively. We start with the invertible linear transformation S on $\mathbb{R}^d$ given by

$$S = \begin{pmatrix} \mathrm{Id}_m & A \\ 0 & \mathrm{Id}_n \end{pmatrix}, \qquad S^{-1} = \begin{pmatrix} \mathrm{Id}_m & -A \\ 0 & \mathrm{Id}_n \end{pmatrix},$$

where $A \in \mathbb{R}^{m \times n}$ is given by

$$A = \begin{pmatrix} \mathrm{Id}_n \\ 0 \end{pmatrix}.$$

The parameters appearing in the description (20) of the factor process X_t can then be specified, and for this it is convenient to introduce the index sets $I = \{1, \ldots, m\}$ and $J = \{m+1, \ldots, d\}$. We write the mean reversion matrix κ in block form as

$$\kappa = \begin{pmatrix} \kappa_{II} & \kappa_{IJ} \\ \kappa_{JI} & \kappa_{JJ} \end{pmatrix},$$

where κ_{IJ} denotes the submatrix whose rows are indexed by I and columns by J, and similarly for κ_{II}, κ_{JI}, κ_{JJ}. We require that κ_{IJ} satisfy the restriction

$$\kappa_{IJ} = \kappa_{II} A - A \kappa_{JJ} + A \kappa_{JI} A. \tag{21}$$

The level of mean reversion is taken to be a vector $\theta = (\theta_I, \theta_J) \in \mathbb{R}^m \times \mathbb{R}^n$, and we fix some volatility parameters $\sigma_i > 0$, $i = 1, \dots, d$. To guarantee that a solution to (20) exists, we impose the standard admissibility conditions that $\kappa\theta \in \mathbb{R}^d_+$ and the off-diagonal elements of κ be nonpositive, see e.g. Filipović (2009, Theorem 10.2).

To confirm that this model indeed exhibits USV, we consider the dynamics of the transformed state vector $\widehat{X}_t = SX_t = (Z_t, U_t)$. The transformed parameters are obtained from (9). Due to (21) and the form of S we get

$$\widehat{\kappa} = S\kappa S^{-1} = \begin{pmatrix} \kappa_{II} + A\kappa_{JI} & 0 \\ \kappa_{JI} & \kappa_{JJ} - \kappa_{JI}A \end{pmatrix}, \quad \widehat{\theta} = \begin{pmatrix} \theta_I + A\theta_J \\ \theta_J \end{pmatrix}.$$

The following result shows that this specification gives rise to at least n USV factors.

Proposition 4.2. *The dimension of the term structure kernel is at least n, $\dim \mathcal{U} \geq n$, with equality if $\widehat{\kappa}_{ZZ} = \kappa_{II} + A\kappa_{JI}$ satisfies Theorem 2.4*(iii)*. In this case, if $\sigma_i \neq \sigma_{m+i}$ for $i = 1, \dots, n$, then all the unspanned factors are in fact USV factors.*

For our empirical analysis, we employ the following parsimonious specification that falls within the class described above.

Definition 4.3. *The* LRSQ(m, n) specification *is obtained by letting in the above construction $\kappa_{JI} = 0$ and $\kappa_{JJ} = A^\top \kappa_{II} A$ (this is the upper left $n \times n$ block of κ_{II}).*

As an illustration, consider the *LRSQ(1,1)* specification, where we have one term structure factor and one unspanned factor. It shows in particular that a linear-rational term structure model may exhibit USV even in the two-factor case.[12]

Example 4.4. Under the *LRSQ(1,1)* specification the mean reversion matrix is given by

$$\kappa = \begin{pmatrix} \kappa_{11} & 0 \\ 0 & \kappa_{11} \end{pmatrix}.$$

The term structure factor and unspanned factor thus become $Z_t = X_{1t} + X_{2t}$ and $U_t = X_{2t}$, respectively. The transformed mean reversion matrix $\widehat{\kappa}$ coincides with κ,

$$\widehat{\kappa} = \begin{pmatrix} \kappa_{11} & 0 \\ 0 & \kappa_{11} \end{pmatrix},$$

and the corresponding volatility matrix is

$$\widehat{\sigma}(z, u) = \begin{pmatrix} \sigma_1 \sqrt{z_1 - u_1} & \sigma_2 \sqrt{u_1} \\ 0 & \sigma_2 \sqrt{u_1} \end{pmatrix}.$$

[12]This contradicts the statement of Collin-Dufresne and Goldstein (2002, Proposition 3), which thus is incorrect.

Thus the transformed diffusion matrix $\widehat{a}(z,u)$ satisfies $\widehat{a}_{11}(z,u) = \sigma_1^2 z + (\sigma_2^2 - \sigma_1^2)u$. This is non-constant in u as long as $\sigma_1 \neq \sigma_2$, as it should in view of Proposition 4.2. In particular, U_t is a USV factor.

Example 4.5. Consider now the *LRSQ(3,1)* specification. In this case we have

$$\kappa = \begin{pmatrix} \kappa_{11} & \kappa_{12} & \kappa_{13} & 0 \\ \kappa_{21} & \kappa_{22} & \kappa_{23} & \kappa_{21} \\ \kappa_{31} & \kappa_{32} & \kappa_{33} & \kappa_{31} \\ 0 & 0 & 0 & \kappa_{11} \end{pmatrix},$$

where it is straightforward to impose admissibility conditions on κ. The term structure factors and unspanned factors become

$$\begin{pmatrix} Z_{1t} \\ Z_{2t} \\ Z_{3t} \\ U_{1t} \end{pmatrix} = SX_t = \begin{pmatrix} X_{1t} + X_{4t} \\ X_{2t} \\ X_{3t} \\ X_{4t} \end{pmatrix},$$

and the transformed mean reversion matrix is given by

$$\widehat{\kappa} = \begin{pmatrix} \kappa_{11} & \kappa_{12} & \kappa_{13} & 0 \\ \kappa_{21} & \kappa_{22} & \kappa_{23} & 0 \\ \kappa_{31} & \kappa_{32} & \kappa_{33} & 0 \\ 0 & 0 & 0 & \kappa_{11} \end{pmatrix}.$$

The corresponding volatility matrix is

$$\widehat{\sigma}(z,u) = \begin{pmatrix} \sigma_1\sqrt{z_1 - u_1} & 0 & 0 & \sigma_4\sqrt{u_1} \\ 0 & \sigma_2\sqrt{z_2} & 0 & 0 \\ 0 & 0 & \sigma_3\sqrt{z_3} & 0 \\ 0 & 0 & 0 & \sigma_4\sqrt{u_1} \end{pmatrix}.$$

We now have $\widehat{a}_{11}(z,u) = \sigma_1^2 z_1 + (\sigma_4^2 - \sigma_1^2)u_1$, which again demonstrates the presence of USV, provided $\sigma_1 \neq \sigma_4$.

4.2 Swaption Pricing

Swaption pricing becomes particularly tractable in the LRSQ model. Since the factor process X_t is affine, the function $\widehat{q}(z)$ in Theorem 2.6 can be expressed using the exponential-affine transform formula that is available for such processes. Computing $\widehat{q}(z)$ then amounts to solving a system of ordinary differential equations, which takes the following well-known form, see Filipović (2009, Theorem 10.3).

Theorem 4.6 (Exponential-Affine Transform Formula)**.** *Suppose X_t follows the affine dynamics* (20)*. Then, for any $x \in \mathbb{R}_+^d$, $t \geq 0$, $u \in \mathbb{C}$, $v \in \mathbb{C}^d$ such that $\mathbb{E}_x[|\exp(v^\top X_t)|] < \infty$ we have*

$$\mathbb{E}_x\left[\mathrm{e}^{u+v^\top X_t}\right] = \mathrm{e}^{\Phi(t)+x^\top\Psi(t)},$$

where $\Phi : \mathbb{R}_+ \to \mathbb{C}$, $\Psi : \mathbb{R}_+ \to \mathbb{C}^d$ solve the system

$$\begin{aligned}
&\Phi'(\tau) = b^\top\Psi(\tau) \\
&\Psi_i'(\tau) = \beta_i^\top\Psi(\tau) + \frac{1}{2}\sigma_i^2\Psi_i(\tau)^2, \qquad i = 1, \ldots, d,
\end{aligned}$$

with initial condition $\Phi(0) = u$, $\Psi(0) = v$. The solution to this system is unique.

In order for Theorem 2.6 to be applicable, it is necessary that some exponential moments of $p_{\text{swap}}(X_t)$ be finite. We therefore remark that for X_t of the form (20), and for any $v \in \mathbb{R}^d$, $x \in \mathbb{R}_+^d$, $t \geq 0$, there is always some $\mu > 0$ (depending on v, x, t) such that $\mathbb{E}_x[\exp(\mu v^\top X_t)] < \infty$. While it may be difficult *a priori* to decide how small μ should be, the choice is easy in practice since numerical methods diverge if μ is too large, resulting in easily detectable outliers.

5 Data and Estimation

5.1 Swaps and Swaptions

We estimate the model on a panel data set consisting of swaps and swaptions. At each observation date, we observe rates on spot-starting swap contracts with maturities of one, two, three, five, seven, and ten years, respectively. We also observe prices on swaptions with three-month option maturities, the same six swap maturities, and strikes equal to the forward swap rates. Such at-the-money-forward (ATMF) swaptions are the most liquid. We convert swaption prices into normal implied volatilities using (17) with zero-coupon bonds bootstrapped from the swap curve. The data is from Bloomberg and consists of composite quotes computed from quotes that Bloomberg collects from major banks and inter-dealer brokers. The sample period consists of 827 weekly observations from January 29, 1997 to November 28, 2012.

Table 1 shows summary statistics of swap rates (Panel A) and swaption IVs (Panel B). The term structure of swap rates is upward-sloping, on average, while the standard deviation of swap rates decreases with maturity. Time series of the 1-year, 5-year, and 10-year swap rate are displayed in Panel A1 of Figure 1. The 1-year swap rate fluctuates between a minimum of 0.32 percent (on October 17, 2012) and a maximum of 7.51 percent (on May 17, 2000), while the longer-term swap rates exhibit less variation. A principal component analysis (PCA) of weekly changes in swap rates shows that the first three factors explain 90, 7, and 2 percent, respectively, of the variation.

The term structure of swaption IVs is hump-shaped, on average, increasing from 81 bps at the 1-year swap maturity to 107 bps at the 7-year swap maturity. The standard deviation of swaption IVs is also a hump-shaped function of maturity. Time series of swaption IVs at the 1-year, 5-year, and 10-year swap maturities are displayed in Panel B1 of Figure 1. The swaption IV at the 1-year swap maturity fluctuates between a minimum of 17 bps (on October 10, 2012) and a maximum of 224 bps (on October 15, 2008), while swaption IVs at longer swap maturities fluctuate in a tighter range. Swaptions also display a high degree of commonality, with the first three factors from a PCA of weekly changes in swaption IVs explaining 87, 7, and 2 percent, respectively, of the variation.

5.2 Volatility Dynamics at the Zero Lower Bound

A large literature has investigated the dynamics of interest rate volatility. A particular focus has been on the extent to which variation in volatility is related to variation in the term structure, with most papers finding that a significant component of volatility is only weakly related to term structure movements.[13] Here, we revisit this issue in the context of interest rates being close to the zero lower bound. We focus on the volatility of the 1-year swap rate, since this is the rate that is nearest to the zero lower bound during the sample period. Figure 2 shows the swaption IV at the 1-year swap maturity (in basis points) plotted against the 1-year swap rate. It strongly indicates that volatility becomes more level-dependent as the underlying interest rate approaches the zero lower bound.

To investigate the issue more formally, we regress weekly changes in the swaption IV at the 1-year swap maturity on weekly changes in the 1-year swap rate (including a constant); i.e.,

$$\Delta\sigma_{N,t} = \beta_0 + \beta_1 \Delta S_t + \epsilon_t. \tag{22}$$

[13]See Collin-Dufresne and Goldstein (2002) and subsequent papers by Heidari and Wu (2003), Andersen and Benzoni (2010), Li and Zhao (2006), Li and Zhao (2009), Trolle and Schwartz (2009), and Collin-Dufresne, Goldstein, and Jones (2009), among others. The issue is not without controversy, however, with Fan, Gupta, and Ritchken (2003), Jacobs and Karoui (2009), and Bikbov and Chernov (2009) providing a sceptical appraisal of the evidence.

Result are displayed in the upper part of Table 2, with Newey and West (1987) t-statistics using four lags in parentheses. The first column shows results using the entire sample period. β_1 is positive and statistically significant (t-statistic of 2.77); however, the R^2 is small at 0.05.[14] That is, unconditionally, a large fraction of the variation in volatility is unrelated to variation in rates, consistent with the existing unspanned stochastic volatility literature.

The second to sixth column shows results conditional on the 1-year swap rate being in the intervals 0-0.01, 0.01-0.02, 0.02-0.03, 0.03-0.04, and 0.04-0.08, respectively. A clear pattern emerges. At low interest rates, β_1 is positive and highly statistically significant (t-statistic of 10.1), and the R^2 is very high at 0.46. In other words, there is a strong and positive relation between volatility and rate changes, when rates are close to the zero lower bound. However, as interest rates increase, the relation between volatility and rate changes becomes progressively weaker. Both β_1 and R^2 decrease, and when the 1-year swap rate is above 0.03, the R^2 is essentially zero. Capturing the increasing level-dependence in volatility as the underlying interest rate approaches the zero lower bound poses a significant challenge for dynamic term structure models.

5.3 Model Specifications

For our empirical evaluation we use an $LRSQ(m,n)$ specification, see Definition 4.3, with m term structure factors and n USV factors. We always set $m=3$ and consider specifications with $n=1$ (volatility of Z_{1t} containing an unspanned component), $n=2$ (volatility of Z_{1t} and Z_{2t} containing unspanned components), and $n=3$ (volatility of all term structure factors containing unspanned components).

5.4 Maximum Likelihood Estimation

We estimate the model specifications using maximum likelihood in conjunction with Kalman filtering. For this purpose, we cast the model in state space form with a measurement equation describing the relation between the state variables and the

[14]Trolle and Schwartz (2013) also document a positive level-dependence in swaption IVs. An earlier literature has estimated generalized diffusion models for the short-term interest rate; see, e.g. Chan, Karolyi, Longstaff, and Sanders (1992), Ait-Sahalia (1996), Conley, Hansen, Luttmer, and Scheinkman (1997), and Stanton (1997). These papers generally find a relatively strong level-dependence in interest rate volatility. However, much of this level-dependence can be attributed to the 1979-1982 monetary policy experiment, which is not representative of the current monetary policy regime.

observable swap rates and swaption IVs, as well as a transition equation describing the discrete-time dynamics of the state variables.

Let X_t denote the vector of state variables and let Y_t denote the vector consisting of the term structure of swap rates and swaption IVs observed at time t. The measurement equation is given by

$$Y_t = h(X_t; \Theta) + u_t, \quad u_t \sim N(0, \Sigma), \tag{23}$$

where h is the pricing function, Θ is the vector of model parameters, and u_t is a vector of i.i.d. Gaussian pricing errors with covariance matrix Σ. To reduce the number of parameters in Σ, we assume that the pricing errors are cross-sectionally uncorrelated (that is, Σ is diagonal), and that one variance, σ^2_{rates}, applies to all pricing errors for swap rates, and that another variance, $\sigma^2_{swaption}$, applies to all pricing errors for swaption IVs.

While the transition density of X_t is unknown, its conditional mean and variance is known in closed form, because X_t follows an affine diffusion process. We approximate the transition density with a Gaussian density with identical first and second moments, in which case the transition equation is of the form

$$X_t = \Phi_0 + \Phi_X X_{t-1} + w_t, \quad w_t \sim N(0, Q_t), \tag{24}$$

where Q_t is an affine function of X_{t-1}.

As both swap rates and swaption IVs are non-linearly related to the state variables, we apply the nonlinear unscented Kalman filter.[15] The Kalman filter produces one-step-ahead forecasts for Y_t, $\hat{Y}_{t|t-1}$, and the corresponding error covariance matrices, $F_{t|t-1}$, from which we construct the log-likelihood function

$$\mathcal{L}(\Theta) = -\frac{1}{2} \sum_{t=1}^{T} \left(\# \log 2\pi + \log |F_{t|t-1}| + (Y_t - \hat{Y}_{t|t-1})^\top F^{-1}_{t|t-1} (Y_t - \hat{Y}_{t|t-1}) \right), \tag{25}$$

where T is the number of observation dates and # is the number of observations in Y_t. The (quasi) maximum likelihood estimator, $\hat{\Theta}$, is then

$$\hat{\Theta} = \arg\max_{\Theta} \mathcal{L}(\Theta). \tag{26}$$

When estimating the model, we found that the upper-triangular elements of $\widehat{\kappa}$ were always very close to zero. The same was true of κ_{31}. To obtain more parsimonious model specifications, we reestimate the models after setting to zero these

[15]Leippold and Wu (2007) appear to be the first to apply the unscented Kalman filter to the estimation of dynamic term structure models. Christoffersen, Jacobs, Karoui, and Mimouni (2009) show that it has very good finite-sample properties when estimating models using swap rates.

elements of the mean-reversion matrix. The likelihood functions were virtually unaffected by this, so we henceforth study these constrained model specifications.

6 Results

6.1 Maximum Likelihood Estimates

Table 3 displays parameter estimates and their asymptotic standard errors. All parameters are statistically significant. The structure of the mean-reversion matrix is such that the first term structure factor drives the mean-reversion level of the second factor, which in turn drives the mean-reversion level of the third factor. In all specifications, the first term structure factor exhibits the most persistence, followed by the third factor, and the second factor. Recall that if the i'th term structure factor exhibits USV, its instantaneous volatility is given by $\sqrt{\sigma_i^2 z_i + (\sigma_{i+3}^2 - \sigma_i^2) u_i}$ We always have $\sigma_{i+3} > \sigma_i$ implying that the volatility of the term structure factors are increasing in the USV factors.

Since all model specifications are stationary (all eigenvalues of κ are positive), α equals the infinite-maturity forward rate. This lies in a range from 0.0540 to 0.0636 across specifications, which appears reasonable. The table also reports the upper bound on possible short rates, which lies in a range from 0.2288 to 0.3195 across specifications. This is larger than the maximum short rate that has been observed historically in the U.S.[16] Furthermore, simulations show that the likelihood of observing short rates close to the upper bound is virtually zero.[17] As such, the upper bound on interest rates is not a restrictive feature of the model framework.

6.2 Factors

Figure 3 displays the estimated factors. The first, second, and third column shows the factors of the *LRSQ(3,1)*, *LRSQ(3,2)*, and *LRSQ(3,3)* model specification, respectively. The first, second, and third row shows (Z_1, U_1), (Z_2, U_2), and (Z_3, U_3), respectively. The factors are highly correlated across specifications, which is indicative of a stable factor structure. U_1 and U_2 are occasionally large relative to Z_1

[16]The historical maximum for the effective federal funds rate is 0.2236 and was reached on July 22, 1981, during the monetary policy experiment.

[17]In a simulation of 10,000 years of weekly data, the maximum short rate is 0.2071, 0.2644, and 0.2564, respectively, for *LRSQ(3,1)*, *LRSQ(3,2)*, and *LRSQ(3,3)*. This is below the upper bounds of 0.2288, 0.3042, and 0.3195, respectively.

and Z_2. In contrast, U_3 is almost always small relative to Z_3, providing a first indication that adding a third USV factor is not important for pricing.

To better understand the factor dynamics, Figure 4 plots the instantaneous volatility of each term structure factor against its level. Again, the first, second, and third column corresponds to the *LRSQ(3,1)*, *LRSQ(3,2)*, and *LRSQ(3,3)* specification, respectively, while the first, second, and third row corresponds to $Z_{1,t}$, $Z_{2,t}$, and $Z_{3,t}$, respectively. The grey areas mark the possible range of factor volatilities, which is given by $\sigma_i\sqrt{z_i}$ to $\sigma_{i+3}\sqrt{z_i}$ in case the i'th term structure factor exhibits USV. For $Z_{1,t}$ and $Z_{2,t}$ and whenever USV is allowed, there appears to be significant variation in factor volatilities that is unrelated to the factor level. For $Z_{3,t}$, allowing for USV appears to have less of an effect.

6.3 Specification Analysis

For each of the model specifications, we compute the fitted swap rates and swaption IVs based on the filtered state variables. We then compute weekly root mean squared pricing errors (RMSEs) separately for swap rates and swaption IVs, thereby constructing two time series of RMSEs. The first three rows in Table 4 reports the sample means of the RMSE time series for the three specifications. To investigate the performance of the model when interest rates are close to the zero lower bound, we also split the sample period into a zero interest rate policy (ZIRP) sample period and a pre-ZIRP sample period. The beginning of ZIRP is taken to be December 16, 2008, when the Federal Reserve reduced the federal funds rate from one percent to a target range of 0 to 1/4 percent. The next two rows report the mean difference in RMSEs between two model specifications along with the associated t-statistics corrected for heteroscedasticity and serial correlation in parenthesis.

Even the most parsimonious *LRSQ(3,1)* specification has a reasonable fit to the data. For instance, for the full sample period, the mean RMSEs for swap rates and swaption IVs are 5.73 bps and 7.31 bps, respectively. Adding one more USV factor decreases the mean RMSEs by 1.86 bps and 1.42 bps, respectively, which is both economically important and strongly statistically significant. Adding an additional USV factor decreases the mean RMSEs by a further 0.13 bps and 0.47 bps, respectively, which is only a modest improvement economically, if still statistically significant.

Comparing across sub-samples, for all the specifications the fit is better in the pre-ZIRP period than in the ZIRP period, particularly for swap rates. Nevertheless, even during the ZIRP period, the model performs well. For instance, in case of the *LRSQ(3,2)* specification, the mean RMSEs for swap rates and swaption IVs are 6.08

bps and 6.19 bps, respectively. The performance of the *LRSQ(3,2)* specification over time is illustrated in Figure 1 which shows the fitted time series of selected swap rates (Panel A2) and swaption IVs (Panel B2) as well as time series of the RMSEs for swap rates (Panel A3) and swaption IVs (Panel B3).[18]

As always, there is a tradeoff between parsimony and (in-sample) pricing performance. *LRSQ(3,1)* is valued for its parsimony and *LRSQ(3,2)* is valued for its better pricing performance. *LRSQ(3,3)* appears overparameterized given the data at hand and will not be considered in the remainder of the paper.

6.4 Capturing Volatility Dynamics at the Zero Lower Bound

We now investigate if the model can capture the volatility dynamics at the zero lower bound discussed in Section 5.2. We focus on the population properties of the model as this is much more demanding than using the fitted data. To infer the population properties, we redo the analysis in Section 5.2 using simulated time series from the *LRSQ(3,1)* and *LRSQ(3,2)* specifications consisting of 520,000 weekly observations. The middle and lower part of Table 2 shows the population regression coefficients and R^2s for the two specifications. In terms of R^2s, the specifications closely match the pattern in the data with an R^2 of approximately 0.50 when rates are close to zero, and a fast decay in the R^2 as rates increase. In terms of the regression coefficient, the specifications do not quite match the degree of level-dependence, but do match the decay in the regression coefficient as rates increase.

7 Conclusion

We introduce the class of linear rational term structure models, where the state price density is modeled such that bond prices become linear-rational functions of the current state. This class is highly tractable with several distinct advantages: i) ensures non-negative interest rates, ii) easily accommodates unspanned factors affecting volatility and risk premia, and iii) has analytical solutions to swaptions. A parsimonious specification of the model with three term structure factors and one, or possibly two, unspanned factors has a very good fit to both interest rate swaps and swaptions since 1997. In particular, the model captures well the dynamics of the term structure and volatility during the recent period of near-zero interest rates.

[18]As a further check, Table 5 reports RMSEs of individual swap rates and swaption IVs. The quality of the fit appears to be relatively uniform across swap maturities.

A Proofs

A.1 Proofs for Section 2

The following lemma directly implies formula (4). It is also used in the proof of Lemma A.10.

Lemma A.1. *Assume that X is of the form* (2) *with integrable starting point X_0. Then for any bounded stopping time ρ and any deterministic $\tau \geq 0$, the random variable $X_{\rho+\tau}$ is integrable, and we have*

$$\mathbb{E}[X_{\rho+\tau} \mid \mathcal{F}_\rho] = \theta + \mathrm{e}^{-\kappa\tau}(X_\rho - \theta).$$

Proof. We first prove the result for $\rho = 0$. An application of Itô's formula shows that the process

$$Y_t = \theta + \mathrm{e}^{-\kappa(\tau-t)}(X_t - \theta)$$

satisfies $\mathrm{d}Y_t = \mathrm{e}^{-\kappa(\tau-t)}\mathrm{d}M_t$, and hence is a local martingale. It is in fact a true martingale. Indeed, integration by parts yields

$$Y_t = Y_0 + \mathrm{e}^{-\kappa(T-t)}M_t - \int_0^t M_s\kappa\mathrm{e}^{-\kappa(T-s)}\mathrm{d}s,$$

from which the integrability of X_0 and L^1-boundedness of the martingale M imply that Y is bounded in L^1. Fubini's theorem then yields, for any $0 \leq t \leq u$,

$$\begin{aligned}
\mathbb{E}\left[Y_u \mid \mathcal{F}_t\right] &= Y_0 + \mathrm{e}^{-\kappa(T-u)}M_t - \int_0^u M_{s\wedge t}\kappa\mathrm{e}^{-\kappa(T-s)}\mathrm{d}s \\
&= Y_t + M_t\left[\mathrm{e}^{-\kappa(T-u)} - \mathrm{e}^{-\kappa(T-t)} - \int_t^u \kappa\mathrm{e}^{-\kappa(T-s)}\mathrm{d}s\right] \\
&= Y_t,
\end{aligned}$$

showing that Y is a true martingale. Since $Y_\tau = X_\tau$ it follows that

$$\mathbb{E}[X_\tau \mid \mathcal{F}_0] = Y_0 = \theta + \mathrm{e}^{-\kappa\tau}(X_0 - \theta),$$

as claimed. If ρ is a bounded stopping time, then the L^1-boundedness of Y, and hence of X, implies that X_ρ is integrable. The result then follows by applying the $\rho = 0$ case to the process $(X_{\rho+s})_{s\geq 0}$ and filtration $(\mathcal{F}_{\rho+s})_{s\geq 0}$. □

Proof of Proposition 2.2. Observe, by taking the orthogonal complement in (8), that we must prove $\mathcal{U}^\perp = \mathrm{span}\big\{(\kappa^\top)^p\psi : p = 0,\dots,d-1\big\}$. By the Cayley-Hamilton theorem (see Horn and Johnson (1990, Theorem 2.4.2)) we may equivalently let p range over all nonnegative integers. In other words, we need to prove

$$\mathrm{span}\,\{\nabla F(\tau,x) : \tau \ge 0, x \in E\} = \mathrm{span}\,\big\{(\kappa^\top)^p\psi : p \ge 0\big\}\,. \tag{27}$$

Denote the left side by $\mathcal{S}$. A direct computation shows that the gradient of F is given by

$$\nabla F(\tau,x) = \frac{\mathrm{e}^{-\alpha\tau}}{\phi+\psi^\top x}\left[\mathrm{e}^{-\kappa^\top\tau}\psi - \mathrm{e}^{\alpha\tau}F(\tau,x)\psi\right], \tag{28}$$

whence $\mathcal{S} = \mathrm{span}\{\mathrm{e}^{-\kappa^\top\tau}\psi - \mathrm{e}^{\alpha\tau}F(\tau,x)\psi : \tau \ge 0, x \in E\}$. By the non-triviality assumption there are $x,y \in E$ and $\tau \ge 0$ such that $F(\tau,x) \ne F(\tau,y)$. It follows that $\mathrm{e}^{\alpha\tau}(F(\tau,x) - F(\tau,y))\psi$, and hence ψ itself, lies in $\mathcal{S}$. We deduce that $\mathcal{S} = \mathrm{span}\{\mathrm{e}^{-\kappa^\top\tau}\psi : \tau \ge 0\}$, which coincides with the right side of (27), as desired. □

Proof of Corollary 2.3. Write $\Lambda = \mathrm{Diag}(\lambda_1,\dots,\lambda_d)$ and consider the matrix

$$A = \begin{bmatrix}\psi & \kappa^\top\psi & \cdots & (\kappa^\top)^{d-1}\psi\end{bmatrix}.$$

Writing $\widehat{\psi} = S^{-\top}\psi$, the determinant of A is given by

$$\begin{aligned}
\det A &= \det\left(S^\top\right)\det\begin{pmatrix}\widehat{\psi} & \Lambda\widehat{\psi} & \cdots & \Lambda^{d-1}\widehat{\psi}\end{pmatrix}\\
&= \det\left(S^\top\right)\,\widehat{\psi}_1\cdots\widehat{\psi}_d\,\det\begin{pmatrix}1 & \lambda_1 & \cdots & \lambda_1^{d-1}\\ \vdots & \vdots & & \vdots\\ 1 & \lambda_d & \cdots & \lambda_d^{d-1}\end{pmatrix}\\
&= \det\left(S^\top\right)\,\widehat{\psi}_1\cdots\widehat{\psi}_d \prod_{1\le i<j\le d}(\lambda_j-\lambda_i),
\end{aligned}$$

where the last equality uses the formula for the determinant of the Vandermonde matrix. Proposition 2.2 now shows that the term structure kernel is trivial precisely when all eigenvalues of κ are distinct and all components of $\widehat{\psi}$ are nonzero, as was to be shown. □

Proof of Theorem 2.4. The case $n = 0$ is immediate, so we consider the case $n \ge 1$. We write $\widehat{\mathcal{U}} = S(\mathcal{U})$ and assume (11) holds. That is, we have

$$\widehat{\mathcal{U}} = \{0\}\times\mathbb{R}^n \subset \mathbb{R}^m\times\mathbb{R}^n. \tag{29}$$

On the other hand, Proposition 2.2 yields

$$\widehat{\mathcal{U}} = \left\{\widehat{\xi} \in \mathbb{R}^d : \widehat{\psi}^\top \widehat{\kappa}^p \widehat{\xi} = 0 \text{ for } p = 0, 1, \dots, d-1\right\}. \tag{30}$$

Partition $\widehat{\psi}$ and $\widehat{\kappa}$ to conform with the product structure of $\mathbb{R}^m \times \mathbb{R}^n$:

$$\widehat{\psi} = \begin{pmatrix} \widehat{\psi}_Z \\ \widehat{\psi}_U \end{pmatrix} \in \mathbb{R}^{m+n}, \qquad \widehat{\kappa} = \begin{pmatrix} \widehat{\kappa}_{ZZ} & \widehat{\kappa}_{ZU} \\ \widehat{\kappa}_{UZ} & \widehat{\kappa}_{UU} \end{pmatrix} \in \mathbb{R}^{(m+n)\times(m+n)}.$$

It then follows from (29) and (30) that we have $\widehat{\psi}_U^\top \widehat{\xi}_U = 0$ for any $\widehat{\xi}_U \in \mathbb{R}^n$. Hence $\widehat{\psi}_U = 0$, which proves (i). As a consequence, we have

$$\widehat{\psi}^\top \widehat{\kappa}^p = \begin{pmatrix} \widehat{\psi}_Z^\top \widehat{\kappa}_{ZZ}^p & \qquad \widehat{\psi}_Z^\top \widehat{\kappa}_{ZZ}^{p-1} \widehat{\kappa}_{ZU} \end{pmatrix} \tag{31}$$

for $p = 1$. Suppose we know (31) holds for some $p \geq 1$. Then (29) and (30) imply, for any $\widehat{\xi}_U \in \mathbb{R}^n$,

$$0 = \widehat{\psi}^\top \widehat{\kappa}^p \begin{pmatrix} 0 \\ \widehat{\xi}_U \end{pmatrix} = \widehat{\psi}_Z^\top \widehat{\kappa}_{ZZ}^{p-1} \widehat{\kappa}_{ZU} \widehat{\xi}_U,$$

and hence

$$\widehat{\psi}_Z^\top \widehat{\kappa}_{ZZ}^{p-1} \widehat{\kappa}_{ZU} = 0. \tag{32}$$

Multiplying both sides of (31) by $\widehat{\kappa}$ from the right then shows that (31) holds also for $p+1$. It follows by induction that (31) and (32) hold for all $p \geq 1$.

Now, pick any $\widehat{\xi}_Z$ such that $\widehat{\psi}_Z^\top \widehat{\kappa}_{ZZ}^p \widehat{\xi}_Z = 0$ for all $p \geq 0$. Then the vector $\widehat{\xi} = (\widehat{\xi}_Z, 0) \in \mathbb{R}^m \times \mathbb{R}^n$ satisfies $\widehat{\psi}^\top \widehat{\kappa}^p \widehat{\xi} = 0$ for all $p \geq 0$. Hence $\widehat{\xi} \in \widehat{\mathcal{U}}$ by (30), and then $\widehat{\xi}_Z = 0$ by (29). This proves (iii). Finally, since (32) holds for all $p \geq 1$, the range of $\widehat{\kappa}_{ZU}$ lies in the kernel of $\widehat{\psi}_Z^\top \widehat{\kappa}_{ZZ}^{p-1}$ for all $p \geq 1$. But by (iii) this implies that the range of $\widehat{\kappa}_{ZU}$ consists of the zero vector, which is to say that $\widehat{\kappa}_{ZU} = 0$. This proves (ii).

We now prove the converse part of the theorem, and assume that $\widehat{\kappa}$ and $\widehat{\psi}$ given in (9) and (10) satisfy (i)–(iii). We first show that (i)–(ii) imply $S(\mathcal{U}) \supset \{0\} \times \mathbb{R}^n$. Letting $\{e_1, \dots, e_d\}$ denote the canonical basis of $\mathbb{R}^d$, we have for $p = 0, \dots, d-1$,

$$\psi^\top \kappa^p S^{-1} e_i = \widehat{\psi}^\top \widehat{\kappa}^p e_i.$$

Note that $\widehat{\kappa}^p$ has the same block triangular structure as $\widehat{\kappa}$ for all $p \geq 1$. Thus for $i = m+1, \dots, d$, the right side above is zero for all $p \geq 0$, so $S^{-1} e_i \in \mathcal{U}$. We deduce that $S(\mathcal{U}) \supset \{0\} \times \mathbb{R}^n$ holds, as claimed. Suppose now in addition that (iii) holds, and consider a vector ξ in the span of $S^{-1}e_1, \ldots, S^{-1}e_m$. Then $S\xi = (\widehat{\xi}_Z, 0) \in \mathbb{R}^m \times \mathbb{R}^n$. By (i) and (ii) we have

$$\psi^\top \kappa^p \xi = \widehat{\psi}^\top \widehat{\kappa}^p S\xi = \widehat{\psi}_Z^\top \widehat{\kappa}_{ZZ}^p \widehat{\xi}_Z.$$

If $\xi \in \mathcal{U}$, the left side is zero for all $p = 0, \ldots, d-1$. Hence so is the right side, which by (iii) implies $\xi = 0$. We deduce that $\mathcal{U} = \operatorname{span}\{S^{-1}e_{m+1}, \ldots, S^{-1}e_d\}$ and hence $S(\mathcal{U}) = \{0\} \times \mathbb{R}^n$, as claimed. □

The proof of the converse direction of Theorem 2.4 directly leads to the following result, which is a useful sufficient condition that guarantees the existence of at least a given number of unspanned factors.

Lemma A.2. *Let $m, n \geq 0$ be integers with $m + n = d$. If the transformed model parameters* (9)–(10) *satisfy* (i)–(ii) *in Theorem 2.4, we have $S(\mathcal{U}) \supset \{0\} \times \mathbb{R}^n$. In this case, we have* $\dim \mathcal{U} \geq n$.

Proof of Proposition 2.5. We first assume $\mathcal{U} = \{0\}$, so that $m = d$. Expanding $F(\tau, x)$ as a power series in τ shows that for any fixed $x, y \in E$ we have $F(\tau, x) = F(\tau, y)$ for all $\tau \geq 0$ if and only if

$$\frac{\psi^\top \kappa^p (x - \theta)}{\phi + \psi^\top x} = \frac{\psi^\top \kappa^p (y - \theta)}{\phi + \psi^\top y}, \qquad p \geq 1. \tag{33}$$

To prove sufficiency, assume κ is invertible and $\phi + \psi^\top \theta \neq 0$. Then pick x, y such that (33) is satisfied; we must prove that $x = y$. Lemma A.5 implies that $\kappa^\top \psi, \ldots, (\kappa^\top)^d \psi$ span $\mathbb{R}^d$, so we may find coefficients $a_1, \ldots, a_d$ so that $\psi = \sum_{p=1}^d a_p (\kappa^\top)^p \psi$. Multiplying both sides of (33) by a_p and summing over $p = 1, \ldots, d$ yields

$$\frac{\psi^\top (x - \theta)}{\phi + \psi^\top x} = \frac{\psi^\top (y - \theta)}{\phi + \psi^\top y}, \qquad p \geq 1,$$

or, equivalently, $\psi^\top (x - y)(\phi + \psi^\top \theta) = 0$. Since $\phi + \psi^\top \theta \neq 0$ we then deduce from (33) that $\psi^\top \kappa^p (x - y) = 0$ for all $p \geq 0$, which by the aforementioned spanning property of $\psi^\top \kappa^p$, $p \geq 0$, implies $x = y$ as required. This finishes the proof of the first assertion.

To prove necessity, we argue by contradiction and suppose it is not true that κ is invertible and $\phi + \psi^\top \theta \neq 0$. There are two cases. First, assume κ is not invertible. We claim that there is an element $\eta \in \ker \kappa$ such that $\theta + s\eta$ lies in the set $\{x \in \mathbb{R}^d : \phi + \psi^\top x \neq 0\}$ for all large s. Indeed, if this were not the case we would have $\ker \kappa \subset \ker \psi^\top$, which would contradict $\mathcal{U} = \{0\}$. So such an η exists. Now simply take $x = \theta + s_1\eta$, $y = \theta + s_2\eta$ for large enough $s_1 \neq s_2$—clearly (33) holds for this choice, proving that injectivity fails.

The second case is where κ is invertible, but $\phi + \psi^\top\theta = 0$. In particular $\phi + \psi^\top x = \psi^\top(x-\theta)$. Together with the fact that $\kappa^\top\psi, \ldots, (\kappa^\top)^d\psi$ span $\mathbb{R}^d$ (see Lemma A.5), this shows that (33) is equivalent to

$$\frac{x-\theta}{\psi^\top(x-\theta)} = \frac{y-\theta}{\psi^\top(y-\theta)}.$$

We deduce that $F(\tau, x)$ is constant along rays of the form $\theta + s(x-\theta)$, where x is any point in the state space, and thus that injectivity fails.

The proof of the proposition is now complete for the case $\mathcal{U} = \{0\}$. The general case where $\mathcal{U}$ it not necessarily trivial follows by applying the $\mathcal{U} = \{0\}$ case to the model with factor process Z and state price density $\zeta_t = \mathrm{e}^{-\alpha t}(\widehat{\phi} + \widehat{\psi}_Z^\top Z_t)$. Indeed, this model has a trivial term structure kernel in view of Theorem 2.4 (iii). $\square$

Proof of Theorem 2.6. The proof uses the following identity from Fourier analysis, valid for any $\mu > 0$ and $s \in \mathbb{R}$ (see for instance Bateman and Erdélyi (1954, Formula 3.2(3))):

$$s^+ = \frac{1}{2\pi}\int_{\mathbb{R}} \mathrm{e}^{(\mu+\mathrm{i}\lambda)s}\frac{1}{(\mu+\mathrm{i}\lambda)^2}\mathrm{d}\lambda. \tag{34}$$

Let $q(\mathrm{d}s)$ denote the conditional distribution of the random variable $p_{\mathrm{swap}}(X_{T_0})$, given $\mathcal{F}_t$, so that

$$\widehat{q}(z) = \int_{\mathbb{R}} e^{zs}q(\mathrm{d}s)$$

for every $z \in \mathbb{C}$ such that the right side is well-defined and finite. Pick $\mu > 0$ such that $\int_{\mathbb{R}} \mathrm{e}^{\mu s}q(\mathrm{d}s) < \infty$. Then,

$$\begin{aligned}\int_{\mathbb{R}^2}\left|\mathrm{e}^{(\mu+\mathrm{i}\lambda)s}\frac{1}{(\mu+\mathrm{i}\lambda)^2}\right|\mathrm{d}\lambda\otimes q(\mathrm{d}s) &= \int_{\mathbb{R}^2}\frac{\mathrm{e}^{\mu s}}{\mu^2+\lambda^2}\mathrm{d}\lambda\otimes q(\mathrm{d}s)\\ &= \int_{\mathbb{R}}\mathrm{e}^{\mu s}q(\mathrm{d}s)\int_{\mathbb{R}}\frac{1}{\mu^2+\lambda^2}\mathrm{d}\lambda < \infty,\end{aligned}$$

where the second equality follows from Tonelli's theorem. This justifies applying Fubini's theorem in the following calculation, which uses the identity (34) on the second line:

$$\begin{aligned}
\mathbb{E}\left[p_{\mathrm{swap}}(X_{T_0})^+ \mid \mathcal{F}_t\right] &= \int_{\mathbb{R}} s^+ q(\mathrm{d}s) \\
&= \int_{\mathbb{R}} \left(\frac{1}{2\pi} \int_{\mathbb{R}} \mathrm{e}^{(\mu+\mathrm{i}\lambda)s} \frac{1}{(\mu+\mathrm{i}\lambda)^2} \mathrm{d}\lambda \right) q(\mathrm{d}s) \\
&= \frac{1}{2\pi} \int_{\mathbb{R}} \frac{\widehat{q}(\mu+\mathrm{i}\lambda)}{(\mu+\mathrm{i}\lambda)^2} \mathrm{d}\lambda \\
&= \frac{1}{\pi} \int_0^\infty \mathrm{Re}\left[\frac{\widehat{q}(\mu+\mathrm{i}\lambda)}{(\mu+\mathrm{i}\lambda)^2} \right] \mathrm{d}\lambda.
\end{aligned}$$

Here the last equality uses that the left, and hence right, side is real, together with the observation that the real part of $(\mu+\mathrm{i}\lambda)^{-2}\widehat{q}(\mu+\mathrm{i}\lambda)$ is an even function of λ (this follows from a brief calculation.) The resulting expression for the conditional expectation, together with (16), gives the result. $\square$

A.2 Proofs for Section 3

The proof of Theorem 3.2 requires some notation and two lemmas. For a multiindex $\mathbf{k} = (k_1, \ldots, k_d) \in \mathbb{N}_0^d$ we write $|\mathbf{k}| = k_1 + \cdots + k_d$, $x^{\mathbf{k}} = x_1^{k_1} \cdots x_d^{k_d}$, and $\partial^{\mathbf{k}} = \partial^{|\mathbf{k}|}/\partial x_1^{k_1} \cdots \partial x_d^{k_d}$.

Lemma A.3. *Assume $\sigma(X_t)$ is invertible $\mathrm{d}t \otimes \mathrm{d}\mathbb{P}$-almost surely. For any function $f \in C^{1,\infty}(\mathbb{R}_+ \times E)$ we have*

$$\{f(t, X_t) = 0\} \subset \bigcap_{\mathbf{k} \in \mathbb{N}_0^d} \left\{\partial^{\mathbf{k}} f(t, X_t) = 0\right\},$$

up to a $\mathrm{d}t \otimes \mathrm{d}\mathbb{P}$-nullset.

Proof. Let $n \geq 0$ and suppose we have, up to a nullset,

$$\{f(t, X_t) = 0\} \subset \{\partial^{\mathbf{k}} f(t, X_t) = 0\} \tag{35}$$

for all $\mathbf{k} \in \mathbb{N}_0^d$ with $|\mathbf{k}| = n$. Fix such a $\mathbf{k}$ and set $g(t,x) = \partial^{\mathbf{k}} f(t,x)$. The occupation time formula, see Revuz and Yor (1999, Corollary VI.1.6), yields

$$\mathbf{1}_{\{g(t,X_t)=0\}} \nabla g(t, X_t)^\top a(X_t) \nabla g(t, X_t) = 0 \quad \mathrm{d}t \otimes \mathrm{d}\mathbb{P}\text{-a.s.}$$

This implies $\{g(t, X_t) = 0\} \subset \{\nabla g(t, X_t)^\top a(X_t) \nabla g(t, X_t) = 0\}$ up to a nullset. Since $a(X_t)$ is invertible $\mathrm{d}t \otimes \mathrm{d}\mathbb{P}$-almost surely we get, again up to a nullset,

$$\{g(t, X_t) = 0\} \subset \{\nabla g(t, X_t) = 0\}.$$

We deduce that (35) holds for all $\mathbf{k} \in \mathbb{N}_0^d$ with $|\mathbf{k}| = n+1$. Since (35) is trivially true for $n = 0$, the result follows by induction. $\square$

Lemma A.4. *For any $x \in E$ we have the identity*

$$\operatorname{span}\{\nabla F(\tau, x) : \tau \geq 0\} = \operatorname{span}\left\{(\kappa^\top)^p \psi - \frac{\psi^\top \kappa^p (x-\theta)}{\psi + \psi^\top x}\psi : p = 1, \dots, d\right\}.$$

Proof. This follows directly from the power series expansion in τ of the expression (28) for $\nabla F(\tau, x)$. $\square$

Lemma A.5. *Assume $\phi + \psi^\top \theta \neq 0$ and consider any $x \in E$. The following conditions are equivalent:*

(i) *The vectors $(\kappa^\top)^p \psi$, $p = 1, \dots, d$, are linearly independent;*

(ii) *κ is invertible and $\mathcal{U} = \{0\}$;*

(iii) $\operatorname{span}\{\nabla F(\tau, x) : \tau \geq 0\} = \mathbb{R}^d$.

Moreover, conditions (i) *and* (ii) *are equivalent even if $\psi + \psi^\top \theta = 0$.*

Proof. The equivalence of (i) and (ii) is deduced from the identity

$$\det\left(\kappa^\top \psi \ \cdots \ (\kappa^\top)^d \psi\right) = \det\left(\kappa^\top\right) \det\left(\psi \ \cdots \ (\kappa^\top)^{d-1}\psi\right),$$

together with Proposition 2.2, which in particular states that the second determinant on the right side is zero if and only if $\mathcal{U} = \{0\}$.

We next prove that (i) implies (iii). Since $(\kappa^\top)^p \psi$, $p = 1, \dots, d$, span $\mathbb{R}^d$, we can find $a_1, \dots, a_d$ such that $\psi = \sum_{p=1}^d a_p (\kappa^\top)^p \psi$. Together with the representation in Lemma A.4 we deduce that the vector

$$\sum_{p=1}^d a_p \left[(\kappa^\top)^p \psi - \frac{\psi^\top \kappa^p (x-\theta)}{\phi + \psi^\top x}\psi\right] = \psi - \frac{\psi^\top (x-\theta)}{\phi + \psi^\top x}\psi = \frac{\phi + \psi^\top \theta}{\phi + \psi^\top x}\psi$$

lies in $\operatorname{span}\{\nabla F(\tau, x) : \tau \geq 0\}$. So does ψ since $\phi + \psi^\top \theta \neq 0$, and it follows that we have $\operatorname{span}\{\nabla F(\tau, x) : \tau \geq 0\} \supset \operatorname{span}\{(\kappa^\top)^p \psi : p = 1, \dots, d\} = \mathbb{R}^d$. This proves (iii).

It remains to prove that (iii) implies (i). To this end, we use the hypothesis together with the representation in Lemma A.4 to find $a_1, \dots, a_p$ such that

$$\psi = \sum_{p=1}^d a_p \left[(\kappa^\top)^p \psi - \frac{\psi^\top \kappa^p (x-\theta)}{\phi + \psi^\top x}\psi\right].$$

Re-arranging this expression and writing $u = \sum_{p=1}^{d} a_p (\kappa^\top)^p \psi$ yields

$$\psi \left(1 + \frac{u^\top (x - \theta)}{\phi + \psi^\top x} \right) = u.$$

It follows that ψ lies in $\operatorname{span}\{(\kappa^\top)^p \psi : p = 1, \dots, d\}$, which consequently is equal to $\operatorname{span}\{\nabla F(\tau, x) : \tau \geq 0\} = \mathbb{R}^d$. This proves that (i) holds. □

Proof of Theorem 3.2. By a standard argument involving the martingale representation theorem, bond market completeness holds if and only if for any given $T \geq 0$ there exist maturities $T_i \geq T$, $i = 1, \dots, m$, such that $\mathrm{d}t \otimes \mathrm{d}\mathbb{P}$-almost surely the volatilility vectors $\nu(T_i - t, X_t)$ span $\mathbb{R}^d$. Since $\sigma(X_t)$ is invertible $\mathrm{d}t \otimes \mathrm{d}\mathbb{P}$-almost surely this happens if and only if $\mathrm{d}t \otimes \mathrm{d}\mathbb{P}$-almost surely the vectors $\nabla F(T_i - t, X_t)$ span $\mathbb{R}^d$. This shows, in particular, the implication "(i) $\Longrightarrow$ (ii)".

To prove "(ii) $\Longrightarrow$ (iii)", suppose (iii) fails. Lemma A.4 then implies that for each $x \in E$, $\operatorname{span}\{\nabla F(\tau, x) : \tau \geq 0\}$ is not all of $\mathbb{R}^d$. Thus (ii) fails.

It remains to prove "(iii) $\Longrightarrow$ (i)", so we assume that κ is invertible and $\mathcal{U} = \{0\}$. Choose maturities $T_1, \dots, T_d$ greater than or equal to T so that the function

$$g(t, x) = \det \Big(\nabla F(T_1 - t, x) \ \cdots \ \nabla F(T_d - t, x) \Big)$$

is not identically zero. This is possible by Lemma A.4. A calculation yields

$$\nabla F(\tau, x) = \frac{\mathrm{e}^{-\alpha \tau}}{(\psi + \psi^\top x)^2} \eta(\tau, x),$$

where $\eta(\tau, x)$ is a vector of first degree polynomials in x whose coefficients are analytic functions of τ. Defining

$$f(t, x) = \det \Big(\eta(T_1 - t, x) \ \cdots \ \eta(T_d - t, x) \Big),$$

we have $g(t, x) = 0$ if and only if $f(t, x) = 0$. Hence $f(t, x)$ is not identically zero. Our goal is to strengthen this to the statement that $\{f(t, X_t) = 0\}$ is a $\mathrm{d}t \otimes \mathrm{d}\mathbb{P}$-nullset. Indeed, then $\{g(t, X_t) = 0\}$ is also a $\mathrm{d}t \otimes \mathrm{d}\mathbb{P}$-nullset, implying that completeness holds. To prove that $\{f(t, X_t) = 0\}$ is a $\mathrm{d}t \otimes \mathrm{d}\mathbb{P}$-nullset, note that $f(t, x)$ is of the form

$$f(t, x) = \sum_{|\mathbf{k}| \leq n} c_{\mathbf{k}}(t) x^{\mathbf{k}},$$

where $n = \max_{0 \leq t \leq T} \deg f(t, \cdot) < \infty$. Lemma A.3 implies

$$\{f(t, X_t) = 0\} \subset \bigcap_{|\mathbf{k}| = n} \{c_{\mathbf{k}}(t) = 0\},$$

up to a nullset. Assume for contradiction that the left side is of positive $\mathrm{d}t \otimes \mathrm{d}\mathbb{P}$-measure. Then so is the right side, whence all the $c_{\mathbf{k}}$ (which are deterministic) vanish on a t-set of positive Lebesgue measure. The zero set of each $c_{\mathbf{k}}$ must thus contain an accumulation point, so that, by analyticity, they are all identically zero, see Rudin (1987, Theorem 10.18). Hence we have either $\max_{0\le t\le T} \deg f(t,\cdot) \le n-1$ (if $n \ge 1$) or $f(t,x) \equiv 0$ (if $n=0$). In both cases we obtain a contradiction, which shows that $\{f(t,X_t)=0\}$ is a $\mathrm{d}t \otimes \mathrm{d}\mathbb{P}$-nullset, as required.

Finally, the equivalence of (iii) and (iv) follows easily from Proposition 2.5. The theorem is proved. $\square$

Proposition A.7 below gives a description of the variance-covariance kernel $\mathcal{W}$. Its proof requires the following lemma.

Lemma A.6. *Assume $\phi + \psi^\top\theta \neq 0$, and consider any $x \in E$. The following conditions are equivalent.*

- *(i)* $\mathcal{U}^\perp = \operatorname{span}\{\nabla F(\tau,x) : \tau \ge 0\}$,
- *(ii)* $\psi \in \operatorname{span}\big\{(\kappa^\top)^p\psi : p = 1,\dots,d\big\}$.

Proof. The proof is a straightforward adaptation of the proof of the equivalence "(i)$\Longleftrightarrow$(iii)" in Lemma A.5, and therefore omitted. $\square$

Proposition A.7. *The variance-covariance kernel satisfies*

$$\mathcal{U} \cap \mathcal{W} \quad \subset \quad \mathcal{U} \cap \bigcap_{\eta\in\mathcal{U}^\perp} \ker \eta^\top a(\cdot)\eta$$

with equality if $\phi + \psi^\top\theta \neq 0$ and $\psi \in \operatorname{span}\big\{(\kappa^\top)^p\psi : p = 1,\dots,d\big\}$.

Proof of Proposition A.7. Consider an arbitrary vector $\xi \in \mathcal{U}\cap\mathcal{W}$. Since $F(\tau, x+s\xi)$ is constant in s, we have that $\nabla G(\tau_1,\tau_2,x)^\top\xi = 0$ if and only if $\nabla \overline{G}(\tau_1,\tau_2,x)^\top\xi = 0$, where we define

$$\overline{G}(\tau_1,\tau_2,x) = \nabla F(\tau_1,x)^\top a(x)\nabla F(\tau_2,x).$$

Now, for any $x \in E$ and $\tau \ge 0$, the chain rule yields

$$\frac{\mathrm{d}}{\mathrm{d}s}\nabla F(\tau, x+s\xi)\Big|_{s=0} = \nabla F(\tau,x)^\top\xi = 0.$$

Hence, by the product rule,

$$\frac{\mathrm{d}}{\mathrm{d}s}\overline{G}(\tau_1,\tau_2,x+s\xi)\Big|_{s=0} = \frac{\mathrm{d}}{\mathrm{d}s}\nabla F(\tau_1,x)^\top a(x+s\xi)\nabla F(\tau_2,x)\Big|_{s=0}.$$

The right side is zero for all $x \in E$, $\tau_1, \tau_2 \geq 0$ if and only if, for every $x \in E$,

$$\frac{\mathrm{d}}{\mathrm{d}s}\eta_1^\top a(x+s\xi)\eta_2\Big|_{s=0} = 0$$

holds for all $\eta_1, \eta_2 \in \operatorname{span}\{\nabla F(\tau, x) : \tau \geq 0\}$. But we always have

$$\operatorname{span}\{\nabla F(\tau, x) : \tau \geq 0\} \subset \mathcal{U}^\perp,$$

with equality if $\phi + \psi^\top\theta \neq 0$ and $\psi \in \operatorname{span}\{(\kappa^\top)^p\psi : p = 1, \ldots, d\}$, due to Lemma A.6. This proves that the inclusion

$$\mathcal{W} \subset \left\{\xi \in \mathcal{U} : \frac{\mathrm{d}}{\mathrm{d}s}\eta_1^\top a(x+s\xi)\eta_2\Big|_{s=0} = 0 \text{ for all } x \in E, \eta_1, \eta_2 \in \mathcal{U}^\perp\right\} \tag{36}$$

holds, with equality if $\phi + \psi^\top\theta \neq 0$ and $\psi \in \operatorname{span}\{(\kappa^\top)^p\psi : p = 1, \ldots, d\}$. Finally, the identity

$$\eta_1^\top A\eta_2 = \frac{1}{2}\Big[\eta_1^\top A\eta_1 + \eta_2^\top A\eta_2 - (\eta_1 - \eta_2)^\top A(\eta_1 - \eta_2)\Big],$$

valid for any symmetric matrix A, implies that the right side of (36) is equal to

$$\left\{\xi \in \mathcal{U} : \frac{\mathrm{d}}{\mathrm{d}s}\eta^\top a(x+s\xi)\eta\Big|_{s=0} = 0 \text{ for all } x \in E, \eta \in \mathcal{U}^\perp\right\}.$$

Since this set is equal to $\mathcal{U} \cap \bigcap_{\eta \in \mathcal{U}^\perp} \ker \eta^\top a(\cdot)\eta$, the proof is complete. □

Remark A.8. *Note that the set on the right side in Proposition A.7 is equal to*

$$\bigcap_{\eta_1, \eta_2 \in \mathcal{U}^\perp} \ker \eta_1^\top a(\cdot)\eta_2 \ \cap\ \mathcal{U},$$

see Equation (36) *in the proof of Proposition A.7.*

Building on this remark we now discuss how the USV factors and residual factors affect the volatility of the transformed factor process $\widehat{X}_t = (Z_t, U_t)$, where $U_t = (V_t, W_t)$. The dynamics of $\widehat{X}_t$ can be written

$$\mathrm{d}\begin{pmatrix} Z_t \\ U_t \end{pmatrix} = \begin{pmatrix} \widehat{\kappa}_{ZZ}(\widehat{\theta}_Z - Z_t) \\ \widehat{\kappa}_{UZ}(\widehat{\theta}_Z - Z_t) + \widehat{\kappa}_{UU}(\widehat{\theta}_U - U_t) \end{pmatrix} \mathrm{d}t + \widehat{\sigma}(Z_t, V_t, W_t)\mathrm{d}B_t,$$

where $\widehat{\kappa}$ and $\widehat{\theta}$ are given by (9), and $\widehat{\sigma}(z,v,w) = S\sigma(S^{-1}(z,v,w))$. The corresponding diffusion matrix is

$$\widehat{a}(z,v,w) = \widehat{\sigma}\widehat{\sigma}^\top(z,v,w) = Sa(S^{-1}(z,v,w))S^\top.$$

Consider now those components of $\widehat{a}(z,v,w)$ that are related to the volatility and covariation of the term structure factors Z_t,

$$\widehat{a}_{ij}(z,v,w) = e_i^\top \widehat{a}(z,v,w)e_j = \eta_i^\top a(S^{-1}\widehat{x})\eta_j, \qquad i,j \in \{1,\ldots,m\},$$

where we set $\eta_i = S^\top e_i \in \mathcal{U}^\perp$. In view of Proposition A.7 and Remark A.8, we see that in the generic case, the functions $\widehat{a}_{ij}(z,v,w)$ are all constant in w, but not all constant in v. In other words, the volatilities and covariations of the term structure factors Z_t are not directly affected by the residual factors W_t, but are affected by the USV factors V_t. This fact provides a further justification for our terminology.

Example A.9. We now provide an example illustrating that residual factors may affect the distribution of future bond prices, despite having no instantaneous impact on the current term structure or bond return volatilities.

Consider a three-factor linear-rational model with state space $E = \mathbb{R}_+ \times \mathbb{R}^2$ and factor process given by

$$\begin{aligned}
\mathrm{d}X_{1t} &= \lambda_1(1 - X_{1t})\mathrm{d}t + X_{1t}\varphi(X_{2t})\mathrm{d}B_{1t} \\
\mathrm{d}X_{2t} &= -\lambda_2 X_{2t}\mathrm{d}t + X_{3t}\mathrm{d}B_{2t} \\
\mathrm{d}X_{3t} &= -\lambda_3 X_{3t}\mathrm{d}t + \mathrm{d}B_{3t},
\end{aligned}$$

where $\lambda_i > 0$ ($i = 1,2,3$), and $\varphi : \mathbb{R} \to [1,2]$ is a strictly increasing, differentiable function with $s\varphi'(s)$ bounded. Then $x \mapsto x_1\varphi(x_2)$ is Lipschitz continuous, and it follows that there is a unique strong solution to the above equation, starting from any point $x \in E$. Note that X_{1t} necessarily stays nonnegative since its drift is positive and the diffusion component vanishes at the origin. The state price density is taken to be

$$\zeta_t = \mathrm{e}^{-\alpha t}(1 + X_{1t})$$

for some α. The short rate is then given by $r_t = \alpha - \lambda_1(1 - X_{1t})/(1 + X_{1t})$, see (6), and therefore, see (7), we pick

$$\alpha = \alpha^* = \sup_{x\in E} \lambda_1 \frac{1 - x_1}{1 + x_1} = \lambda_1.$$

Since also $\alpha_* = -\lambda_1$, this gives a short rate contained in $[0, 2\lambda_1]$. What are the unspanned factors in this model? The term structure kernel, given by (8), reduces to

$$\mathcal{U} = \{\xi \in \mathbb{R}^3 : \xi_1 = 0\} = \operatorname{span}\{e_2, e_3\}.$$

Furthermore, the transformation S can be taken to be the identity. Thus X_{2t} is a USV factor by Corollary 3.4, since $a_{11}(x+se_2) = x_1^2\varphi(x_2+s)^2$ is non-constant in s. On the other hand, X_{3t} is a residual factor, since $a_{11}(x+se_3) = x_1^2\varphi(x_2)^2$ is constant in s. At the same time, however, $a_{22}(x+se_3)$ varies with s. The message is the following: while a perturbation of the residual factor X_{3t} has no immediate impact on the bond prices or their volatilities and covariations, such a perturbation does affect the volatility of the USV factor X_{2t}. Therefore it affects the future distribution of this factor, and hence also the future distribution of the volatility of the term structure factor X_{1t}. This in turn alters the future distribution of bond prices. In conclusion, derivatives prices may, in general, be sensitive to residual factors, and thus contain information about their current values. This may happen despite the fact that the residual factors are neither term structure factors, nor USV factors.

A.3 Proofs for Section 4

We begin with the following general result regarding linear-rational term structure models whose state space is the nonnegative orthant. It provides us with a canonical form for the LRSQ model which renders it identifiable. We let $\mathbf{1}_p$ denote the vector in $\mathbb{R}^d$ whose first p ($p \le d$) components are ones, and the remaining components are zeros. As before, we write $\mathbf{1} = \mathbf{1}_d$.

Lemma A.10. *Consider any linear-rational term structure model with state space $E = \mathbb{R}^d_+$ and short rate bounded from below. After coordinatewise scaling of the factor process and permutation of its components, it still has the representation (2), and the state price density can be written*

$$\zeta_t = \mathrm{e}^{-\alpha t}(1 + \mathbf{1}_p^\top X_t)$$

for some $p \le d$. The extremal values in (7) are given by $\alpha^* = \max \mathcal{S}$ and $\alpha_* = \min \mathcal{S}$, where

$$\mathcal{S} = \left\{\mathbf{1}_p^\top \kappa\theta, -\mathbf{1}_p^\top \kappa_1, \ldots, -\mathbf{1}_p^\top \kappa_d\right\},$$

and where κ_i denotes the i:th column of κ. Moreover, the submatrix $\kappa_{1\ldots p, p+1\ldots d}$ is zero.

The proof of Lemma A.10 uses the following auxiliary result. We always assume that the state space E is *minimal* in the sense that $\mathbb{P}(X_t \in U \text{ for some } t \geq 0) > 0$ holds for any relatively open subset $U \subset E$.

Lemma A.11. *Assume X is a semimartingale of the form* (2) *whose minimal state space is $\mathbb{R}^d_+$. Then $\kappa_{ij} \leq 0$ for all $i \neq j$.*

Proof. Let $G(\tau, x)$ denote the solution to the linear differential equation

$$\partial_\tau G(\tau, x) = \kappa(\theta - G(\tau, x)), \qquad G(0, x) = x,$$

so that, by Lemma A.1, $\mathbb{E}[X_{\rho+\tau} \mid \mathcal{F}_\rho] = G(\tau, X_\rho)$ holds for any bounded stopping time ρ and any (deterministic) $\tau \geq 0$. Pick $i, j \in \{1, \ldots, d\}$ with $i \neq j$, and assume for contradiction that $\kappa_{ij} > 0$. Then, for $\lambda > 0$ large enough, we have

$$\partial_\tau G_i(0, \lambda e_j) = e_i^\top \kappa(\theta - \lambda e_j) = e_i^\top \kappa\theta - \lambda\kappa_{ij} < -2,$$

where e_i (e_j) denotes the i:th (j:th) unit vector, and G_i is the i:th component of G. By continuity there is some $\varepsilon > 0$ such that $\partial_\tau G_i(\tau, \lambda e_j) \leq -2$ for all $\tau \in [0, 2\varepsilon]$. Hence $G_i(\tau, \lambda e_j) = 0 + \int_0^\tau \partial_\tau G_i(s, \lambda e_j) \mathrm{d}s \leq -2\tau$ for all $\tau \in [0, 2\varepsilon]$. By continuity of $(\tau, x) \mapsto G(\tau, x)$ there is some $r > 0$ such that

$$G_i(\tau, x) \leq -\tau \quad \text{holds for all} \quad \tau \in [0, \varepsilon],\ x \in B(\lambda e_j, r),$$

where $B(x, r)$ is the ball of radius r centered at x. Now define

$$\rho = n \wedge \inf\{t \geq 0 : X_t \in B(\lambda e_j, r)\}, \quad A = \{X_\rho \in B(\lambda e_j, r)\},$$

where n is chosen large enough that $\mathbb{P}(A) > 0$. The assumption that $\mathbb{R}^d_+$ is a minimal state space implies that such an n exists. Then

$$\mathbb{E}\left[\mathbf{1}_A X_{i,\rho+\varepsilon}\right] = \mathbb{E}\left[\mathbf{1}_A \mathbb{E}\left[X_{i,\rho+\varepsilon} \mid \mathcal{F}_\rho\right]\right] = \mathbb{E}\left[\mathbf{1}_A G_i(\varepsilon, X_\rho)\right] \leq -\varepsilon\mathbb{P}(A) < 0,$$

whence $\mathbb{P}(X_{i,\rho+\varepsilon} < 0) > 0$, which is the desired contradiction. The lemma is proved. □

Proof of Lemma A.10. By assumption, the state price density is of the form $\zeta_t = \mathrm{e}^{-\alpha t}(\phi + \psi^\top X_t)$. Since $\phi + \psi^\top x$ is assumed positive on E, we must have $\psi \in \mathbb{R}^d_+$ and $\phi > 0$. Dividing ζ_t by ϕ does not affect any model prices, so we may take $\phi = 1$. Moreover, after permuting and scaling the components, X is still of the form (2) and takes values in E, so we can assume $\psi = \mathbf{1}_p$. The short rate is given by $r_t = \alpha - \rho(X_t)$, where

$$\rho(x) = \frac{\mathbf{1}_p^\top \kappa\theta - \mathbf{1}_p^\top \kappa x}{1 + \mathbf{1}_p^\top x}. \tag{37}$$

To see that $\kappa_{1\ldots p,p+1\ldots d}$ vanishes, suppose $p < d$, pick $j \in \{p+1,\ldots,d\}$, and set $x = \lambda e_j$, where $\lambda > 0$ is arbitrary and e_j is the j:th unit vector. Inserting this choice into (37) yields

$$\rho(\lambda e_j) = \mathbf{1}_p^\top \kappa\theta - \lambda \mathbf{1}_p^\top \kappa_j.$$

Since ρ is bounded above on E we have $\mathbf{1}_p^\top \kappa_j \geq 0$. But by Lemma A.11, we also have $\kappa_{ij} \leq 0$ for all $i \in \{1,\ldots,p\}$, and this implies that $\kappa_{1\ldots p,p+1\ldots d}$ is zero.

The expression for α^* can now be obtained from (7) by observing that for each $x \in E$, $\rho(x)$ is a convex combination of the numbers $\mathbf{1}_p^\top \kappa\theta$, $-\mathbf{1}_p^\top \kappa_1$, …, $-\mathbf{1}_p^\top \kappa_d$. This immediately implies that the maximum of these numbers is an upper bound on $\rho(x)$, and by choosing $x \in E$ suitably, $\rho(x)$ can be made arbitrarily close to this upper bound, which thus equals α^*. The expression for α_* is derived in an analogous manner. The lemma is proved. □

Proof of Theorem 4.1. Assume first that the short rate is bounded from below. Since the factor process X_t remains a square-root process after coordinatewise scaling and permutation of its components, Lemma A.10 then implies that we may take $\zeta_t = \mathrm{e}^{-\alpha t}(1 + \mathbf{1}_p^\top X_t)$ for some $p \leq d$ without loss of generality. Moreover, the submatrix $\kappa_{1\ldots p,p+1\ldots d}$ vanishes, so $(X_{1t},\ldots,X_{pt})$ is an autonomous square-root process on the smaller state space $\mathbb{R}_+^p$. Since ζ_t only depends on the first p components of X_t, the pricing model is unaffected if we exclude the last $d-p$ components, and this proves that we may take $p = d$, as desired.

Conversely, if $\zeta_t = \mathrm{e}^{-\alpha t}(1 + \mathbf{1}^\top X_t)$, the right side of (37) is bounded, implying that the short rate is bounded as well.

Finally, the expressions for α^* and α_* follow directly from Lemma A.10. □

Proof of Proposition 4.2. The assertion about $\dim\mathcal{U}$ follows from Theorem 2.4 and Lemma A.2. To prove that all unspanned factors are USV factors we apply Corollary 3.4. To this end it suffices to consider the entries $\widehat{a}_{ii}(z,u)$ of the diffusion matrix of the transformed factor process $\widehat{X}_t = (Z_t, U_t)$. Using that $\widehat{a}(z,u) = S\sigma\sigma^\top(S^{-1}(z,u))S^\top$, a calculation yields

$$\widehat{a}_{ii}(z,u) = \sigma_i^2 z_i + (\sigma_{m+i}^2 - \sigma_i^2)u_i, \qquad i = 1,\ldots,n,$$

which is non-constant in u_i since $\sigma_i \neq \sigma_{m+i}$. Since this holds for all $i = 1,\ldots,n$, Corollary 3.4 implies that all unspanned factors are USV factors. □

	Mean	Median	Std.	Min	Max
Swap rates					
1 year	3.33	3.04	2.21	0.32	7.51
2 years	3.60	3.63	2.08	0.36	7.65
3 years	3.86	4.07	1.96	0.43	7.70
5 years	4.28	4.40	1.75	0.74	7.75
7 years	4.56	4.65	1.59	1.14	7.77
10 years	4.84	4.91	1.45	1.55	7.80
Swaption IV					
1 year	81.0	76.5	29.2	17.3	224.4
2 years	95.5	92.8	32.5	21.3	212.5
3 years	100.3	97.6	31.8	24.3	208.2
5 years	106.9	102.9	30.8	41.3	203.4
7 years	107.4	102.8	28.3	57.1	206.2
10 years	106.9	103.4	26.9	57.2	208.3

Table 1: Summary statistics.
The table reports the mean, median, standard deviation, minimum, and maximum of each time series. Swap rates are reported in percentages. Swaption normal implied volatilities are reported in basis points. Each time series consists of 827 weekly observations from January 29, 1997 to November 28, 2012.

		<0.08	0-0.01	0.01-0.02	0.02-0.03	0.03-0.04	0.04-0.08
Data	$\hat{\beta}_1$	0.17***	1.03 ***	0.59***	0.21	0.04	-0.04
		(2.77)	(10.10)	(2.94)	(1.55)	(0.34)	(-0.65)
	R^2	0.05	0.46	0.26	0.09	0.00	0.00
LRSQ(3,1)	β_1	0.02	0.59	0.23	0.11	0.04	-0.04
	R^2	0.00	0.50	0.37	0.13	0.01	0.01
LRSQ(3,2)	β_1	0.06	0.47	0.23	0.13	0.06	-0.01
	R^2	0.02	0.48	0.38	0.18	0.03	0.00

Table 2: Level-dependence in volatility of 1-year swap rate.
The table report results from regressing weekly changes in the swaption IV at the 1-year swap maturity on weekly changes in the 1-year swap rate (including a constant). The first column shows results using data where the 1-year swap rate is below 0.08. The second to sixth column shows results conditional on the 1-year swap rate being in the intervals 0-0.01, 0.01-0.02, 0.02-0.03, 0.03-0.04, and 0.04-0.08, respectively. The upper part of the table shows results for the sample with t-statistics, corrected for heteroscedasticity and serial correlation up to 4 lags using the method of Newey and West (1987), in parentheses. The middle and lower part of the table shows the population regression coefficients and R^2s for two model specifications. These are based on simulated time series consisting of 520,000 weekly observations.

	LRSQ(3,1)	*LRSQ(3,2)*	*LRSQ(3,3)*
κ_1	0.0630 (0.0005)	0.0836 (0.0007)	0.0856 (0.0009)
κ_2	0.4377 (0.0035)	0.3598 (0.0030)	0.3393 (0.0027)
κ_3	0.1652 (0.0016)	0.2502 (0.0024)	0.2645 (0.0026)
$\kappa_{2,1}$	-0.1266 (0.0006)	-0.1372 (0.0009)	-0.1406 (0.0010)
$\kappa_{3,2}$	-0.5012 (0.0037)	-0.4138 (0.0031)	-0.3944 (0.0027)
θ_1	0.6709 (0.0260)	0.3022 (0.0146)	0.4765 (0.0163)
θ_2	0.2903 (0.0044)	0.1152 (0.0116)	0.1982 (0.0044)
θ_3	0.8810 (0.0156)	0.4083 (0.0074)	0.2955 (0.0063)
θ_4	0.3275 (0.0229)	0.3436 (0.0179)	0.1629 (0.0096)
θ_5		0.1310 (0.0117)	0.0675 (0.0033)
θ_6			0.1006 (0.0042)
σ_1	0.2269 (0.0039)	0.1901 (0.0039)	0.1441 (0.0056)
σ_2	0.6882 (0.0038)	0.4288 (0.0036)	0.4116 (0.0037)
σ_3	0.1229 (0.0015)	0.1008 (0.0010)	0.0966 (0.0010)
σ_4	1.8097 (0.0157)	1.4648 (0.0118)	0.9054 (0.0100)
σ_5		1.0167 (0.0099)	0.9777 (0.0121)
σ_6			0.4839 (0.0061)
σ_{rates}	7.3765 (0.0539)	5.0105 (0.0542)	4.8279 (0.0498)
$\sigma_{swaptions}$	8.4015 (0.0552)	7.0472 (0.0433)	6.6027 (0.0519)
α	0.0636	0.0540	0.0550
$\sup r_t$	0.2288	0.3042	0.3195
$\log L \times 10^{-4}$	5.4037	5.5930	5.6514

Table 3: Maximum likelihood estimates.

The table reports parameter estimates with asymptotic standard errors are in parentheses. σ_{rates} denotes the standard deviation of swap rate pricing errors and $\sigma_{swaptions}$ denotes the standard deviation of swaption pricing errors in terms of normal implied volatilities. Both σ_{rates} and $\sigma_{options}$ are measured in basis points. α is chosen as the smallest value that guarantees a nonnegative short rate. $\sup r_t$ is the upper bound on possible short rates. $\log L$ denotes the log-likelihood value. The sample period consists of 827 weekly observations from January 29, 1997 to November 28, 2012.

Specification	Full sample		Pre-ZIRP		ZIRP	
	Swaps	Swaptions	Swaps	Swaptions	Swaps	Swaptions
LRSQ(3,1)	5.73	7.31	5.16	7.28	7.43	7.41
LRSQ(3,2)	3.87	5.89	3.13	5.80	6.08	6.19
LRSQ(3,3)	3.74	5.42	3.01	5.21	5.92	6.03
LRSQ(3,2)-LRSQ(3,1)	-1.86^{***}	-1.42^{***}	-2.03^{***}	-1.48^{***}	-1.36^{***}	-1.22^{***}
	(-6.13)	(-7.19)	(-5.62)	(-7.10)	(-2.86)	(-2.73)
LRSQ(3,3)-LRSQ(3,2)	-0.13^{***}	-0.47^{**}	-0.12^{**}	-0.58^{**}	-0.16	-0.16
	(-2.58)	(-2.22)	(-2.14)	(-2.18)	(-1.21)	(-0.48)

Table 4: Comparison of model specifications.
The table reports means of time series of the root mean squared pricing errors (RMSE) of swap rates and normal implied swaption volatilities. Units are basis points. t-statistics, corrected for heteroscedasticity and serial correlation up to 50 lags using the method of Newey and West (1987), are in parentheses. $*$, $**$, and $***$ denote significance at the 10%, 5%, and 1% level, respectively. The full sample period consists of 827 weekly observations from January 29, 1997 to November 28, 2012. The zero interest rate policy (ZIRP) sample period consists of 207 weekly observations after December 16, 2008. The pre-ZIRP sample period consists of 620 weekly observations before December 16, 2008.

Specification	Full sample		Pre-ZIRP		ZIRP	
	Swaps	Swaptions	Swaps	Swaptions	Swaps	Swaptions
LRSQ(3,1)						
1 year	9.03	6.38	6.48	5.80	4.56	6.51
2 years	9.96	7.55	6.34	7.79	7.94	10.30
3 years	8.85	6.01	6.51	4.36	2.85	6.07
5 years	10.29	7.19	5.73	8.01	8.11	10.00
7 years	9.51	7.35	6.37	8.77	7.63	7.69
10 years	8.91	8.52	7.88	7.12	7.39	11.12
LRSQ(3,2)						
1 year	5.19	4.72	4.10	4.85	3.57	4.75
2 years	9.27	6.65	5.24	6.09	4.38	7.52
3 years	4.25	4.39	3.16	3.57	2.61	4.04
5 years	9.56	6.74	4.85	5.78	4.05	6.87
7 years	7.30	5.57	6.08	7.45	5.50	6.41
10 years	8.33	6.39	6.27	6.92	5.24	9.18
LRSQ(3,3)						
1 year	4.95	4.68	3.74	4.63	3.44	4.50
2 years	7.92	6.49	4.77	5.03	4.47	7.47
3 years	3.91	4.27	3.00	3.52	2.52	3.39
5 years	8.25	6.41	3.96	4.78	3.97	5.93
7 years	7.20	5.74	5.36	6.95	5.29	6.81
10 years	6.86	6.74	6.61	5.73	5.71	10.82

Table 5: Individual RMSEs.
The table reports root mean squared pricing errors (RMSE) of individual swap rates and normal implied swaption volatilities. Units are basis points. The full sample period consists of 827 weekly observations from January 29, 1997 to November 28, 2012. The zero interest rate policy (ZIRP) sample period consists of 207 weekly observations after December 16, 2008. The pre-ZIRP sample period consists of 620 weekly observations before December 16, 2008.

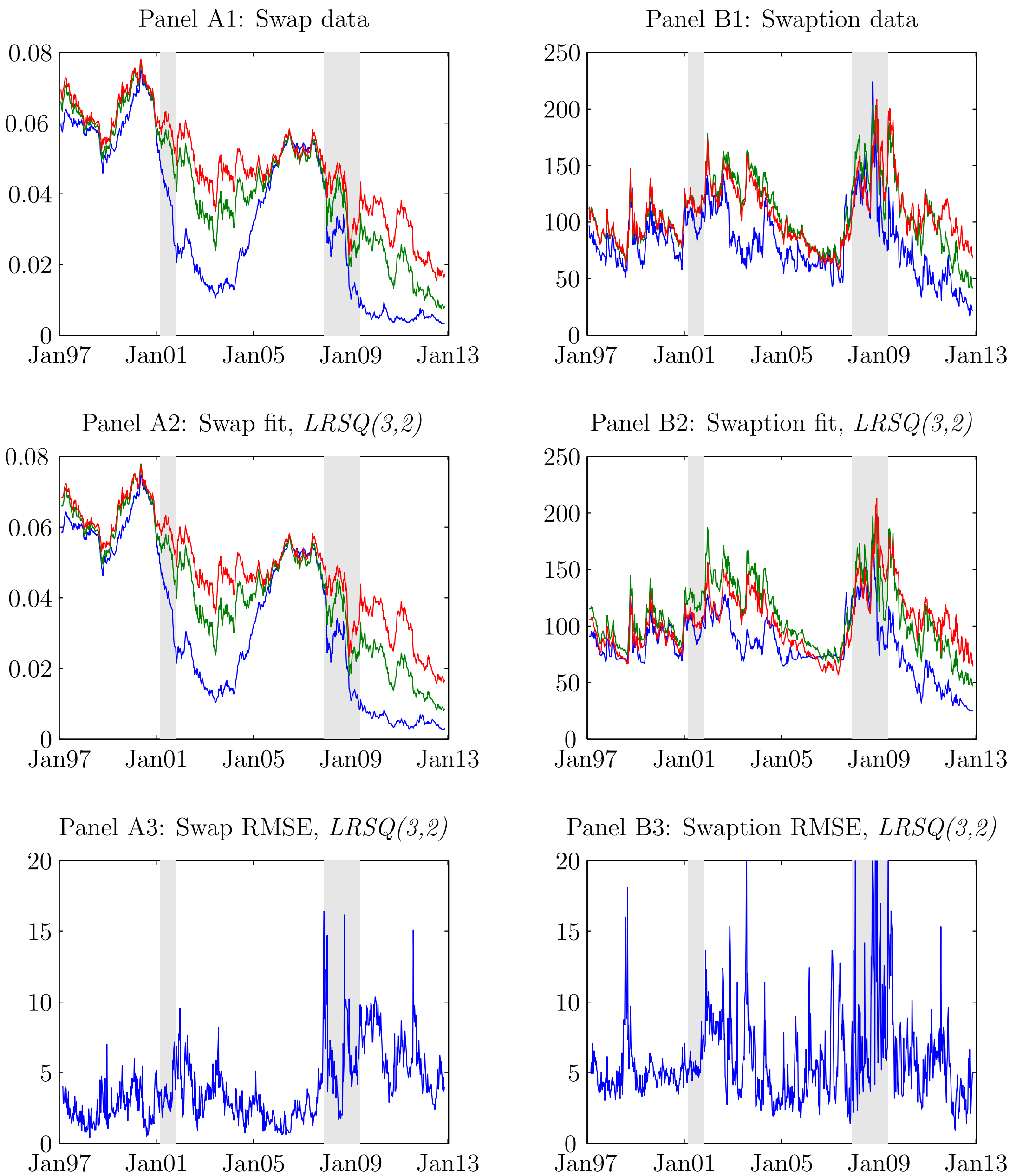

Figure 1: Data and fit
Panel A1 shows time series of the 1-year, 5-year, and 10-year swap rate. Panel B1 shows time series of the normal implied volatilities on 3-month options on the 1-year, 5-year, and 10-year swap rate. Panels A2 and B2 shows the fit to the swap rates and swaption implied volatilities in case of the *LRSQ(3,2)* specification. Panels A3 and B3 shows time series of the root mean squared pricing errors (RMSE) of swap rates and swaption implied volatilities, respectively. The units in Panels B1, B2, A3, and B3 are basis points. The grey areas mark the two NBER-designated recessions from March 2001 to November 2001 and from December 2007 to June 2009, respectively. Each time series consists of 827 weekly observations from January 29, 1997 to November 28, 2012.

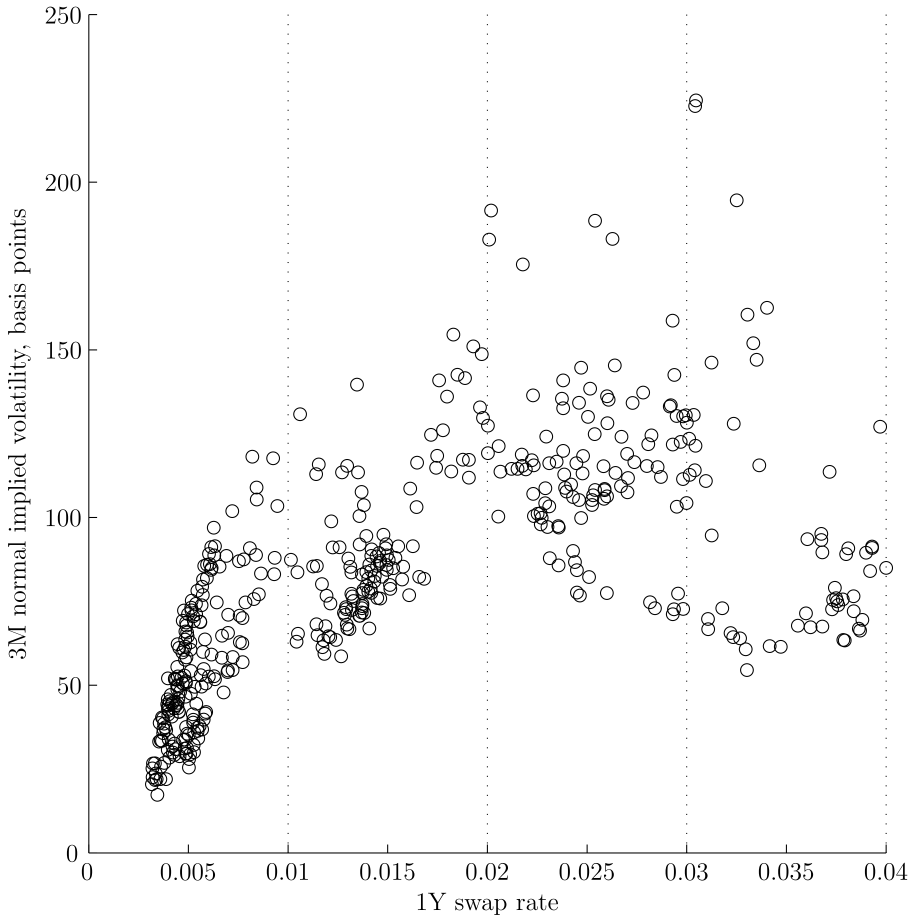

Figure 2: Level-dependence in volatility of 1-year swap rate
The figure shows the normal implied volatility of the 3-month option on the 1-year swap rate (in basis points) plotted against the 1-year swap rate.

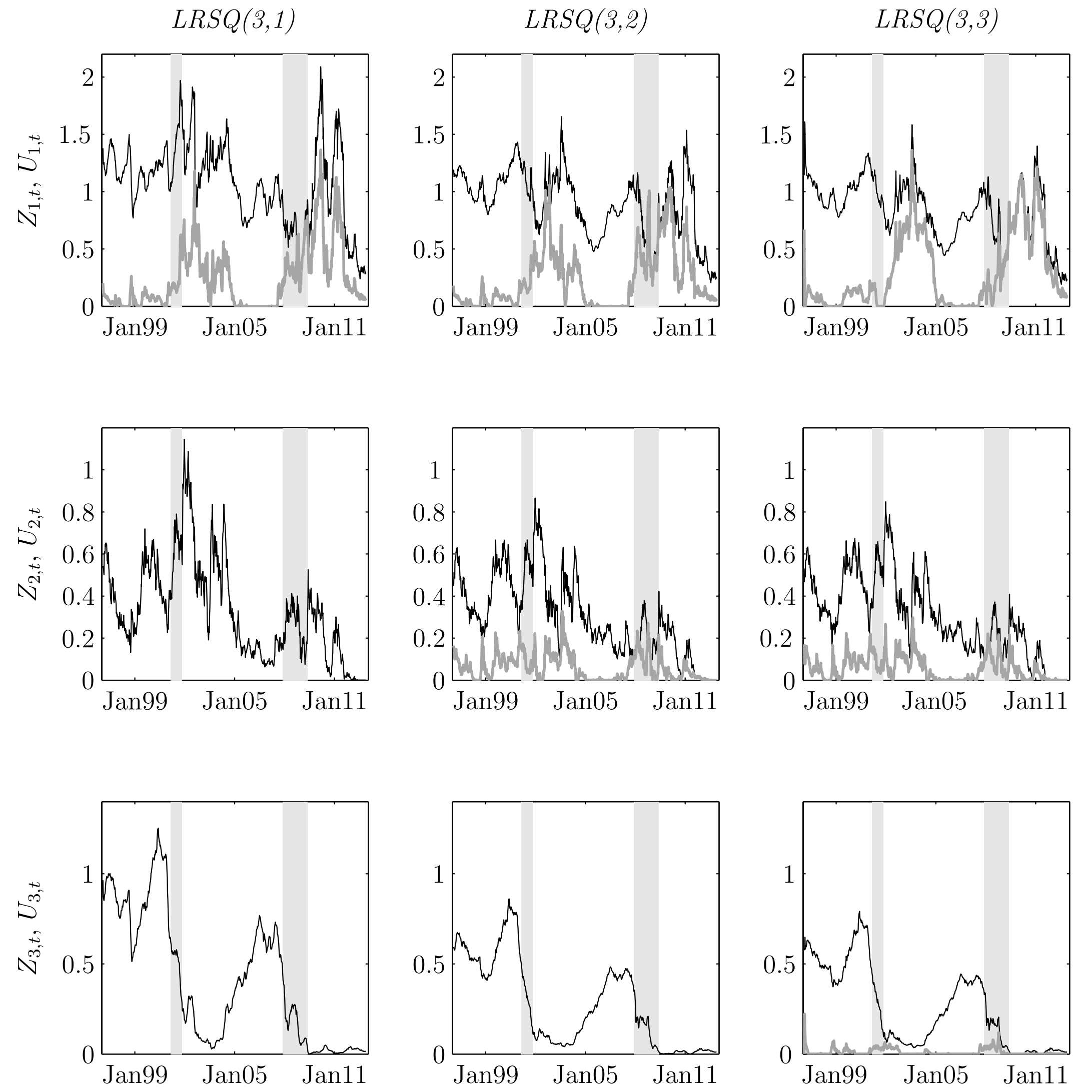

Figure 3: Estimated factors
The figure displays time series of the estimated factors. The first, second, and third column shows the factors of the *LRSQ(3,1)*, *LRSQ(3,2)*, and *LRSQ(3,3)* specification, respectively. The first row displays $Z_{1,t}$ and $U_{1,t}$. The second row displays $Z_{2,t}$ and possibly $U_{2,t}$. The third row displays $Z_{3,t}$ and possibly $U_{3,t}$. The thin black lines show the term structure factors, $Z_{1,t}$, $Z_{2,t}$, and $Z_{3,t}$. The thick grey lines show the unspanned stochastic volatility factors, $U_{1,t}$, $U_{2,t}$, and $U_{3,t}$. The grey areas mark the two NBER-designated recessions from March 2001 to November 2001 and from December 2007 to June 2009, respectively. Each time series consists of 827 weekly observations from January 29, 1997 to November 28, 2012.

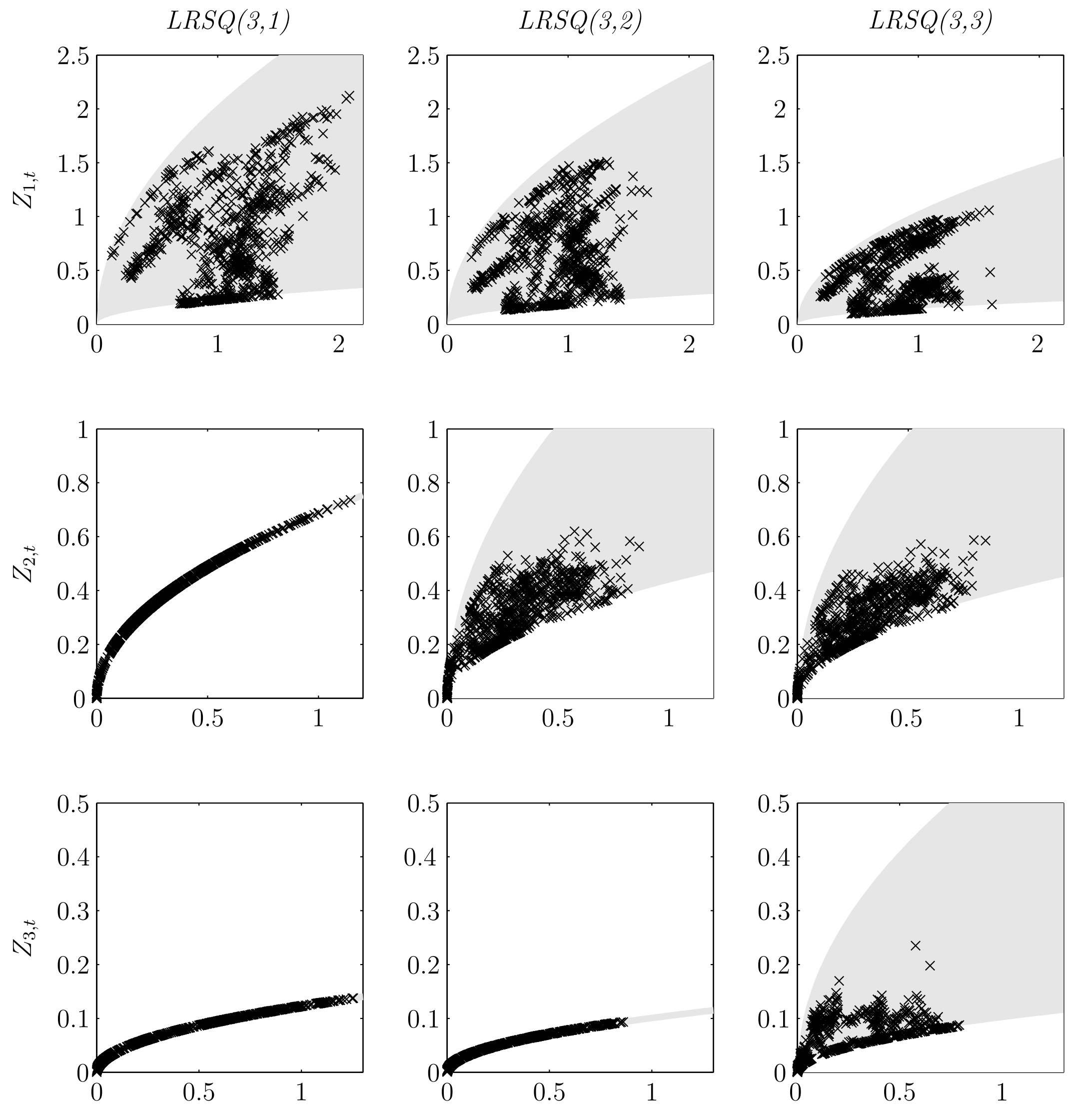

Figure 4: Level-dependence in volatility of the term structure factors
For each term structure factor, its instantaneous volatility is plotted against its level. The first, second, and third column correspond to the *LRSQ(3,1)*, *LRSQ(3,2)*, and *LRSQ(3,3)* specification, respectively. The first, second, and third row correspond to $Z_{1,t}$, $Z_{2,t}$, and $Z_{3,t}$, respectively. Each plot contains 827 weekly observations from January 29, 1997 to November 28, 2012. The grey areas mark the possible range of factor volatilities for a given factor level.

References

Ait-Sahalia, Y. (1996). Testing Continuous-Time Models of the Spot Interest Rate. *Review of Financial Studies 9*, 385–426.

Andersen, T. G. and L. Benzoni (2010). Do bonds span volatility risk in the U.S. Treasury market? A specification test for affine term structure models. *Journal of Finance 65*, 603–653.

Bateman, H. and A. Erdélyi (1954). *Tables of integral transforms: Based, in part, on notes left by Harry Bateman*. New York: McGraw-Hill.

Bauer, M. and G. Rudebusch (2013). Monetary policy expectations at the zero lower bound. Working paper, Federal Reserve Bank of San Francisco.

Bikbov, R. and M. Chernov (2009). Unspanned stochastic volatility in affine models: Evidence from Eurodollar futures and options. *Management Science 55*, 1292–1305.

Black, F. (1995). Interest rates as options. *Journal of Finance 50*, 1371–1376.

Carr, P., X. Gabaix, and L. Wu (2009). Linearity-generating processes, unspanned stochastic volatility, and interest-rate option pricing. Working paper, New York University.

Chan, K., G. Karolyi, F. Longstaff, and A. Sanders (1992). An empirical comparison of alternative models of the short-term interest rate. *Journal of Finance 47*, 1209–1227.

Christensen, J. and G. Rudebusch (2013). Estimating shadow-rate term structure models with near-zero yields. Working paper, Federal Reserve Bank of San Francisco.

Christoffersen, P., K. Jacobs, L. Karoui, and K. Mimouni (2009). Non-linear filtering in affine term structure models: Evidence from the term structure of swap rates. Working paper, McGill University.

Collin-Dufresne, P. and R. Goldstein (2002). Do bonds span the fixed income markets? Theory and evidence for unspanned stochastic volatility. *Journal of Finance 57*, 1685–1730.

Collin-Dufresne, P., R. Goldstein, and C. Jones (2009). Can interest rate volatility be extracted from the cross section of bond yields? *Journal of Financial Economics 94*, 47–66.

Conley, T. G., L. P. Hansen, E. Luttmer, and J. A. Scheinkman (1997). Short-term interest rates as subordinated diffusions. *Review of Financial Studies 10*, 525–577.

Constantinides, G. (1992). A theory of the nominal term structure of interest rates. *Review of Financial Studies 5*, 531–552.

Corp, H. (2012). *Interest Rate Swaps and Other Derivatives*. Columbia University Press.

Dai, Q. and K. Singleton (2000). Specification analysis of affine term structure models. *Journal of Finance 55*, 1943–1978.

Duffie, D. and R. Kan (1996). A yield-factor model of interest rates. *Mathematical Finance 6*, 379–406.

Fan, R., A. Gupta, and P. Ritchken (2003). Hedging in the possible presence of unspanned stochastic volatility: Evidence from swaption markets. *Journal of Finance 58*, 2219–2248.

Filipović, D. (2009). *Term Structure Models – A Graduate Course*. Springer.

Filipovic, D. and A. Trolle (2013). The term structure of interbank risk. *Journal of Financial Economics 109*, 707–733.

Flesaker, B. and L. P. Hughston (1996). Positive Interest. *Risk 9*(1), 46–49.

Gabaix, X. (2009). Linearity-generating processes: A modelling tool yielding closed forms for asset prices. Preprint.

Heidari, M. and L. Wu (2003). Are interest rate derivatives spanned by the term structure of interest rates? *Journal of Fixed Income 13*, 75–86.

Horn, R. and C. Johnson (1990). *Matrix Analysis*. Cambridge University Press.

Jacobs, K. and L. Karoui (2009). Conditional volatility in affine term structure models: Evidence from Treasury and swap markets. *Journal of Financial Economics 91*, 288–318.

Johannes, M. and S. Sundaresan (2007). The impact of collateralization on swap rates. *Journal of Finance 62*, 383–410.

Kim, D. H. and K. J. Singleton (2012). Term structure models and the zero bound: An empirical investigation of Japanese yields. *Journal of Econometrics 170*, 32–49.

Leippold, M. and L. Wu (2007). Design and estimation of multi-currency quadratic models. *Review of Finance 11*, 167–207.

Li, H. and F. Zhao (2006). Unspanned stochastic volatlity: Evidence from hedging interest rate drivatives. *Journal of Finance 61*, 341–378.

Li, H. and F. Zhao (2009). Nonparametric estimation of state-price densities implicit in interest rate cap prices. *Review of Financial Studies 22*, 4335–4376.

Newey, W. and K. West (1987). A simple, positive semi-definite, heteroscedasticity, and autocorrelation consistent covariance matrix. *Econometrica 55*, 703–708.

Revuz, D. and M. Yor (1999). *Continuous martingales and Brownian motion* (3rd ed ed.), Volume 293. Berlin: Springer.

Rogers, L. C. G. (1997). The potential approach to the term structure of interest rates and foreign exchange rates. *Mathematical Finance 2*, 157–164.

Rudin, W. (1987). *Real and Complex Analysis* (3rd ed ed.). New York: McGraw-Hill.

Stanton, R. (1997). A nonparametric model of term structure dynamics and the market price of interest rate risk. *Journal of Finance 52*, 1973–2002.

Trolle, A. B. and E. S. Schwartz (2009). A general stochastic volatility model for the pricing of interest rate derivatives. *Review of Financial Studies 22*, 2007–2057.

Trolle, A. B. and E. S. Schwartz (2013). The swaption cube. Working paper, EPFL and UCLA.